Decision Feedback Multiuser Detection: A Systematic Approach

Mahesh K. Varanasi, *Senior Member, IEEE*

***Abstract*— A systematic approach to decision feedback multiuser detection is introduced for the joint detection of symbols of K simultaneously transmitting users of a synchronous correlated waveform multiple-access (CWMA) channel with Gaussian noise. A new performance criterion called symmetric energy is defined which is a low-noise indicator of the joint error rate that at least one user is detected erroneously. Even the best linear detectors can perform poorly in terms of symmetric energy compared to the maximum-likelihood detector. A general class of decision feedback detectors is defined with $O(K)$ implementational complexity per user. The symmetric energy of arbitrary DFD's and bounds on their asymptotic effective energy (AEE) performance are obtained along with an exact bit-error rate and AEE analysis for the decorrelating DFD. The optimum DFD that maximizes symmetric energy is obtained. Each one of the $K!$ optimum, decorrelating, and conventional DFD's, that correspond to the $K!$ orders in which the users can be detected, are shown to outperform the linear optimum, decorrelating, and conventional detectors, respectively, in terms of symmetric energy. Moreover, algorithms are obtained for determining the choice of order of detection for the three DFD's which guarantee that they uniformly (user-wise) outperform their linear counterparts. In addition to optimality in symmetric energy, it is also shown that under certain conditions, the optimum DFD achieves the AEE performance of the exponentially complex maximum-likelihood detector for all users simultaneously. None of the results of this paper make the perfect feedback assumption. The implications of our work on power control for multiuser detection are also discussed.**

***Index Terms*— Decision feedback, error analysis, minimax methods, multiaccess communication, multiuser detection.**

I. Introduction

Multiuser detection for the symbol-synchronous Gaussian correlated waveform multiple-access (CWMA) channel[1] has already received considerable attention for over a decade now, and justifiably so: its simplicity allows a rigorous and systematic study of the subject, which in turn provides new research directions for the analysis and design of multiuser receivers for more complex and realistic systems that may involve noncoherent reception, symbol asynchronism, coding, fading, multipath, etc. Such research will determine the design of receivers for practical wireless communication systems that must operate in the presence of interuser interference (IUI). IUI may be intentional in CWMA and bandwidth-efficient multiple-access (BEMA) systems as proposed by the author and Guess in [47], or unintentional in cellular TDMA systems, where it arises due to channel distortion and/or intercell interference resulting from frequency reuse (cf. Haas and Belfiore [15], Caire *et al.* [3]).

Tutorial articles on multiuser detection are available with extensive reference lists (cf. Verdú [51], Duel-Hallen [8], and Moshavi [23]). Our brief survey here offers another perspective with a discussion of a few key results while focusing on linear modulation and the Gaussian multiuser channel.

Optimum coherent multiuser detectors for the symbol-synchronous case can be obtained by applying the theory of minimum probability of error detection for simple hypothesis testing (cf. Poor [25], Lupas and Verdú [20]). The indecomposable upper bound on bit-error rate of the maximum-likelihood detector is due to Verdú [49] and a bit-error rate analysis for that detector can be found in Varanasi and Aazhang [37]. Maximum-likelihood detection is NP-hard [20] unless the signal correlations have a special structure as was found for nonpositive correlations by Ulukus and Yates in [33]. Earlier work in multiuser detection by Etten [11] and Verdú [49] considered optimum sequence detection for the asynchronous multiuser channel via the Viterbi algorithm. Suboptimal, lower complexity alternatives are the sequential decoding algorithm by Xie, Rushforth, and Short [57] and the cyclic decision feedback sequence detector by Fain and Varanasi [12]. In [49], it was recognized that optimum bit-error rate performance for each user in the multiuser Gaussian channel is akin to performance achievable over a single-user Gaussian channel with a fraction of the actual energy. This fraction was called asymptotic efficiency [50]. The asymptotic efficiency of the cyclic decision feedback sequence detector was studied in [12].

Suboptimum multiuser detection based on interference cancellation from tentative decisions was introduced in the form of multistage detection by Varanasi and Aazhang in [34]–[37] (see also Kohno, Imai, and Hatori [19]). The two-stage detector makes a decision at the second stage on a user's symbol by estimating and then subtracting from that user's matched

Manuscript received March 11, 1997; revised June 15, 1998. This work was supported in part by the NSF under Grants NCR-9406069 and NCR-9725778. The material in this paper was presented in part at the 31st Annual Conference on Information Sciences and Systems, Johns Hopkins University, Baltimore, MD, March 1997, and at the IEEE International Symposium on Information Theory, Ulm, Germany, June–July, 1997.

The author is with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309 USA.

Communicated by R. Kohno, Associate Editor for Detection.

Publisher Item Identifier S 0018-9448(99)00056-5.

[1] The name CWMA is used instead of the usual code-division multiple access (CDMA). Even though the name CDMA is commonly used in the multiuser detection literature, CWMA describes the multiple-access technique under consideration more accurately. It also avoids the confusion that results from using the same name CDMA for both CWMA and the CDMA technique described by the IS-95 system [32] and Qualcomm's embodiment of it [53].

filter output, the multiple-access interference contributed by other users based on tentative decisions of the conventional (decorrelating) detector made at the first stage. Since the desired signal-to-noise ratio (SNR) is highest at the matched-filter output, it is the effect of interference doubling from users that are incorrectly detected at the penultimate stage, that ultimately limits the performance of the multistage detector. This was quantified via an error probability analysis for the two-stage detector for the general K-user channel in [35] and [37]. There has been considerable research on multistage detection since then (cf. Hegarty and Vojcic [16], Gray, Kocic, and Brady [14], Ghazi-Moghadam, Nelson, and Kaveh [13], Shi, Du, and Driessen [30], Buehrer and Woerner [2], Divsalar, Simon, and Raphaeli [6], and Zhang and Brady [62]).

Meanwhile, the simple decorrelating detectors of Lupas and Verdú [20] (see also Schneider [27]) for coherent detection and Varanasi and Aazhang [38] for differentially coherent detection were obtained by optimizing near–far resistance among linear and bilinear detectors, respectively. Near–far resistance is defined as the worst asymptotic efficiency over all the interfering users' energies for coherent detection and over complex amplitudes for differentially coherent reception. Those decorrelating detectors were also found to achieve the optimum near–far resistance achievable by the exponentially complex coherent maximum-likelihood detector. The linear detector that optimizes asymptotic efficiency was also obtained in [20]. In asynchronous multiuser channels, the decorrelating detectors generalize to multi-input, multi-output equalizers that perform the inversion of the channel transfer matrix as obtained for baseband channels by Lupas and Verdú in [21] and for passband channels with differentially coherent detection by Varanasi in [40]. These equalizers were also obtained by optimizing near–far resistance. The idea of decorrelation has also received considerable attention in subsequent research (cf. Iltis and Mailaender [17], Tsatsanis [31], Juntti and Aazhang [18], and Wang and Poor [54]).

Based on the more classical minimum mean-square error (MMSE) criterion, linear MMSE multiuser detectors were derived and their performance was studied for both synchronous and asynchronous channels (cf. Xie, Short, and Rushforth [58], Yang and Roy [60], Madhow and Honig [22], and Poor and Verdú [26]). While achieving the same asymptotic (high-SNR) performance as the decorrelator, the MMSE detector mitigates to some extent, for low–medium SNR's and for systems where the channel is not invertible, the interference limitedness of the decorrelator. While decorrelative and MMSE strategies are at once simple and attractive from a performance standpoint for low–moderate waveform correlations, their performance can degrade substantially for systems with higher bandwidth efficiencies. The problem here is that the structural constraint that the detector be linear, is too severe.

Group detection was introduced in [39] and [43], where a group or subset of users are jointly detected. Rather than pursuing an approach based on performance optimization, the classical theory of detection was revisited to obtain suboptimal solutions. In particular, it was proposed that the generalized likelihood ratio test (cf. [25]) be applied under the assumption (pretense) that the energies (or complex amplitudes) of users not in the group are unknown. The resulting group detectors involve projecting the received signal onto the orthogonal complement of the signal subspace spanned by interfering users. They were found to optimize the worst case asymptotic efficiency over signal energies (or complex amplitudes) of users not in that group, also known as group near–far resistance [39]. Incidentally, group detectors lend themselves to a simple interpretation as being a cascade of a decorrelator followed by an optimum detector for the outputs of the users in the group of interest. The particular case of unit group size specializes to the decorrelator and gave a new justification for it as a generalized likelihood detector when the energies (or complex amplitudes) of all other interfering users were unknown. As the group size is varied, group detectors achieve a performance ranging between and including those of the decorrelator and the optimum detector [39] for a commensurate change in complexity. Group sequence detection generalizes group detection to asynchronous channels [42]. In particular, in a cascade structure of a linear filter followed by a Viterbi-like sequence detection algorithm, the problem of optimizing the linear pre-equalizer with a fixed complexity of the sequence detection algorithm was addressed therein. The idea of projecting the received signal onto the orthogonal complement of the interfering signal subspace was also studied for synchronous channels as matched subspace detectors by Scharf and Friedlander in [28] and as projection receivers by Schlegel, Roy, Alexander, and Xiang in [29].

The idea of successive cancellation dates back to Wyner [56] and Cover [4], and was proposed in the context of an information-theoretic study of a scalar-output Gaussian multiaccess channel. It involves decoding the users sequentially in a given order. The first user is decoded by regarding the interference from other users as noise. The decoded and re-encoded symbols of the first user are then subtracted from the received data and the second user is decoded by regarding the interference from the remaining users as noise, and so on. Interestingly, this strategy achieves the total capacity of the Gaussian multiaccess channel. The application of that idea to a multiple-access channel with low-rate convolutional codes can be found in Viterbi [52] (see also Yoon, Kohno, and Imai [61]). Following the ideas of successive cancellation and using the insights on the performance of multistage detection given in [37], the decorrelating decision feedback detector (D-DFD) was proposed by Duel-Hallen [7]. While it was described in [7] by a pair of feedforward–feedback transformations with a feedforward transformation to whiten the noise and a feedback transformation to eliminate the interference from already detected users, the D-DFD can more fruitfully be interpreted as a detector that cancels interference from stronger users based on hard decisions of those users and from weaker users by decorrelation in a user-expurgated channel. The latter involves a signal-to-noise ratio penalty which is easily quantified. The former involves a penalty due to interference doubling from incorrect decisions which may be particularly significant when the stronger users are not sufficiently strong. More recently, the maximum signal-to-interference ratio (SIR) (or equivalently, the minimum mean-square error) decision feedback detector was obtained by Varanasi and Guess [48]. In particular, [48] considers the problem with coding and shows that maximum SIR equalization with successive decoding at the receiver achieves the total capacity at the vertices of the capacity region of the vector Gaussian multiple-access channel.

For asynchronous channels, the multivariate decision feedback equalizer that minimizes total mean-square error (as well as the geometric mean-squared error) was obtained by Yang and Roy [59]. Generalizing the D-DFD of the synchronous channel, a zero-forcing multiuser decision feedback equalizer (MDFE) was obtained by Duel-Hallen in [9] based on a classical result in minimum-phase multivariate spectral factorization by Wiener and Masani [55] and Davis [5]. The noise-whitening approach of [9] and the total MSE performance optimization in [59] can be regarded as multivariate extensions of the zero-forcing and MMSE decision feedback equalization for single-user ISI channels. Neither of those works, however, account for the effects of error propagation in analysis or design. While ignoring the effects of error propagation finds justification for single-user ISI channels (cf. Duttweiler *et al.* [10], Altekar and Beaulieu [1]), the error propagation problem is quite different in multiuser channels. The MDFE based on the so-called partial spectral factorization by Duel-Hallen [9] and the MDFE proposed by Varanasi in [44], with the latter involving only minimum-phase spectral factorizations of certain scalar-valued spectra, address the error propagation issue through a careful specification of the MDFE's to reduce the effects of interference doubling due to imperfect feedback. For the MDFE in [44], the problem of optimum power control under average power constraints was solved for the two-user case without making the assumption of perfect feedback.

In this paper, a new performance-optimization-based approach is introduced for decision feedback multiuser detection which includes all users' performances at once without the option of power control and without making the perfect feedback assumption. Furthermore, while a decision feedback multiuser detector is usually understood as detecting the users in the decreasing order of signal energies, this paper takes the view that there are indeed $K!$ decision feedback detectors, one for every order in which users can be detected. The order of detection is shown to be a valuable degree of freedom (with no counterpart in single-user ISI channels) if it can be exploited properly. Finally, the issue of a comparative performance analysis between a linear detector and its nonlinear decision feedback counterpart is resolved through results that take all users' performances into account at once.

The rest of this paper is organized as follows. Section II contains a brief review of multiuser detection for the symbol-synchronous channel. In Section III, we define a new performance measure called *symmetric energy* which is an indicator of the overall performance of a multiuser detector in terms of the joint error rate (JER) defined as the probability that at least one user is detected erroneously. Section III-A gives a detailed study of nondecision-feedback multiuser detectors via the symmetric energy measure for the two-user channel with and without power control. It also provides the motivation for the rest of this paper. Section IV is on decision feedback multiuser detection. In Section IV-A, we define a class of decision feedback detectors (DFD's) that includes as particular cases, all linear detectors, the conventional DFD (C-DFD), also known as the successive canceler, the decorrelating DFD (D-DFD) [7], and the maximum signal-to-interference ratio DFD (MSIR-DFD) [48]. Sections IV-B through IV-D contain several results on the D-DFD including an exact bit-error rate (BER) and asymptotic effective energy (AEE) analysis, the robustness of D-DFD in terms of symmetric energy to the selection of order of detection, and an algorithm for determining an order that guarantees superior user-wise AEE performance compared to the decorrelator. In Section IV-E, we obtain a closed-form expression for the symmetric energy of arbitrary DFD's. In Section IV-F, we obtain the optimum DFD (O-DFD) that maximizes symmetric energy among all DFD's. In Section IV-G, the O-DFD is shown to be robust in terms of symmetric energy to the choice of order of detection. Section IV-H contains a detailed study of decision feedback detectors relative to optimum and linear optimum detectors for the two-user channel with and without power control. Section IV-I gives an algorithm for determining the order of detection so that the O-DFD (C-DFD) user-wise outperforms the linear optimum (conventional) detector in the AEE measure. Section IV-J consists of a per-user AEE analysis of the O-DFD for the two-user channel. Sections IV-K derives sufficient conditions under which the O-DFD achieves the per-user AEE performance of its genie-aided version which has perfect feedback. Section IV-L gives sufficient conditions under which the O-DFD achieves the AEE performance of the maximum-likelihood detector for every user simultaneously. Section V concludes this paper.

II. Preliminaries

This section describes the CWMA channel model. The maximum-likelihood (ML), the two-stage [37], and the linear detectors of [20] and their per-user bit-error rate (BER) and asymptotic effective energy (AEE) performances are given.

A. System Model

A discrete-time equivalent model for the matched-filter outputs at the receiver is given by the K-length vector

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{b} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{b} \in \{-1, +1\}^K$ denotes the K-length vector of bits transmitted by the K active users (the results in this paper can be extended to arbitrary PAM signaling). The matrix $\boldsymbol{H} = \boldsymbol{W}^{1/2}\boldsymbol{R}\boldsymbol{W}^{1/2}$ is a signature waveform correlation matrix and is thus nonnegative definite. $\boldsymbol{R}$ is the normalized correlation matrix so that $\mathrm{diag}\{\boldsymbol{R}\} = \boldsymbol{I}$. $\boldsymbol{W}$ is a diagonal matrix whose kth diagonal element w_k is the received signal energy per bit of the kth user. $\boldsymbol{n}$ is a real-valued zero-mean Gaussian random vector with a covariance matrix that is equal to $\sigma^2\boldsymbol{H}$. This model applies to baseband [20] and passband [43] additive Gaussian noise channels. In the passband channel, $\boldsymbol{R}$ depends on the carrier phases of all users. Models such as $\boldsymbol{z} = \boldsymbol{G}\boldsymbol{b} + \boldsymbol{w}$ where $\boldsymbol{w}$ is zero-mean Gaussian with a full-rank correlation matrix $\sigma^2\boldsymbol{F}$ can be reduced to the above form by forming the sufficient statistic $\boldsymbol{G}^T\boldsymbol{F}^{-1}\boldsymbol{z}$ which admits a model identical to that in (1) with $\boldsymbol{H} = \boldsymbol{G}^T \boldsymbol{F}^{-1} \boldsymbol{G}$. In particular, the Rician fading channel described in [41] reduces to the model in (1) with the matrix $\boldsymbol{H}$ depending on the ratios of fading-to-additive noise variances in addition to the signal energies and correlations.

In the examples of this paper, we consider heavily correlated signaling. Not only is it easier to clearly distinguish between the various suboptimum detectors in this case, but highly correlated (even linearly dependent) signals result when spreading and coding are properly traded as in bandwidth-efficient multiple-access (BEMA) systems [47].

B. Optimum, Linear, and Two-Stage Detectors

This section is a brief review of [20] and [37]. A multiuser detector for the CWMA model in (1) can be described as a vector-valued map ϕ: $R^K \to \{-1, +1\}^K$ that acts on $\boldsymbol{y}$ to produce one of 2^K possible bit decisions for all users jointly according to $\hat{\boldsymbol{b}} = \phi(\boldsymbol{y})$ whose kth element we denote as $\hat{b}_k = \phi^k(\boldsymbol{y})$ for $k = 1, 2, \cdots, K$. For instance, the maximum-likelihood detector [20] is the map

$$\phi_{\mathrm{ML}}: \quad \hat{\boldsymbol{b}} = \arg \max_{\boldsymbol{b} \in \{-1, +1\}^K} 2\boldsymbol{y}^T \boldsymbol{b} - \boldsymbol{b}^T \boldsymbol{H} \boldsymbol{b} \qquad (2)$$

which minimizes among all detectors the probability that not all users' decisions are correct. However, ϕ_{ML} is in general exponentially complex to implement. We therefore restrict attention to multiuser detectors that are easily implementable. The simplest restriction is seen in a linear detector which, parameterized by a matrix $\boldsymbol{F}$, makes decisions separately for each user by taking the sign of an inner product of the kth row of $\boldsymbol{F}$ (denoted as $\boldsymbol{f}_k^T$) and $\boldsymbol{y}$ so that $\hat{b}_k = \mathrm{sgn}\,[\boldsymbol{f}_k^T \boldsymbol{y}]$ for $k = 1, 2, \cdots, K$. The particular case where $\boldsymbol{F}$ is equal to $\boldsymbol{I}$ (the identity matrix), or $\boldsymbol{Q} \triangleq \boldsymbol{H}^{-1}$, or $(\boldsymbol{H} + \sigma^2 \boldsymbol{I})^{-1}$, is called the conventional, or decorrelating [20], or the minimum mean-squared error (MMSE) detector [58], respectively. We denote these detectors as ϕ_C, ϕ_D, and ϕ_{MMSE}. The two-stage detector with decorrelating first stage [37], denoted as ϕ_{2S}, is defined according to the maps

$$\phi_{2S}^k(\boldsymbol{y}) = \hat{b}_k = \mathrm{sgn}\,[y_k - \sum_{j \neq k} H_{kj} \phi_D^j(\boldsymbol{y})]$$

where $\phi_D^j(\boldsymbol{y})$ is the decision of the linear decorrelating detector for user j.

C. Asymptotic Effective Energy

Definition: Let $\mathcal{E}_k(\phi)$ denote the event (in the probability space in which $\boldsymbol{b}$ and $\boldsymbol{n}$ are defined) that the detector ϕ detects user i erroneously. Let the effective energy $e_k(\sigma, \phi)$ corresponding to the probability of $\mathcal{E}_k(\phi)$ be defined implicitly via the equation

$$\Pr(\mathcal{E}_k(\phi)) = Q(\sqrt{e_k(\sigma, \phi)}/\sigma).$$

The *asymptotic effective energy* (AEE) for user i, which we denote as $E_k(\phi)$, is defined as

$$E_k(\phi) \triangleq \lim_{\sigma \to 0} e_k(\sigma, \phi) = \sup \left\{ e \geq 0;\ \lim_{\sigma \to 0} \frac{\Pr(\mathcal{E}_k(\phi))}{Q(\sqrt{e}/\sigma)} < \infty \right\}. \qquad (3)$$

$E_k(\phi)$ is thus the energy required by the matched-filter detector in a single-user channel to achieve in high-SNR regimes, the bit-error rate of the multiuser detector for user k in the multiuser channel. It is equal to the product of the actual energy (w_k) and asymptotic efficiency as defined in [20] (which we denote as $\eta_k(\phi)$), hence $E_k(\phi) = w_k \eta_k(\phi)$. The exponentially complex ML detector is optimal in asymptotic efficiency (and hence in AEE) for each user among all detectors. The AEE for user k is equal to

$$E_k(\phi_{\mathrm{ML}}) = \min_{\boldsymbol{e} \in \{-1, 0, 1\}^K;\, e_k \neq 0} \boldsymbol{e}^T \boldsymbol{H} \boldsymbol{e}. \qquad (4)$$

The AEE for user k of an arbitrary linear detector corresponding to $\boldsymbol{F}$, denoted as $\phi(\boldsymbol{F})$, depends on $\boldsymbol{F}$ only through its kth row $\boldsymbol{f}_k^T$ and is given as

$$E_k(\phi(\boldsymbol{F})) = \frac{1}{\boldsymbol{f}_k^T \boldsymbol{H} \boldsymbol{f}_k} \max{}^2 \left\{ 0,\ \boldsymbol{f}_k^T \boldsymbol{h}_k - \sum_{j \neq k} |\boldsymbol{f}_k^T \boldsymbol{h}_j| \right\} \qquad (5)$$

where $\boldsymbol{h}_j$ denotes the jth column of $\boldsymbol{H}$.

III. Symmetric Energy

When the function of a multiuser detector is to detect all (or a subset of) users, it is reasonable to require that its error-rate performance be below a certain threshold for all such users simultaneously. We therefore base our measure on the joint error rate (JER) which is the probability that at least one user is detected erroneously. Such a measure accounts for all users' performance at once, representing an important departure from the traditional per-user performance measures such as signal-to-interference ratio (cf. [47]), asymptotic efficiency, and near–far resistance [20].

Definition: Let $\mathcal{E}(\phi)$ denote the event that the detector ϕ does not detect all users correctly. Let the effective energy $e(\sigma, \phi)$ corresponding to the probability of $\mathcal{E}(\phi)$ be defined implicitly via the equation

$$\Pr(\mathcal{E}(\phi)) = Q(\sqrt{e(\sigma, \phi)}/\sigma).$$

The *symmetric energy* is defined as the limit of the effective energy $e(\sigma, \phi)$ as $\sigma \to 0$ so that

$$E(\phi) \triangleq \lim_{\sigma \to 0} e(\sigma, \phi) = \sup \left\{ e \geq 0;\ \lim_{\sigma \to 0} \frac{\Pr(\mathcal{E}(\phi))}{Q(\sqrt{e}/\sigma)} < \infty \right\}. \qquad (6)$$

The symmetric energy $E(\phi)$ is thus the energy required by the matched-filter detector in a single-user channel to achieve in high-SNR regimes, the joint error rate of the multiuser detector ϕ operating in the multiuser channel.

Since the maximum-likelihood detector in (2) minimizes $\Pr(\mathcal{E}(\phi))$ over all detectors, it is optimal in symmetric energy when there are no constraints on structure or complexity. Therefore, the symmetric energy $E(\phi_{\mathrm{ML}})$ is the benchmark against which all other detectors' symmetric energies must be compared.

Definition: The relative symmetric energy (RSE) of a multiuser detector is the ratio of its symmetric energy to that of the maximum-likelihood detector expressed in decibels, i.e., $R(\phi, \phi_{\mathrm{ML}}) = 10 \log_{10}(E(\phi)/E(\phi_{\mathrm{ML}}))$.

The RSE of a multiuser detector is hence a nonpositive number that quantifies the SNR gap (in decibels) between its JER and the minimum achievable JER for high SNR. One could also use any other detector as a benchmark, in which case we have $R(\phi_1, \phi_2) = 10 \log_{10}(E(\phi_1)/E(\phi_2))$, which denotes the SNR gain (loss if negative) of detector ϕ_1 over ϕ_2.

Lemma 1: The symmetric energy of any multiuser detector for the CWMA channel is equal to its minimum asymptotic effective energy (MAEE), i.e.,

$$E(\phi) = \min_{k=1,2,\cdots,K} E_k(\phi). \tag{7}$$

Proof: Note that $\mathcal{E}(\phi) = \cup_{i=1}^{K} \mathcal{E}_i(\phi)$. Since $\mathcal{E}_k(\phi) \subseteq \mathcal{E}(\phi)$, $\Pr(\mathcal{E}_k(\phi))$ is a lower bound on $\Pr(\mathcal{E}(\phi))$. We can hence obtain an upper bound on $E(\phi)$ by substituting the single-event lower bound in place of $\Pr(\mathcal{E}(\phi))$ in (6). Using (3) then, we have $E(\phi) \le E_k(\phi)$. Moreover, since this bound is valid for any $k = 1, 2, \cdots, K$, by choosing the best among the K upper bounds, we have

$$E(\phi) \le \min_{k=1,2,\cdots,K} E_k(\phi). \tag{8}$$

Since the sum of single-event probabilities $\sum_{i=1}^{K} \Pr(\mathcal{E}_i(\phi))$ is an upper (union) bound on $\Pr(\mathcal{E}(\phi))$, we can obtain a lower bound on $E(\phi)$ by substituting the union bound in place of $\Pr(\mathcal{E}(\phi))$ in (6). In the computation of the limit of the ratio of this union bound to $Q(\sqrt{e}/\sigma)$, the term in the union bound that corresponds to (asymptotically) the slowest rate of exponential (in SNR) decay will dominate the union bound so that it alone determines the lower bound. Consequently, the lower bound is

$$E(\phi) \ge \min_{k=1,2,\cdots,K} E_k(\phi). \tag{9}$$

The upper and lower bounds in (8) and (9) coincide. □

The symmetric energy measure was introduced by the author in the context of power control for multiuser detection [44] where optimum power allocations were obtained that maximize symmetric energy under average power constraints for each of several multiuser detectors for Gaussian and fading channels. The dual concept of minimum average energy was defined in [46] as the minimum average energy (or equivalently, minimum sum of transmit powers of all users) required by a multiuser detector to achieve a given symmetric energy performance. The problem of optimizing symmetric energy among decision feedback detectors without the option of power control was solved (among other results) in a conference version of this paper [45].

As a simple consequence of Lemma 1, the symmetric energy of detectors for which AEE's are known, are easily obtained. For instance, using the formula for the AEE for the maximum-likelihood detector in (4), we have

$$E(\phi_{\mathrm{ML}}) = \min_{\boldsymbol{e} \in \{-1,0,1\}^K - \{\boldsymbol{0}\}} \boldsymbol{e}^T \boldsymbol{H} \boldsymbol{e}. \tag{10}$$

Similarly, the symmetric energy of any linear detector can be obtained from (5) by taking the minimum AEE of the K users. The linear optimal detector was obtained in [20] by optimizing $\eta_k(\phi(\boldsymbol{F}))$ over all vectors $\boldsymbol{f}_k$ for each user and hence this detector also optimizes $E_k(\phi(\boldsymbol{F}))$ (see (5)) for every user separately. It follows then that the linear optimal detector also optimizes symmetric energy among all linear detectors. An interesting property of the linear optimum detector is that for a sufficiently strong user, it achieves the AEE of the maximum-likelihood detector. In particular, it was shown in [20] that the kth user achieves optimum AEE if

$$\sqrt{w_k} \ge \max_{j=1,\cdots,K} \left(\frac{1}{|R_{kj}|} \sum_{i \ne k} \sqrt{w_i} |R_{ij}| \right). \tag{11}$$

Unfortunately, if one user is sufficiently strong to satisfy the sufficient condition to achieve optimum AEE, the other $K-1$ users will not. Moreover, for sufficiently weak users, the linear optimum detector coincides with the decorrelator [20]. The suboptimality of the decorrelator, coupled with the fact that it is for a weak user that the linear optimum detector coincides with the decorrelator, results in the symmetric energy of the linear optimum detector often being equal to that of the decorrelator. The symmetric energy of the decorrelator, which is also equal to that of the MMSE detector (the MMSE detector converges to the decorrelator in the limit as $\sigma \to 0$), is obtained by letting $\boldsymbol{F} = \boldsymbol{Q}$ in (5) and taking the minimum over $k = 1, \cdots K$ so that

$$E(\phi_{\mathrm{MMSE}}) = E(\phi_D) = \min_{k=1,2,\cdots,K} [Q_{kk}]^{-1}. \tag{12}$$

A. The Two-User Channel

A simple two-user study of the symmetric energy of the linear, two-stage, and the optimum detectors gives clear insights into their relative performances with and without power control and also serves to motivate decision feedback detection. We specify $E_k(\phi)$ for a two-user channel for the aforementioned detectors in terms of the energies w_1, w_2, and the normalized correlation $R_{12} = \rho$. We denote the other user's index as j so that all formulae apply for the cases $(k, j) \in \{(1, 2), (2, 1)\}$. The AEE's of the ML, conventional, decorrelating, and linear optimum detector for user k are denoted as $E_k(\phi_C)$, $E_k(\phi_D)$, $E_k(\phi_{\mathrm{LO}})$, and $E_k(\phi_{\mathrm{ML}})$, respectively. For $(k, j) \in \{(1, 2), (2, 1)\}$

$$E_k(\phi_{\mathrm{ML}}) = \min\{w_k, w_k + w_j - 2|\rho|\sqrt{w_k w_j}\} \tag{13}$$

$$E_k(\phi_C) = \max{}^2\{0, \sqrt{w_k} - \sqrt{w_j}|\rho|\} \tag{14}$$

$$E_k(\phi_D) = w_k(1-\rho^2) \tag{15}$$

$$E_k(\phi_{\mathrm{LO}}) = \begin{cases} E_k(\phi_{\mathrm{ML}}), & \text{if } \sqrt{\frac{w_j}{w_k}} < |\rho| \\ E_k(\phi_D), & \text{otherwise.} \end{cases} \tag{16}$$

Using (16), it can be verified that the symmetric energy of the linear optimum detector is equal to that of the decorrelating detector, i.e., $E(\phi_{\mathrm{LO}}) = E(\phi_D) = (1-\rho^2)\min\{w_1, w_2\}$.

The BER of the two-stage detector was derived for the general K-user problem in [37]. It was shown there that the key to the exact analysis of the BER is the statistical independence of

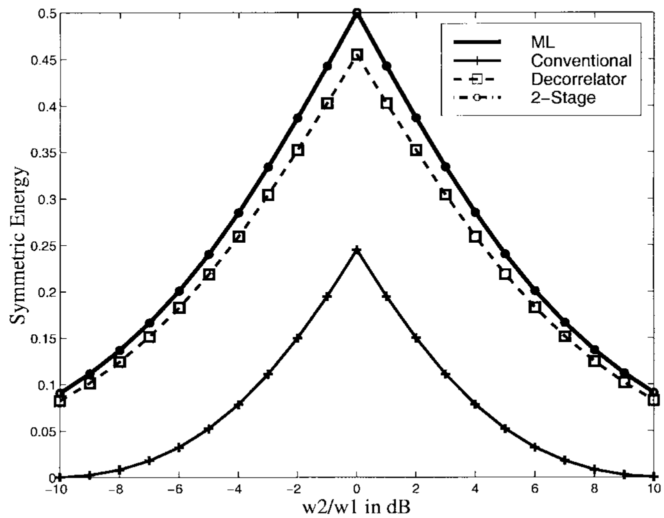

Fig. 1. A symmetric energy comparison between the linear, two-stage, and optimum detectors for a two-user channel with $\rho = 0.3$ and $w_1 + w_2 = 1$. The conventional detector performs poorly even when the correlation is low and the powers are balanced.

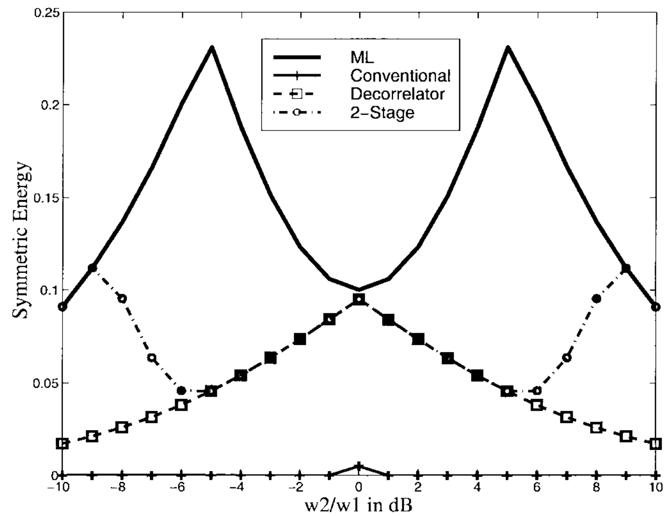

Fig. 2. A symmetric energy comparison of linear, two-stage, and optimum detectors for a two-user channel with $|\rho| = 0.9$ and $w_1 + w_2 = 1$. There is a yawning gap between the linear optimum and optimum detectors. Optimal power control for high-performance detectors can in general require very disparate energies.

the residual interference (resulting from imperfect interference cancellation) and the additive Gaussian noise components of the decision statistic

$$D_k = y_k - \sum_{j \neq k} H_{kj} \phi_D^j(\boldsymbol{y}).$$

When the result there is specialized to the two-user case, with the normalized correlation $R_{12} = \rho$, the bit-error probability for the kth user is

$$P_k(\phi_{2S}) = Q\left(\frac{\sqrt{w_k}}{\sigma}\right)\left[1 - Q\left(\frac{\sqrt{(1-\rho^2)w_j}}{\sigma}\right)\right] + Q\left(\frac{\sqrt{(1-\rho^2)w_j}}{\sigma}\right) Q\left(\frac{\left(\sqrt{w_k} - 2|\rho|\sqrt{w_j}\right)}{\sigma}\right). \quad (17)$$

The asymptotic effective energy is easily obtained as

$$E_k(\phi_{2S}) = \begin{cases} \min\{w_k,\, (1+3\rho^2)w_j - 4|\rho|\sqrt{w_j w_k} + w_k\}, & \text{if } 0 \leq \sqrt{\frac{w_j}{w_k}} \leq \frac{1}{2|\rho|} \\ \min\{w_k,\, (1-\rho^2)w_j\}, & \text{if } \sqrt{\frac{w_j}{w_k}} > \frac{1}{2|\rho|}. \end{cases} \quad (18)$$

The symmetric energy $E(\phi_{2S}) = \min\{E_1(\phi_{2S}),\, E_2(\phi_{2S})\}$.

Fig. 1 shows the symmetric energies of the conventional, decorrelating (and hence linear optimum), two-stage, and optimum detectors for a low correlation of $\rho = 0.3$ as a function of the energy ratio w_2/w_1 in decibels. We fix $w_1 + w_2 = 1$. The multiuser detectors show a marked improvement over the conventional detector over the entire range. Moreover, the two-stage detector uniformly outperforms the best linear detector and, in fact, achieves optimum symmetric energy for all energy ratios w_2/w_1. This property of the two-stage detector can be shown to remain true for all $\rho \leq 1/3$.

In Fig. 2, we consider a high correlation example with $\rho = 0.9$ and again we fix $w_1 + w_2 = 1$. The conventional detector becomes interference-limited for almost all energy ratios. There is a large gap between the decorrelator and the ML detector in spite of the fact that the decorrelator (or optimal linear and MMSE) is optimum in symmetric energy among all linear detectors. Therefore, restriction to linear detectors is too severe. The nonlinear two-stage detector also degrades significantly relative to the optimum detector for this high correlation example. However, it does at least as well as the decorrelating detector, improving upon the decorrelator's performance significantly for unequal energies or near–far situations. The decorrelator performs best relative to the ML detector when user energies are equal and degrades in near–far conditions. This may seem surprising at first given that the decorrelator achieves the near–far resistance (worst asymptotic efficiency over energies of interfering users) of the maximum-likelihood detector [20]. The problem is that near–far resistance is a per-user measure and a conservative one at that. Even in the rare instance where the signal energies have nearly worst case distribution for a particular user, so that the optimum detector for that user can perform not much better than the decorrelator, the other users can achieve significantly better performance with the optimum detector than with the decorrelator for the same signal energy distribution. In other words, *the signal energies cannot be worst case for all users simultaneously*, and designing the detector for this unrealizable scenario is too conservative.

Let us consider the implications of Figs. 1 and 2 when power control is an option with multiuser detection. With $w_1 + w_2 = 1$, we seek to distribute powers between the two users so as to achieve maximum symmetric energy or minimum JER in the high-SNR regions. We see for $\rho = 0.3$ that the power control strategy for each of the three detectors is one that makes the powers equal. For $\rho = 0.9$, however, the optimal power control for the optimum detector requires very disparate power levels (the two-stage detector favors equal or very disparate powers). Elaborate power control methods are usually implemented in CDMA systems to balance powers but an equipower distribution is essentially *ad hoc* and can be far from optimal as seen from the performance of the optimum detector in Fig. 2. *In general, optimum power distributions depend critically, and sometimes even in a complicated way, on the structure of the correlations between signals and the detector used*. Such distributions were obtained in [44] for several detectors that maximized symmetric energy under total (sum) power constraint for general K-user Gaussian as well as fading channels.

Let us consider the dual problem. In practice, it is required that each user's error-rate performance be below a certain threshold P. This can be assured by requiring a quality-of-service (QoS) constraint that the symmetric energy be above a certain threshold E which is determined by solving the equation $P = Q(\sqrt{E}/\sigma)$. Suppose that we require $E = 0.5$ in the two-user example of Fig. 1. This can be achieved only by the optimum and two-stage detectors with optimum power allocation that balances powers to set $w_1 = w_2 = 0.5$. Similarly, for $\rho = 0.9$ in Fig. 2, a QoS specification of $E = 0.225$ for $w_1 + w_2 = 1$ is delivered only by the optimum detector and that too with power control that tightly maintains a power imbalance of nearly 5 dB between the powers of the two users. As the QoS requirement is lowered, however, note that power control for the optimum detector need not be very precise. For instance, with $E = 0.1$, all power ratios between $-$ 9 and 9 dB will work. At this level, the two-stage detector will also meet the specification but only with tight power control which maintains a 9-dB power imbalance. *There is thus a tradeoff between the quality of a multiuser detector (as measured by its symmetric energy) and the accuracy of power control needed*. The higher the quality of the multiuser detector, the laxer are the demands placed on accurate power control. This motivates the search for multiuser detectors that achieve high symmetric energy performance.

Fig. 3 corresponds to a two-user channel with $\rho = 0.9$ and depicts boundaries for the sum of powers $w_1 + w_2$ for each detector above which that detector achieves a symmetric energy that is at least as large as a fixed threshold (which we choose here to be equal to 0.2). It is seen that a sum of powers that is greater than 1 is sufficient for the optimum detector with power control that tightly maintains a power imbalance of nearly 5 dB. A greater expenditure of total power (say 1.5) allows a wider range of power distributions to achieve the QoS specification, thereby easing the requirement on the accuracy of power control. *For a given detector, there is thus a tradeoff between the total power expended and accuracy of power control*. The higher the total power expended, the laxer is the requirement on accuracy of power control to achieve a given QoS specification. *Similarly, there is a tradeoff between the quality of a multiuser detector (as measured by its symmetric energy) and the minimum total energy required to achieve a given QoS specification*. The higher the quality of the multiuser detector, the lower is the total power expended. This is another motivation for the search for multiuser detectors that achieve high symmetric energy performance.

Low power consumption is critical for battery-operated hand-held and lap-top wireless communication devices while maintaining a given QoS performance specification for every

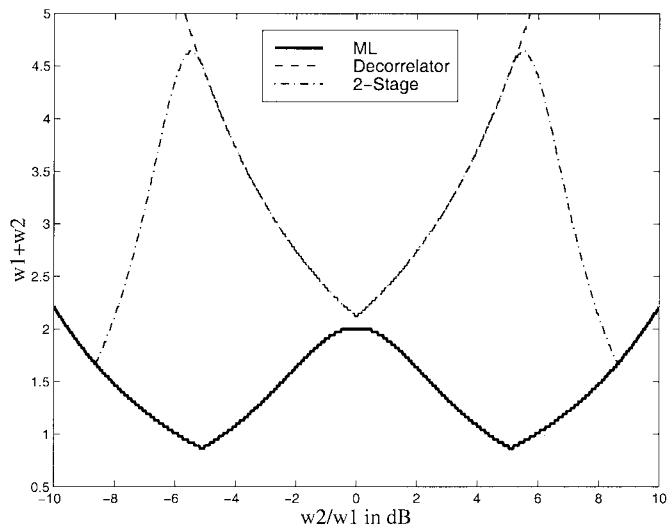

Fig. 3. A comparison of the boundaries of the total energy above which the optimum, two-stage, and decorrelating detectors will achieve a QoS-type symmetric energy criterion (that is, $E = 0.2$) for a two-user channel with $|\rho| = 0.9$.

user in the system. It is, therefore, of interest to obtain optimal power control strategies that minimize power consumption (say, by minimizing the sum of transmit powers) while delivering a symmetric energy that is above a fixed threshold E. Such problems were solved for linear and some decision feedback detectors in [46].

IV. Decision Feedback Multiuser Detection

The motivation for considering decision feedback detection is three-fold. The optimum detector is exponentially complex, the linear detectors perform poorly relative to the optimum detector, and decision feedback detectors have a much greater potential while lending themselves to analysis and optimization.

A. Description and Examples

We give a description of decision feedback detectors for PAM signaling. A decision feedback detector (DFD) detects users 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 parameterized by feedforward and feedback matrices $\boldsymbol{F}$ and $\boldsymbol{B}$ with $\boldsymbol{B}$ being strictly lower triangular. A detailed structure of a DFD is shown in Fig. 4. The decision statistic for the first user is $D_1 = \boldsymbol{f}_1^T \boldsymbol{y}$ which is quantized by the PAM slicer to obtain $\hat{b}_1$. The decision statistic of the second user is formed by computing the inner product $\boldsymbol{f}_2^T \boldsymbol{y}$ and then subtracting from it $B_{21}\hat{b}_1$ resulting in the decision statistic $D_2 = \boldsymbol{f}_2^T \boldsymbol{y} - B_{21}\hat{b}_1$ which is then quantized by a PAM slicer to obtain $\hat{b}_2$, and so on. In forming the decision statistic for the kth user, the inner product $\boldsymbol{f}_k^T \boldsymbol{y}$ is formed after which the interference from the already detected "past" users (i.e., users $1, \cdots, k-1$) is mitigated by subtracting from that inner product a linear combination of the past user's symbols $\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_{k-1}$. The coefficients of the linear combination are the $k-1$ nonzero elements of the kth row of $\boldsymbol{B}$. The

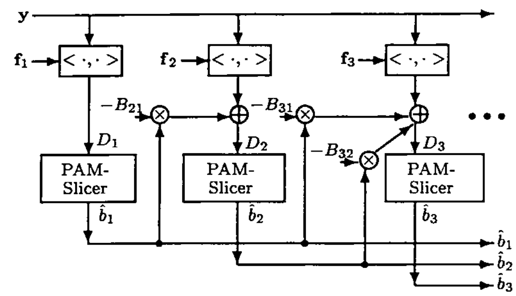

Fig. 4. Decision feedback multiuser detection.

resulting decision statistic

$$D_k = \boldsymbol{f}_k^T \boldsymbol{y} - \sum_{j=1}^{k-1} B_{kj} \hat{b}_j$$

goes through a minimum Euclidean PAM slicer to produce the kth user's decision $\hat{b}_k$ according to

$$\hat{b}_k = \arg \min_{x_k \in \mathcal{F}} |D_k - \boldsymbol{f}_k^T \boldsymbol{h}_k x_k|^2 \tag{19}$$

where $\mathcal{F}$ is the PAM alphabet. The function of the feedforward vectors $\boldsymbol{f}_k$ is to minimize the effects of multiple-access interference from the as yet undecoded ("future") users $k+1$ to K, and the residual interference after subtraction from the already decoded ("past") users 1 to $k-1$. The function of the feedback coefficients $\{B_{kj}\}_{j=1}^{k-1}$ is to mitigate the interference contributed by the past users. Assuming without loss of generality that $\boldsymbol{f}_k^T \boldsymbol{h}_k > 0$ for each k, the PAM slicer in (19) simplifies in the binary antipodal case to a sign detector

$$\phi(\boldsymbol{F}, \boldsymbol{B})\colon \hat{b}_k = \mathrm{sgn}\left[\sum_{j=1}^{K} F_{kj} y_j - \sum_{j=1}^{k-1} B_{kj} \hat{b}_j\right]. \tag{20}$$

It is easily verified that the particular choice of the feedback transformation $\boldsymbol{B} = \mathcal{L}(\boldsymbol{FH})$, where $\mathcal{L}(\cdot)$ is the strictly lower triangular part of the matrix argument, is such that the interference due to past users would be completely canceled if past-user decisions were perfect. Given an arbitrary $\boldsymbol{F}$ and a strictly lower triangular $\boldsymbol{B}$, the implementation of (20) requires, on average, $3K/2$ additions and multiplications per user.

Linear detectors are degenerate decision feedback detectors with $\boldsymbol{B} = \boldsymbol{0}$. For instance, ϕ_C, ϕ_D, and ϕ_{MMSE}, are the decision feedback detectors corresponding to the $(\boldsymbol{F}, \boldsymbol{B})$ pairs $(\boldsymbol{I}, \boldsymbol{0})$, $(\boldsymbol{H}^{-1}, \boldsymbol{0})$, and $((\boldsymbol{H} + \sigma^2 \boldsymbol{I})^{-1}, \boldsymbol{0})$, respectively.

A simple example of a nonlinear DFD is the conventional DFD (C-DFD) denoted as $\phi_{\text{C-DF}}$. It corresponds to the $(\boldsymbol{F}, \boldsymbol{B})$ pair $(\boldsymbol{I}, \mathcal{L}(\boldsymbol{H}))$. This detector captures the idea of successive cancellation that has appeared in several papers (cf. [4], [24], [52], and [61]). There is no attempt at mitigating interference via the feedforward matrix but the feedback matrix is such that the past-user interference would be completely canceled when their decisions are perfect. The nondecision feedback counterpart of $\phi_{\text{C-DF}}$ is the conventional detector.

The decorrelating DFD (D-DFD) is defined by the $(\boldsymbol{F}, \boldsymbol{B})$ pair $(\boldsymbol{L}^{-T}, \mathcal{L}(\boldsymbol{L}))$ where $\boldsymbol{H} = \boldsymbol{L}^T \boldsymbol{L}$ is a Cholesky factorization of $\boldsymbol{H}$ with $\boldsymbol{L}$ being lower triangular. The feedforward transformation whitens the noise and the feedback matrix is of the form $\mathcal{L}(\boldsymbol{FH})$. The D-DFD was proposed in [7]. We will denote it as $\phi_{\text{D-DF}}$.

B. The D-DFD

We show that the nondecision feedback counterpart of the D-DFD is the decorrelator. Consider the $(\boldsymbol{F}, \boldsymbol{B})$ matrix pair $(\boldsymbol{S}, \mathcal{L}(\boldsymbol{SH}))$, where $\boldsymbol{S}$ is an upper triangular matrix defined as follows: the nonzero elements of the kth row of $\boldsymbol{S}$ are identically equal to the first row of the decorrelator defined for a user-expurgated channel consisting of only users indexed from k to K (i.e., the first row of the inverse of the "south-east" block of dimension $K-k+1 \times K-k+1$ of a 2×2 block partition of the matrix $\boldsymbol{H}$).[2]

Lemma 2: The decision feedback detector $(\boldsymbol{S}, \mathcal{L}(\boldsymbol{SH}))$ is equivalent to the D-DFD $(\boldsymbol{L}^{-T}, \mathcal{L}(\boldsymbol{L}))$.

Proof: Let $\boldsymbol{SH} = \boldsymbol{G}$. It is easy to see that $\boldsymbol{G}$ is a lower triangular matrix with unit diagonal elements. Using the Cholesky decomposition of $\boldsymbol{H}$ we have $\boldsymbol{SL}^T = \boldsymbol{GL}^{-1}$. Both $\boldsymbol{S}$ and $\boldsymbol{L}^T$ are upper triangular and so is their product. Similarly, both $\boldsymbol{G}$ and $\boldsymbol{L}^{-1}$ are lower triangular and so is their product. But since the products are equal, they must both be equal to a diagonal matrix. This diagonal matrix is denoted as Λ. Therefore, $\boldsymbol{G} = \Lambda \boldsymbol{L}$ but since $\mathrm{diag}\,(\boldsymbol{G}) = \boldsymbol{I}$, we have that $\Lambda = (\mathrm{diag}\,(\boldsymbol{L}))^{-1}$. Now, $\boldsymbol{S} = \boldsymbol{GH}^{-1}$, whence it follows that $\boldsymbol{S} = (\mathrm{diag}\,(\boldsymbol{L}))^{-1} \boldsymbol{L}^{-T}$. It only remains to invoke the equivalence stated in Footnote 2. □

The advantage of the $(\boldsymbol{S}, \mathcal{L}(\boldsymbol{SH}))$ specification is that it is more intuitive. The multiple-access interference is nonlinearly canceled from users with lower indices based on hard decisions on those users' symbols and it is canceled from users with higher indices by the process of decorrelation under the assumption that the nonlinear interference cancellation from lower indexed users was perfect. The D-DFD can therefore be thought of as a decision feedback counterpart of the decorrelator.

C. BER and AEE Analysis of the D-DFD

One of the main difficulties in the characterization of performance of decision feedback detectors is the error propagation problem. Easy-to-interpret upper and lower bounds on the performance of decision feedback group detectors were obtained in [43] which we state here for the particular case of the D-DFD

$$\min_{i=1,2,\cdots,k} L_{ii}^2 \le E_k(\phi_{\text{D-DF}}) \le \min_{\boldsymbol{e} \in \{-1,0,1\}^k,\, e_k \neq 0} \boldsymbol{e}^T \boldsymbol{Ce} \le L_{kk}^2 \tag{21}$$

where the $k \times k$ matrix $\boldsymbol{C}$ is the Schur complement $\boldsymbol{C} \triangleq \boldsymbol{H}_{11} - \boldsymbol{H}_{12} \boldsymbol{H}_{22}^{-1} \boldsymbol{H}_{21}$ in a 2×2 block-partition of $\boldsymbol{H}$ with $\boldsymbol{H}_{11}$ being

[2] The sequential group detector obtained in [43] when all group sizes are equal to one, reduces to a DFD corresponding to the pair $(\boldsymbol{P}, \mathcal{L}(\boldsymbol{PH}))$ where $\boldsymbol{P} = \mathrm{diag}^2(\boldsymbol{L})\boldsymbol{S}$. Note, however, that the DFD $(\boldsymbol{F}, \boldsymbol{B})$ is equivalent to $(\boldsymbol{DF}, \boldsymbol{DB})$ for any diagonal and positive-definite matrix $\boldsymbol{D}$. Hence, the $(\boldsymbol{S}, \mathcal{L}(\boldsymbol{SH}))$ DFD is equivalent to the sequential group detector of [43]. The equivalence of this detector to the D-DFD proved in Lemma 2 was claimed in [43] without proof.

the "north–west" block of dimension $k \times k$ with the other matrices being of compatible dimensions. The weaker of the two upper bounds L_{kk}^2 is the AEE of the D-DFD under the assumption that past user decisions are perfect (cf. (25)).

Interesting properties of the AEE can be derived from the bounds in (21). In particular, it is easily verified that the symmetric energy of the D-DFD can be both upper- and lower-bounded by $\min_{k=1,2,\cdots,K} L_{kk}^2$, so that

$$E(\phi_{\text{D-DF}}) = \min_{k=1,2,\cdots,K} L_{kk}^2. \tag{22}$$

More generally, we also have by the same argument, a closed-form expression for the minimum of the first k users' AEE's (and hence a more detailed information about the performance of the D-DFD than just its symmetric energy) as

$$\min_{i=1,2,\cdots,k} E_i(\phi_{\text{D-DF}}) = \min_{i=1,2,\cdots,k} L_{ii}^2. \tag{23}$$

The techniques in [43] are, however, not applicable for non-decorrelating DFD's. In the next section, we will use another approach to obtain generalizations of above results for arbitrary DFD's. Before that, we present an exact BER and AEE analysis of the D-DFD.

The bit-error probability of the D-DFD for user k, denoted as $\Pr(\mathcal{E}_k(\phi_{\text{D-DF}}))$, can be obtained as follows. Let $\boldsymbol{b}_k = [b_1, \cdots, b_k]^T$ and $\hat{\boldsymbol{b}}_k = [\hat{b}_1, \cdots, \hat{b}_k]^T$ be the vectors of bits and bit decisions of the D-DFD up to user k, respectively. Define the error vector $\boldsymbol{e}_k = 2^{-1}(\boldsymbol{b}_k - \hat{\boldsymbol{b}}_k)$ and the set of admissible error vectors corresponding to $\boldsymbol{b}_k$ as

$$\mathcal{E}_k(\boldsymbol{b}_k) = \{\boldsymbol{e}_k \subset \{-1, 0, 1\}^k;\ e_k = b_k \\ \&\, e_m \in \{b_m, 0\} m = 1, 2, \cdots, k-1\}.$$

Then, using the law of total probability

$$\Pr(\mathcal{E}_k(\phi_{\text{D-DF}})) = 2^{-k} \sum_{\boldsymbol{b}_k \in \{-1,1\}^k} \sum_{\boldsymbol{e}_k \in \mathcal{E}_k(\boldsymbol{b}_k)} \cdot \Pr(\hat{b}_k \neq b_k | \hat{\boldsymbol{b}}_{k-1} = \boldsymbol{b}_{k-1} - 2\boldsymbol{e}_{k-1}) \cdot \Pr(\hat{\boldsymbol{b}}_{k-1} = \boldsymbol{b}_{k-1} - 2\boldsymbol{e}_{k-1}). \tag{24}$$

The first term in the product of two probabilities can be easily obtained as follows: let the decision statistic

$$\sum_{j=1}^{K} F_{kj} y_j - \sum_{j=1}^{k-1} B_{kj} \hat{b}_j \triangleq D_k.$$

For the D-DFD, substituting $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{b} + \boldsymbol{n}$, we have $D_k = L_{kk} b_k + \mathcal{I}_k + \hat{n}_k$, where $\mathcal{I}_k$ is the residual interference

$$\mathcal{I}_k \triangleq 2 \sum_{j=1}^{k-1} L_{kj} e_j$$

and $\hat{n}_k$ is the kth element of the white-noise vector $\hat{\boldsymbol{n}} = \boldsymbol{L}^{-T} \boldsymbol{n}$ with covariance $\sigma^2 \boldsymbol{I}$. Since the decisions $\hat{\boldsymbol{b}}_{k-1}$ are functions of only the noise components $\hat{n}_1, \cdots, \hat{n}_{k-1}$, which in turn are statistically independent of $\hat{n}_k$, the residual interference $\mathcal{I}_k$ due to those decisions is statistically independent of $\hat{n}_k$ (this argument was first used in the analysis of the two-stage detector with decorrelating first-stage in [37]) so that

$$\Pr(\hat{b}_k \neq b_k | \hat{\boldsymbol{b}}_{k-1} = \boldsymbol{b}_{k-1} - 2\boldsymbol{e}_{k-1}) = Q\left(\frac{L_{kk} + b_k \mathcal{I}_k}{\sigma}\right). \tag{25}$$

Unlike the analysis of the two-stage detector in [37], however, the second term in the product in (24) can be obtained without relying on multivariate Gaussian distribution functions by using the chain rule for probabilities

$$\Pr(\hat{\boldsymbol{b}}_{k-1} = \boldsymbol{b}_{k-1} - 2\boldsymbol{e}_{k-1}) = \prod_{m=1}^{k-1} \Pr(\hat{b}_m = b_m - 2e_m | \hat{\boldsymbol{b}}_{m-1} = \boldsymbol{b}_{m-1} - 2\boldsymbol{e}_{m-1}). \tag{26}$$

Each term in the product can be obtained by noting that $\hat{b}_m = \operatorname{sgn}(D_m)$ where $D_m = L_{mm} b_m + \mathcal{I}_m + w_m$ with

$$\mathcal{I}_m \triangleq 2 \sum_{j=1}^{m-1} L_{mj} e_j.$$

Again due to the fact that $\hat{\boldsymbol{n}}$ has independent components, the decisions $\hat{\boldsymbol{b}}_{m-1}$, and hence the residual interference $\mathcal{I}_m$, are statistically independent of the additive noise term $\hat{n}_m$. Therefore, it is easily shown that

$$\Pr(\hat{b}_m = b_m - 2e_m | \hat{\boldsymbol{b}}_{m-1} = \boldsymbol{b}_{m-1} - 2\boldsymbol{e}_{m-1}) = Q\left(\frac{(-1)^{|e_m|+1}(L_{mm} + b_m \mathcal{I}_m)}{\sigma}\right). \tag{27}$$

Substituting (25) and (27) into (24) yields the bit error probability of the D-DFD as

$$\Pr(\mathcal{E}_k(\phi_{\text{D-DF}})) = 2^{-k} \sum_{\boldsymbol{b}_k \in \{-1,1\}^k} \sum_{\boldsymbol{e}_k \in \mathcal{E}_k(\boldsymbol{b}_k)} \prod_{m=1}^{k} \cdot Q\left(\frac{(-1)^{|e_m|+1}(L_{mm} + b_m \mathcal{I}_m)}{\sigma}\right). \tag{28}$$

The asymptotic effective energy corresponding to each product of Q-function terms in the double summation in (28) is

$$\sum_{m=1}^{k} \max{}^2\{0, (-1)^{|e_m|+1}(L_{mm} + b_m \mathcal{I}_m)\}.$$

Since the term corresponding to the slowest exponential rate of decay dominates the error probability as $\sigma \to 0$, we have that the asymptotic effective energy of user k is

$$E_k(\phi_{\text{D-DF}}) = \min_{\boldsymbol{b}_k \in \{-1,1\}^k} \min_{\boldsymbol{e}_k \in \mathcal{E}_k(\boldsymbol{b}_k)} \sum_{m=1}^{k} \cdot \max{}^2\{0, (-1)^{|e_m|+1}(L_{mm} + b_m \mathcal{I}_m)\}. \tag{29}$$

Equations (28) and (29) give exact formulas for the BER and AEE of the D-DFD. Note, however, that the properties of the AEE's of the D-DFD obtained in [43], i.e., the bounds of (21) and the equalities in (22) and (23) may be tedious to derive by using of the exact formula for AEE in (29).

D. On the Choice of Order of Detection for the D-DFD

There has been considerable interest in linear and decision feedback multiuser detection in recent years. Conventional wisdom has it that decision feedback detection is a good idea only if the energies of the various users are sufficiently disparate, and that, otherwise, linear detection is better because it avoids the complication of error propagation. We will show that a careful analysis reveals otherwise.

Let $\boldsymbol{P}$ be a $K \times K$ permutation matrix where each row and column is a unit vector. The CWMA channel model $\boldsymbol{y} = \boldsymbol{Hb} + \boldsymbol{n}$ can be permuted as $\boldsymbol{y}^P = \boldsymbol{H}^P \boldsymbol{b}^P + \boldsymbol{n}^P$ where $\boldsymbol{y}^P = \boldsymbol{Py}$ and $\boldsymbol{H}^P = \boldsymbol{PHP}^T$ and $\boldsymbol{b}^P = \boldsymbol{Pb}$ and $\boldsymbol{n}^P = \boldsymbol{Pn}$. The D-DFD for the permuted model $\boldsymbol{y}^P = \boldsymbol{H}^P \boldsymbol{b}^P + \boldsymbol{n}^P$ is defined as the D-DFD for the permutation $\boldsymbol{P}$. In other words, for the Cholesky factorization $\boldsymbol{H}^P = \boldsymbol{L}^P (\boldsymbol{L}^P)^T$, the feedforward–feedback matrix pairs $((\boldsymbol{L}^P)^{-T}, \mathcal{L}(\boldsymbol{L}^P))$ when applied to the permuted model $\boldsymbol{y}^P = \boldsymbol{H}^P \boldsymbol{b}^P + \boldsymbol{n}^P$ is the D-DFD for the permutation $\boldsymbol{P}$. We will denote it as $\phi^P_{\text{D-DF}}$. If the kth row of $\boldsymbol{P}$ is a unit row vector with a one in the i_kth position, then users are detected in the order $i_1, i_2, \cdots, i_K$. There are $K!$ permutations and hence $K!$ D-DFD's. Let us denote the set of $K!$ permutation matrices as $\mathcal{P}$.

There does not appear to be a simple relationship between the D-DFD's for two distinct permutations. Their BER and AEE performances can also be quite different. However, the analysis of the previous section can be used to find the BER and AEE of $\phi^P_{\text{D-DF}}$ for any given $\boldsymbol{P}$ by working with the permuted model $\boldsymbol{y}^P = \boldsymbol{H}^P \boldsymbol{b}^P + \boldsymbol{n}^P$. Naturally, it is of interest to know what effect the choice of permutation has on performance. We first consider symmetric energy which gives an indication of joint error rate in high SNR.

Proposition 1: The D-DFD for any permutation $\boldsymbol{P} \in \mathcal{P}$ outperforms the decorrelator in symmetric energy, i.e.,

$$E(\phi^P_{\text{D-DF}}) \geq E(\phi_D), \qquad \text{for all } \boldsymbol{P} \in \mathcal{P}. \tag{30}$$

$\triangle$

Proof: The symmetric energy of $E(\phi^P_{\text{D-DF}})$ is equal to $\min_{m=1,2,\cdots,K} (L^P_{mm})^2$ (see (22)), where L^P_{mm} is the mth diagonal element of $\boldsymbol{L}^P$. Note from (25) that $(L^P_{mm})^2$ is the AEE of user m relative to permutation $\boldsymbol{P}$ (which is user i_m relative to the identity permutation) under the assumption that interference cancellation from past users $m-1, \cdots, 1$ (relative to $\boldsymbol{P}$) is perfect. According to the $(\boldsymbol{S}, \mathcal{L}(\boldsymbol{SH}))$ interpretation of the D-DFD, $(L^P_{mm})^2$ is thus the AEE of a decorrelator in a fictitious user-expurgated channel consisting of users $m, m+1, \cdots, K$ (relative to $\boldsymbol{P}$) which must therefore be no less than the AEE of the decorrelator for user m (relative to $\boldsymbol{P}$) in the actual channel where all K users are active. The latter is equal to $[Q^P_{mm}]^{-1}$, where $\boldsymbol{Q}^P = (\boldsymbol{H}^P)^{-1}$. Thus $(L^P_{mm})^2 \geq [Q^P_{mm}]^{-1}$. Therefore, we have a lower bound on the symmetric energy given as

$$E(\phi^P_{\text{D-DF}}) = \min_{m=1,2,\cdots,K} (L^P_{mm})^2 \tag{31}$$

$$\geq \min_{k=1,2,\cdots,K} [Q^P_{kk}]^{-1} \tag{32}$$

$$= \min_{k=1,2,\cdots,K} [Q_{kk}]^{-1} = E(\phi_D) \tag{33}$$

where the last equality follows from the fact that $Q^P_{mm} = Q_{i_m i_m}$ (with $\boldsymbol{P}$ such that its mth row is the i_mth unit vector), which in turn is a simple consequence of $\boldsymbol{Q} = (\boldsymbol{H}^P)^{-1} = (\boldsymbol{PHP}^T)^{-1} = \boldsymbol{PQP}^T$. $\square$

Thus the D-DFD outperforms the decorrelator no matter what the distribution of energies may be, and independently of the order in which the users are detected. This implies that the JER of the D-DFD will always be no higher than that of the decorrelator for sufficiently high SNR. A result of independent interest in algebra follows from (30).

Lemma 3: For any positive-definite matrix $\boldsymbol{H}$ and any permutation matrix $\boldsymbol{P}$ with the Cholesky factorization $\boldsymbol{PHP}^T = \hat{\boldsymbol{L}}^T \hat{\boldsymbol{L}}$, and $\boldsymbol{Q} = \boldsymbol{H}^{-1}$, we have

$$\min_{i=1,2,\cdots,K} (\hat{L}_{ii})^2 \geq \min_{i=1,2,\cdots,K} [Q_{ii}]^{-1}, \qquad \forall \boldsymbol{P} \in \mathcal{P}. \tag{34}$$

The reader is invited to prove this lemma independently of the terminology of multiuser detection. The inequality $\hat{L}^2_{mm} \geq [\hat{Q}_{mm}]^{-1}$, where $\hat{\boldsymbol{Q}} = \boldsymbol{PH}^{-1}\boldsymbol{P}^T$ can be proved algebraically.

With the order of detection regarded as a *design parameter*, one suspects that there may be stronger results. In light of the result on the lower bound on the AEE in (21), the following question is appropriate. Given a positive-definite matrix $\boldsymbol{H}$, does there exist a permutation matrix $\boldsymbol{P}$ such that the matrix $\boldsymbol{H}^P = \boldsymbol{PHP}^T$ with Cholesky factorization $(\boldsymbol{L}^P)^T \boldsymbol{L}^P$ with diagonal elements of $\boldsymbol{L}^P$ in nonincreasing order? If it does, that permutation would correspond to a highly desirable order of detection for the D-DFD since, for that order, the D-DFD would achieve genie-aided AEE performance for every user, and would therefore be unaffected by error propagation in high-SNR scenarios (see Theorem 3 in the next section). Unfortunately, not even the abundance of permutations guarantees the existence of one such permutation for any given positive-definite matrix. The simple two-user case suffices to produce a counterexample. Neither of the two permutations of the two users will yield the desired property; for instance, when the energy ratio w_1/w_2 lies in the interval $(1 - R^2_{12}, (1 - R^2_{12})^{-1})$.

Consequently, we reformulate our problem. Given that it is more desirable to be able to say something about performance for each user rather than just worst user performance, it is natural to inquire about the existence of an order of detection for which the D-DFD may outperform the decorrelator for *every* user (note that the requirement posed in the previous paragraph would also ensure this user-wise dominance of the D-DFD but it is not necessary). Moreover, if such an order exists, can it be quickly determined since an exhaustive search based on the AEE performance evaluation for $\phi^P_{\text{D-DF}}$ requires that it be done for all $K!$ permutations.

Theorem 1: Order users as follows: select the first user of the new order (denote this user's index as i_1) as one that has the highest AEE among all users if each one of them were to be detected by a decorrelator. For $k = 2, \cdots, K$ select the kth user of the new order (denote this user's index as i_k) as the user that has the highest AEE among the remaining $K-k+1$ users when each of them is detected by a decorrelator for the user-expurgated channel consisting of just those remaining users (i.e., those indexed by $\{1, \cdots, K\} - \{i_1, \cdots, i_{k-1}\}$). When users are ordered as $\{i_1, \cdots, i_K\}$, the D-DFD outperforms the decorrelator in AEE for every user.

Proof: The AEE of the i_kth user when detected by the D-DFD in the order given by the ordering algorithm is greater than or equal to the minimum of the AEE of the previous users i_j, for $j = 1, \cdots, k$, with the i_jth user's AEE evaluated under the assumption of perfect cancellation of interference from users $i_1, \cdots, i_{j-1}$. In what follows, we show that in turn, each of these quantities is greater than or equal to the AEE of the decorrelator for the i_kth user, thereby proving the sought result. Now, the i_1th user, when detected by the D-DFD for the order given by the ordering algorithm, has an AEE that is equal to that of the decorrelator for that user which, by virtue of our ordering algorithm, is at least as high as that of the decorrelator for the i_kth user. Next, the i_jth user for $1 < j < k$ has an AEE under the assumption of perfect cancellation of interference from users $i_1, \cdots, i_{j-1}$, that is greater than or equal to the AEE of the i_kth user for the D-DFD that detects users in a new order that results from exchanging users i_j and i_k in the order dictated by our ordering algorithm, and under the same perfect cancellation assumption, as a consequence of the definition of our ordering algorithm. The latter quantity, which is equal to the effective energy of the decorrelator for the i_kth user in a user-expurgated channel consisting of users indexed by $\{i_j, i_{j+1}, \cdots, i_K\}$ is, as a consequence, greater than or equal to the AEE of the decorrelator for the i_kth user in the full K user channel. Finally, following the last argument, the AEE of the i_kth user detected in the order given by the ordering algorithm under the assumption of perfect cancellation of interference from users $i_1, \cdots, i_{k-1}$ is greater than or equal to that of the decorrelator for the i_kth user. Hence the result. □

The guarantee of the above theorem is not available for the arrangement of users according to the nonincreasing order of energies. It is easy to find a counterexample. It must also be mentioned here that the user-wise dominance of the D-DFD over the decorrelator is the most conservative claim that one can make about the particular ordering rule of the above theorem. As a practical matter, in many instances, this rule tends to also give large differences between the AEE's for each user particularly as the (permuted) user indices increase. We thus have a constructive proof of the following purely algebraic result as implied by Theorem 1.

Lemma 4: Given a positive-definite matrix $\boldsymbol{H}$ of dimension K, there exists a permutation $\boldsymbol{P}$ such that with $\hat{\boldsymbol{H}} = \boldsymbol{P}\boldsymbol{H}\boldsymbol{P}^T$, $\hat{\boldsymbol{Q}} = (\hat{\boldsymbol{H}})^{-1}$ and with the Cholesky factorization $\hat{\boldsymbol{H}} = \hat{\boldsymbol{L}}^T\hat{\boldsymbol{L}}$

$$\min_{i=1,2,\cdots,k} \hat{L}_{ii}^2 \geq [\hat{Q}_{kk}]^{-1}, \qquad k = 1, 2, \cdots, K \tag{35}$$

where $\hat{L}_{ii}$ and $\hat{Q}_{ii}$ are the ith diagonal elements of $\hat{\boldsymbol{L}}$ and $\hat{\boldsymbol{Q}}$, respectively.

E. Performance Analysis of Arbitrary DFD's

When the feedforward matrix is not noise-whitening, the technique for finding the BER and AEE of the D-DFD does not hold. The conditional error probabilities in (25) and (27) are no longer computable because the conditioning on past decisions alters the distribution of the additive noise in the current decision statistic (from prior to posterior) in a way that is difficult to characterize analytically (see [36]). A different approach is necessary.

The following theorem gives a closed-form expression for the symmetric energy of an *arbitrary* decision feedback detector. More generally, it gives an exact expression for the minimum of the AEE's of the first k users.

Theorem 2: For an arbitrary decision feedback detector, the symmetric energy is equal to that of the same detector which is assisted by a genie that ensures that past user decisions are always correct (referred to henceforth as the genie-aided detector). In particular

$$E(\boldsymbol{F}, \boldsymbol{B}) = \min_{1 \leq k \leq K} E_k^g(\boldsymbol{F}, \boldsymbol{B}) \tag{36}$$

where $E_k^g(\boldsymbol{F}, \boldsymbol{B})$ denotes the AEE for user k of the genie-aided version of the detector. More generally, the minimum of the AEE's of the first k users is equal to the minimum of the genie-aided version of that detector's AEE's for those k users, i.e.,

$$\min_{i=1,2,\cdots,k} E_i(\boldsymbol{F}, \boldsymbol{B}) = \min_{i=1,2,\cdots,k} E_i^g(\boldsymbol{F}, \boldsymbol{B}), \quad k = 1, 2, \cdots, K. \tag{37}$$

Proof: Let C_i and C_i^g denote the events (in the probability space over which $\boldsymbol{b}$ and $\boldsymbol{n}$ are defined) that the ith user's decision is correct for an arbitrary decision feedback detector $(\boldsymbol{F}, \boldsymbol{B})$ and its genie-aided version, respectively. The simple but remarkable set equality $\cap_{i=1}^k C_i = \cap_{i=1}^k C_i^g$ can be deduced for any choice of $(\boldsymbol{F}, \boldsymbol{B})$. Consequently

$$\Pr\left(\bigcup_{i=1}^{k} \mathcal{E}_i\right) = \Pr\left(\bigcup_{i=1}^{k} \mathcal{E}_i^g\right) \tag{38}$$

where $\mathcal{E}_i^g$ is the error event that the ith user's decision is incorrect for the genie-aided version of the DFD (i.e., the complement of C_i^g). Equation (38) states that the probabilities that not all of the first k users' decisions are correct are equal for a decision feedback detector and its genie-aided version. Using the argument in the proof of Lemma 1, it can be shown that the AEE corresponding to $\Pr(\cup_{i=1}^k \mathcal{E}_i)$ is equal to $\min_{i=1,2,\dots,k} E_i(\boldsymbol{F}, \boldsymbol{B})$, and similarly, the AEE corresponding to $\Pr(\cup_{i=1}^k \mathcal{E}_i^g)$ is equal to $\min_{i=1,2,\dots,k} E_i^g(\boldsymbol{F}, \boldsymbol{B})$. Hence (37) is true which, for $k = K$, gives the equality of symmetric energies in (36). □

The above result motivates the definition of a performance measure that is an asymptotic indicator of the joint error rate that at least one of a subset of users is detected erroneously. In the context of decision feedback detection, the subset that consists of the first k users is of particular interest. The asymptotic effective energy corresponding to this joint error rate defined as in (6) (with $\cup_{i=1}^{k}\mathcal{E}_i(\phi)$ in place of $\cup_{i=1}^{K}\mathcal{E}_i(\phi)$) will be equal to the smallest AEE from among the group of users $\{1, 2, \cdots, k\}$.

Definition: The *group-symmetric energy* (GSE) denoted as $\Psi_k(\phi)$ for any detector for a given k is defined as the minimum of the first k AEE's, i.e.,

$$\Psi_k(\phi) = \min_{i=1,2,\cdots,k} E_i(\phi). \tag{39}$$

Note that $\Psi_K(\phi) = E(\phi)$ is the symmetric energy. As in the case of AEE and symmetric energy, $\Psi_k(\boldsymbol{F}, \boldsymbol{B})$ denotes the GSE of the DFD corresponding to $(\boldsymbol{F}, \boldsymbol{B})$. Moreover, for specific detectors, we denote it as $\Psi_k(\phi_D)$, $\Psi_k(\phi_{\text{D-DF}})$, etc.

The computation of $\Psi_k(\boldsymbol{F}, \boldsymbol{B})$ (and hence symmetric energy) for any DFD becomes easy as a consequence of (37) because there are no error propagation issues to deal with in the analysis of a genie-aided DFD. Moreover, the GSE $\Psi_k(\boldsymbol{F}, \boldsymbol{B})$ serves as a very useful lower bound on the AEE of an arbitrary DFD for user k.

Corollary 1: For a general DFD, we have

$$\begin{aligned} E_k(\boldsymbol{F}, \boldsymbol{B}) &\geq \Psi_k(\boldsymbol{F}, \boldsymbol{B}) \\ &= \min_{i=1,2,\cdots,k} \frac{1}{\boldsymbol{f}_i^T\boldsymbol{H}\boldsymbol{f}_i} \\ &\quad \cdot \max{}^2\left\{0,\, \boldsymbol{f}_i^T\boldsymbol{h}_i - \sum_{j=1}^{i-1}|\boldsymbol{f}_i^T\boldsymbol{h}_j - B_{ij}| - \sum_{j=i+1}^{K}|\boldsymbol{f}_i^T\boldsymbol{h}_j|\right\} \end{aligned} \tag{40}$$

where the ith term in the minimization over i is equal to $E_i^g(\boldsymbol{F}, \boldsymbol{B})$.

When the result of Corollary 1 is applied to the D-DFD with $(\boldsymbol{F} = \boldsymbol{L}^{-T}, \boldsymbol{B} = \mathcal{L}(\boldsymbol{L}))$, we have the expression for symmetric energy in (22) and the lower bound on AEE of (21), and the expression for GSE in (23), all of which are based on results obtained by using a different analysis technique in [43]. Similarly, for the C-DFD with $(\boldsymbol{F} = \boldsymbol{I}, \boldsymbol{B} = \mathcal{L}(\boldsymbol{H}))$, we have

$$\begin{aligned} E_k(\phi_{\text{C-DF}}) &\geq \Psi_k(\phi_{\text{C-DF}}) \\ &= \min_{i=1,2,\cdots,k} \max{}^2\left\{0,\, \sqrt{w_i} - \frac{1}{\sqrt{w_i}}\sum_{j=i+1}^{K}|H_{ij}|\right\}. \end{aligned} \tag{41}$$

It is intuitive that if the users are detected by a D-DFD in the decreasing order of their energies, and their signal strengths are sufficiently disparate, then error propagation effects should vanish. The following theorem quantifies this statement and gives sufficient conditions under which a general DFD will achieve genie-aided performance.

Theorem 3: Any DFD $(\boldsymbol{F}, \mathcal{L}(\boldsymbol{F}\boldsymbol{H}))$ achieves the AEE performance of its genie-aided version for every user if the genie-aided AEE's are in nonincreasing order. i.e.,

$$E_1^g(\boldsymbol{F}, \boldsymbol{B}) \geq E_2^g(\boldsymbol{F}, \boldsymbol{B}) \geq \cdots \geq E_K^g(\boldsymbol{F}, \boldsymbol{B}).$$

Proof: Consider (37) of Theorem 2. Clearly,

$$E_1(\boldsymbol{F}, \boldsymbol{B}) = E_1^g(\boldsymbol{F}, \boldsymbol{B}).$$

Furthermore, if $E_1^g(\boldsymbol{F}, \boldsymbol{B}) \geq E_2^g(\boldsymbol{F}, \boldsymbol{B})$, then it can be shown using the equality in (37) for $k = 2$ that

$$E_2(\boldsymbol{F}, \boldsymbol{B}) = E_2^g(\boldsymbol{F}, \boldsymbol{B}).$$

Now suppose that $E_i(\boldsymbol{F}, \boldsymbol{B}) = E_i^g(\boldsymbol{F}, \boldsymbol{B})$ for $1 \leq i \leq k-1$. If the genie-aided AEE's are in nonincreasing order, we can use (37) to show that $E_k(\boldsymbol{F}, \boldsymbol{B}) = E_k^g(\boldsymbol{F}, \boldsymbol{B})$. The result of the theorem follows by induction. □

When the above theorem is applied to the D-DFD, the following corollary results.

Corollary 2: The D-DFD will achieve its genie-aided upper bounds on AEE's for every user if the energy ratios satisfy the inequalities

$$w_1\tilde{L}_{11}^2 \geq w_2\tilde{L}_{22}^2 \geq \cdots, \geq w_K\tilde{L}_{KK}^2 \tag{42}$$

where $\tilde{L}_{ii}$ is the ith diagonal element of $\tilde{\boldsymbol{L}}$ which, in turn, is the Cholesky decomposition of the normalized correlation matrix $\boldsymbol{R}$ (note that $\tilde{\boldsymbol{L}} = \boldsymbol{L}\boldsymbol{W}^{-1/2}$).

The conditions in (42) ensure that the genie-aided AEE's are in nonincreasing order by requiring that the user energies be sufficiently disparate. Note that it is difficult to verify Corollary 2 directly from the exact AEE formulas for the D-DFD in (29).

Example 1: Consider a three-user channel with a normalized Toeplitz correlation matrix $\boldsymbol{R}$ with $R_{kl} = \rho^{|k-l|}$. The reader can verify that if $w_1 \geq w_2 \geq w_3(1-\rho^2)^{-1}$, the D-DFD achieves the genie-aided AEE bounds $w_1(1-\rho^2)$, $w_2(1-\rho^2)$, and w_3 for users 1, 2, and 3, respectively.

F. Optimum Decision Feedback Detection (O-DFD)

In this section, we consider the optimization of symmetric energy among all DFD's. We seek the optimum feedforward and feedback matrix pair $(\boldsymbol{F}_*, \boldsymbol{B}_*)$

$$(\boldsymbol{F}_*, \boldsymbol{B}_*) \in \arg\max_{\boldsymbol{F},\boldsymbol{B}} E(\boldsymbol{F}, \boldsymbol{B}). \tag{43}$$

Note that the above optimization includes all linear detectors since they are degenerate cases of decision feedback detectors. We will denote the O-DFD as $\phi_{\text{O-DF}}$.

Theorem 4: The O-DFD $(\boldsymbol{F}_*, \boldsymbol{B}_*)$ defined in (43) is such that $\boldsymbol{F}_*$ is upper triangular with the nonzero part of the ith row determined by coefficients of the linear optimal detector (that optimizes AEE) for the ith user derived for a user-expurgated channel consisting of users indexed from i to K. The nonzero part of the strictly lower triangular feedback matrix $\boldsymbol{B}_*$ is identically equal to $\mathcal{L}(\boldsymbol{F}_*\boldsymbol{H})$.

Proof: The symmetric energy of an arbitrary DFD from (40) is given as

$$\begin{aligned} E(\boldsymbol{F}, \boldsymbol{B}) = &\min_{i=1,2,\cdots,K} \frac{1}{\boldsymbol{f}_i^T\boldsymbol{H}\boldsymbol{f}_i} \\ &\cdot \max{}^2\left\{0,\, \boldsymbol{f}_i^T\boldsymbol{h}_i - \sum_{j=1}^{i-1}|\boldsymbol{f}_i^T\boldsymbol{h}_j - B_{ij}| - \sum_{j=i+1}^{K}|\boldsymbol{f}_i^T\boldsymbol{h}_j|\right\}. \end{aligned} \tag{44}$$

Note that the ith term depends on the feedback matrix only through the ith row of $\boldsymbol{B}$. Moreover, it is maximized by setting $B_{ij} = \boldsymbol{f}_i^T \boldsymbol{h}_j$ for $j = 1, 2, \cdots, i-1$. Therefore, for a given $\boldsymbol{F}$, the optimum feedback matrix is $\mathcal{L}(\boldsymbol{FH})$. In other words, the intuitive rule that feedback must be such that interference from past users is completely canceled when bit decisions are correct, is in fact optimal. For an arbitrary DFD, with this optimum choice of $\boldsymbol{B}$, we have

$$E(\boldsymbol{F}, \mathcal{L}(\boldsymbol{FH})) = \min_{i=1,2,\cdots,K} \frac{1}{\boldsymbol{f}_i^T \boldsymbol{H} \boldsymbol{f}_i} \max{}^2 \left\{ 0,\, \boldsymbol{f}_i^T \boldsymbol{h}_i - \sum_{j=i+1}^{K} |\boldsymbol{f}_i^T \boldsymbol{h}_j| \right\}. \tag{45}$$

Note that the ith term depends on $\boldsymbol{F}$ only through its ith row so that the maximization over $\boldsymbol{F}$ can be performed for each term separately. Considering the maximization over $\boldsymbol{f}_i$ we have from (45)

$$\max_{\boldsymbol{f}_i} \max{}^2 \left\{ 0,\, \frac{\boldsymbol{q}^T \boldsymbol{l}_i - \sum_{j=i+1}^{K} |\boldsymbol{q}^T \boldsymbol{l}_j|}{\sqrt{\boldsymbol{q}^T \boldsymbol{q}}} \right\} \tag{46}$$

where the vector $\boldsymbol{q}$ is defined as $\boldsymbol{L}\boldsymbol{f}_i$, and where $\boldsymbol{l}_j$ is the jth column of $\boldsymbol{L}$. Considering restrictions of the vector $\boldsymbol{f}_i$ that yield a strictly positive value for the effective energy (without loss of generality), it is clear that in optimizing the objective function in (46), and owing to the lower triangularity of $\boldsymbol{L}$, the vector $\boldsymbol{q}$ must always be chosen such that its first $i-1$ elements are identically equal to zero. Using the lower triangularity of $\boldsymbol{L}$ again, this implies that the first $i-1$ elements of $\boldsymbol{f}_i$ are identically equal to zero. This means that the optimum $\boldsymbol{F}$ can, without loss of generality, be chosen to be upper triangular.

Define the "south–east" square block of size $K-i+1$ of a 2×2 partition of the matrix $\boldsymbol{H}$ as $\boldsymbol{H}^{(i)}$. Let the columns of $\boldsymbol{H}^{(i)}$ be indexed from i to K with the jth column denoted as $\tilde{\boldsymbol{h}}_j$. With $\boldsymbol{F}$ taken to be upper triangular (and the feedback matrix equal to $\mathcal{L}(\boldsymbol{FH})$}, (45) becomes

$$E(\boldsymbol{F}, \mathcal{L}(\boldsymbol{FH})) = \min_{i=1,2,\cdots,K} \frac{1}{\tilde{\boldsymbol{f}}_i^T \boldsymbol{H}^{(i)} \tilde{\boldsymbol{f}}_i} \max{}^2 \left\{ 0,\, \tilde{\boldsymbol{f}}_i^T \tilde{\boldsymbol{h}}_i - \sum_{j=i+1}^{K} |\tilde{\boldsymbol{f}}_i^T \tilde{\boldsymbol{h}}_j| \right\} \tag{47}$$

where $\tilde{\boldsymbol{f}}_i$ is the nonzero part of $\boldsymbol{f}_i$ of dimension $K-i+1$ (lower subvector of a two-vector partition of $\boldsymbol{f}_i$). Now, note that the ith term in (47) depends only on $\tilde{\boldsymbol{f}}_i$ so that the optimization of the minimum of the K terms in (47) can be performed independently for each i. By comparing with (5), it is evident that the ith term is identical to the AEE of the linear detector $\tilde{\boldsymbol{f}}_i$ in a fictitious user-expurgated channel over which only users $i, i+1, \cdots, K$ are active and whose system correlation matrix is $\boldsymbol{H}^{(i)}$. The problem thus reduces to that of finding the linear detector in this fictitious channel that optimizes AEE (or equivalently, asymptotic efficiency, see [20] for a solution). Hence the result. □

The suboptimal detectors C-DFD, D-DFD, and MSIR-DFD share with the O-DFD the common property of having upper triangular feedforward matrices and the strictly lower triangular part of $\boldsymbol{FH}$ as their feedback matrices, respectively. The forgoing proof rigorously justifies the choice of an upper triangular feedforward matrix in the case of the D-DFD and the MSIR-DFD and the intuitive choices of the feedback transformations in all three cases. However, neither of these three detectors is in general optimal in symmetric energy.

Note that determining the feedforward transformation for the O-DFD requires even less computation than that for the linear optimum detector since in the O-DFD, the linear optimum detectors for each user have to be determined only for successively user-expurgated channels. While the worst case design complexity of the algorithm that finds the optimum linear detector and hence the O-DFD is exponential, the implementational complexity of both detectors is linear. Once the O-DFD is found, the detection of each user's symbol requires only K additions and K multiplications. In a time-varying environment, the optimum linear detector or the O-DFD must be recomputed as often as the channel model is updated.

There is a stronger sense in which the O-DFD is optimal. This is given in the next theorem. It rules out the possibility that, while optimal in minimum asymptotic effective energy, the O-DFD may perform poorly for certain users relative to other DFD's. The proof must use the more general result on the GSE in (37) rather than the symmetric energy in (36). The rest of the details are the same as in the proof leading up to Theorem 4.

Theorem 5: The O-DFD optimizes, among all decision feedback detectors, the group symmetric energies $\Psi_k(\boldsymbol{F}, \boldsymbol{B})$. In other words, for each $k = 1, 2, \cdots, K$

$$(\boldsymbol{F}_*, \boldsymbol{B}_*) \in \arg \max_{\boldsymbol{F}, \boldsymbol{B}} \min_{i=1,2,\cdots,k} E_i(\boldsymbol{F}, \boldsymbol{B}). \tag{48}$$

As a simple consequence, we can lower-bound the GSE (and hence the symmetric energy) of the O-DFD

$$\Psi_k(\phi_{\text{O-DF}}) \geq \min_{1 \leq i \leq k} L_{ii}^2. \tag{49}$$

G. Robustness of the O-DFD and C-DFD to the Choice of Order

It was shown in Section IV-D that the D-DFD outperforms the decorrelator in symmetric energy independently of the order in which users are detected. We seek a similar result for the O-DFD and C-DFD.

The O-DFD determined as described in Theorem 4 for the permuted model $\boldsymbol{y}^P = \boldsymbol{H}^P \boldsymbol{b}^P + \boldsymbol{n}^P$ (see Section IV-D) is defined as the O-DFD for the permutation $\boldsymbol{P}$. Similarly, the feedforward–feedback matrix pair $(\boldsymbol{I}, \mathcal{L}(\boldsymbol{H}^P))$ is defined as the C-DFD for the permutation $\boldsymbol{P}$. We will denote these detectors as $\phi_{\text{O-DF}}^P$ and $\phi_{\text{C-DF}}^P$, respectively. There are $K!$ O-DFD's and C-DFD's corresponding to each permutation in $\mathcal{P}$.

As in the case of the D-DFD, for two distinct orders, the O-DFD's are distinct. The performance of these O-DFD's can be vastly different as well. The same is true of the C-DFD.

In contrast, the decisions of linear detectors are independent of each other and hence the order of detection is irrelevant to the performance of each user.

The following corollary then illuminates the issue of the relative performances of linear and decision feedback detectors according to the symmetric energy measure.

Corollary 3: The O-DFD outperforms the D-DFD in symmetric energy when both use the same order of detection. It outperforms the linear optimum detector (hence the decorrelator) in symmetric energy *independently* of the order of detection. The C-DFD outperforms the conventional detector in symmetric energy *independently* of the order of detection, i.e.,

$$E(\phi^P_{\text{O-DF}}) \geq E(\phi^P_{\text{D-DF}}) = \min_{i=1,2,\cdots,K} (L^P_{ii})^2, \quad \text{for any } \boldsymbol{P} \in \mathcal{P} \tag{50}$$

$$E(\phi^P_{\text{O-DF}}) \geq E(\phi_{LO}) \geq E(\phi_D), \qquad \text{for all } \boldsymbol{P} \in \mathcal{P} \tag{51}$$

$$E(\phi^P_{\text{D-DF}}) \geq E(\phi_D), \qquad \text{for all } \boldsymbol{P} \in \mathcal{P} \tag{52}$$

$$E(\phi_{\text{C-DF}}) \geq E(\phi_C), \qquad \text{for all } \boldsymbol{P} \in \mathcal{P}. \tag{53}$$

Proof: The inequality in (50) and the first inequality in (51) are true since the O-DFD is optimum in symmetric energy. The second equality in (51) is true since ϕ_{LO} is optimal in symmetric energy among all linear detectors. The third inequality was proved in Section IV-D. The last inequality is left as an exercise for the reader. Note that (51)–(53) hold for all $\boldsymbol{P} \in \mathcal{P}$ since the symmetric energies of linear detectors are independent of $\boldsymbol{P}$. □

The above corollary attests to a robustness of decision feedback multiuser detection for high SNR to the choice of the order of detection. If we adopt the convention that $\Psi_k(\phi^P)$ denotes the GSE of the first k users in the permutation $\boldsymbol{P}$ (for a linear or decision feedback detector, this is the minimum of AEE's of users indexed $i_1, \cdots, i_k$ when the jth row of $\boldsymbol{P}$ is the i_jth unit row vector), then the above corollary can also be generalized to the GSE measure (i.e., (50)–(53) hold with symmetric energies $E(\phi)$ replaced by the corresponding GSE's $\Psi_k(\phi)$). In the case of linear detectors, unlike symmetric energy, the GSE can be different for two distinct permutations since the first k users of the two permutations need not be the same. The symmetric energies of the linear detectors in (51)–(53) must be replaced by $\Psi_k(\phi^P_{\text{LO}})$, $\Psi_k(\phi^P_D)$, and $\Psi_k(\phi^P_C)$, respectively, and instead of those inequalities reading "for all $\boldsymbol{P} \in \mathcal{P}$" they must read "for any $\boldsymbol{P} \in \mathcal{P}$." The proof uses Theorem 5 and follows as in the proof of Corollary 3.

H. The Two-User Channel

In this subsection, we give several examples that give valuable insights into a comparative performance analysis of nonlinear decision feedback detectors and their linear counterparts relative to the optimum detector and the role of power control for multiuser detection [44], [46].

Example 2: Consider the symmetric energies of the O-DFD, the D-DFD, and the C-DFD for a two-user channel.

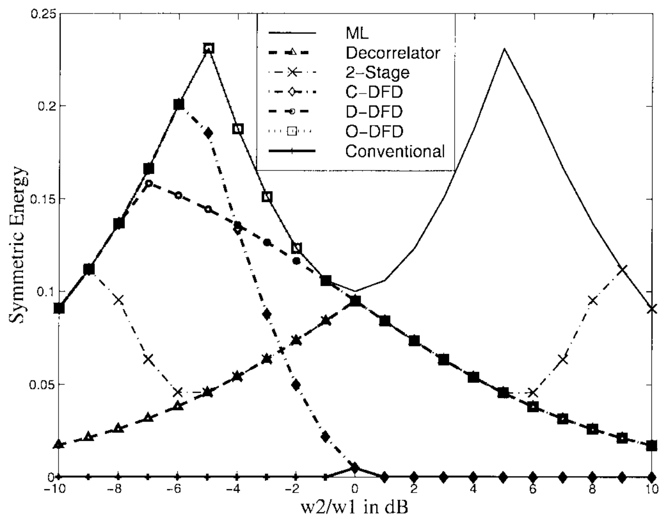

Fig. 5. A symmetric energy comparison of linear, two-stage, decision-feedback, and ML detectors for a two-user channel with $|\rho| = 0.9$ and $w_1 + w_2 = 1$. The first user is detected first for all energy ratios. Decision feedback detectors do at least as well as, and sometimes much better than, their linear counterparts, independently of the order of detection.

We let $\boldsymbol{P} = \boldsymbol{I}$, i.e., user 1 is detected first and user 2 last. As a simple application of Theorem 2, we have $E(\phi_{\text{C-DF}}) = \min\{E_1(\phi_C), w_2\}$, $E(\phi_{\text{D-DF}}) = \min\{E_1(\phi_D), w_2\}$, and $E(\phi_{\text{O-DF}}) = \min\{E_1(\phi_{LO}), w_2\}$ with the AEE's of the linear detectors given in (14)–(16).

Fig. 5 depicts the symmetric energies of the three decision feedback detectors for the two-user channel as a function of w_2/w_1 in decibels with $|\rho| = 0.9$ and $w_1 + w_2 = 1$. The symmetric energies of the conventional, decorrelating, two-stage, and optimum detectors from Fig. 2 are included for easy comparison. Note that there is no gap between the O-DFD and the optimum detector when $w_2 < w_1$. In this region, therefore, the O-DFD improves upon the decorrelator (or linear optimum) performance dramatically. Moreover, it uniformly outperforms the D-DFD (as it must) with the gap between the two being most significant at w_2/w_1 equal to -5 dB. When $w_2 > w_1$, however, the symmetric energies of the D-DFD and the O-DFD coincide with that of the decorrelator. This is attributable to misordering the users. However, they perform no worse than the decorrelator (which in turn performs best among linear detectors) for all choices of energy ratios, thereby verifying the robustness of DFD's in symmetric energy to the choice of order of detection. It is also noteworthy that while the C-DFD outperforms the conventional detector (which is not saying much), with perfectly balanced powers ($w_2/w_1 =$ 0 dB), it has a symmetric energy that is 20 times lower than that of the O-DFD (or the D-DFD or ML), becoming interference-limited (there is a JER floor) when $w_2 > w_1$. Note also that none of the DFD's uniformly outperform the two-stage detector.

Favorable as the result in Corollary 3 may be to decision feedback detection, it does not imply that the selection of order is unimportant. On the contrary, the extent of improvement in performance depends critically on the choice of this order. A poor ordering of the users may give a symmetric energy for the O-DFD, D-DFD, or C-DFD that is equal to that of their

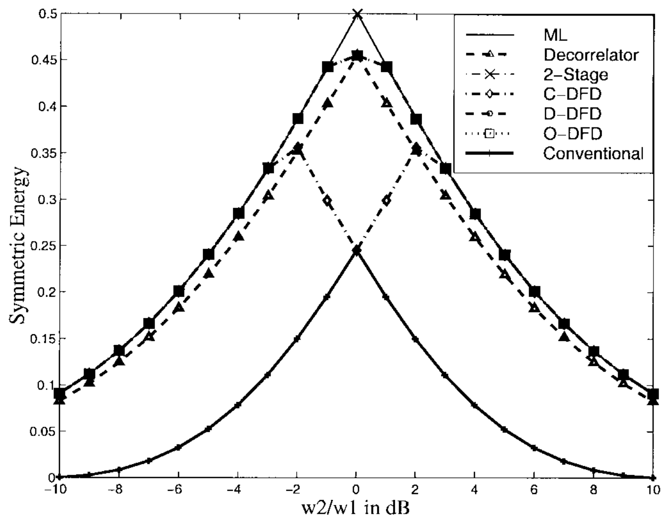

Fig. 6. Symmetric energy comparison of linear, two-stage, and optimum detectors for a two-user channel with $\rho = 0.3$ and $w_1 + w_2 = 1$. For the DFD's, the stronger user is detected first for all energy ratios. The C-DFD performs poorly when powers are balanced and the two-stage detector achieves optimum symmetric energy for all w_2/w_1.

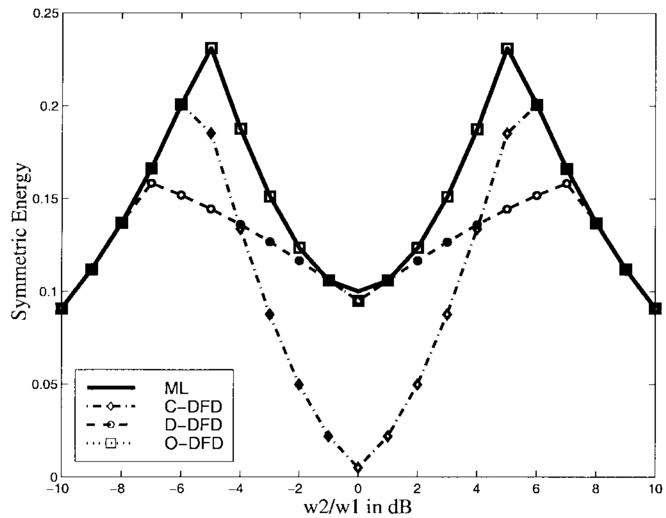

Fig. 7. A symmetric energy comparison of decision feedback detectors and the ML detector for a two-user channel with $|\rho| = 0.9$ and $w_1 + w_2 = 1$ when the stronger user is detected first. The O-DFD is virtually indistinguishable from the optimum detector. The C-DFD performs poorly when energies are not very dissimilar.

corresponding linear detectors as indicated in Fig. 5 for the case $w_2 > w_1$. Worse still, they may have a worse performance than their linear counterparts for some users.

Example 3: Consider the symmetric energy of the C-DFD, D-DFD, and the O-DFD for a two-user channel when each of these detectors detects the stronger user first and the weaker user last. When $w_1 \geq w_2$, the first user is detected first and the symmetric energies are given in Example 2. When $w_2 > w_1$, we have, similarly, $E(\phi_{\text{C-DF}}) = \min\{E_2(\phi_C), w_1\}$, $E(\phi_{\text{D-DF}}) = \min\{E_2(\phi_D), w_1\}$, and $E(\phi_{\text{O-DF}}) = \min\{E_2(\phi_{\text{LO}}), w_1\}$.

Fig. 6 depicts the symmetric energies of the nondecision feedback and decision feedback detectors for a low correlation of $\rho = 0.3$ and $w_1 + w_2 = 1$ and the order of detection determined according to decreasing energies. We noted in Fig. 1 that the gap between linear and optimum detectors is not very significant in this low-correlation example. We see in this figure that the C-DFD performs poorly even in this case, when the powers of the two users are not very dissimilar. On the other hand, O-DFD and the C-DFD have identical performances and coincide with the optimum detector for all but nearly equal powers. When the powers of the two users are balanced, the two-stage detector outperforms the O-DFD. This, however, does not violate the optimality of the O-DFD in symmetric energy among DFD's since the two-stage detector does not belong to the class of DFD's.

Fig. 7 depicts the symmetric energies of the DFD's for the two-user channel with $|\rho| = 0.9$ and $w_1 + w_2 = 1$ and the order of detection determined according to decreasing energies. Note the marked improvement in performance of all DFD's for the case $w_2 > w_1$ relative to that of the DFD's with a fixed order of detection shown in Fig. 5. Switching the order essentially symmetrizes the performance about 0 dB. This figure shows that a proper choice of order of detection is very important for decision feedback detectors. In particular, the performance of the O-DFD becomes virtually indistinguishable from that of the ML detector over the entire range of relative signal strengths and it uniformly dominates the performances of the C-DFD and the D-DFD as expected. The C-DFD performs poorly when the difference between the signal strengths are not very dissimilar making it an unsuitable candidate for a system without power control.

Let us consider the implications of Fig. 7 when power control is an option. To achieve maximum symmetric energy, all the three DFD's require distinctly different power imbalances so that power control algorithms must be designed for the particular detector used. If signature waveforms can be designed to make their correlations low to distinguish one user from another, as in spread-spectrum communications, then why not implement power control to further maximize the distinguishability between different users? Interestingly, with accurate power control, the C-DFD will outperform the D-DFD making it a good candidate for systems where such power control can be accomplished. Optimum power distributions that maximize symmetric energy under total (sum) power constraint were obtained in [44] for the D-DFD for the general K-user synchronous Gaussian channel and for a multiuser DFE for the asynchronous channel proposed therein without making the assumption of perfect feedback for the two-user case.

Fig. 8 corresponds to a two-user channel with $\rho = 0.9$ and depicts boundaries for the sum of powers $w_1 + w_2$ for each DFD above which that detector achieves a symmetric energy that is at least as large as a fixed threshold (which we chose here to be equal to 0.2). This figure must be compared with Fig. 3 which shows the same information for nondecision feedback detectors. Notice the significant lowering of the boundaries for the DFD's relative to their nondecision-feedback counterparts. This implies that in applications where power consumption comes at a premium, as in battery-operated devices (mobiles), decision feedback detectors such as the O-

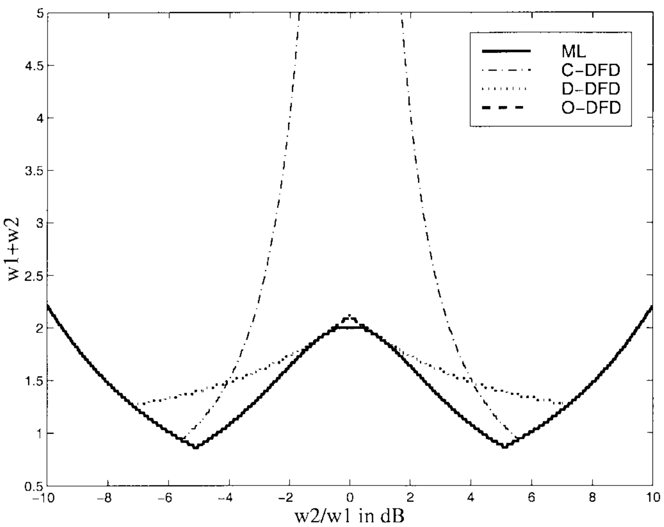

Fig. 8. A comparison of the boundaries of the total energy above which the optimum- and decision feedback detectors will achieve a QoS-type symmetric energy criterion (that is, $E = 0.2$) for a two-user channel with $|\rho| = 0.9$.

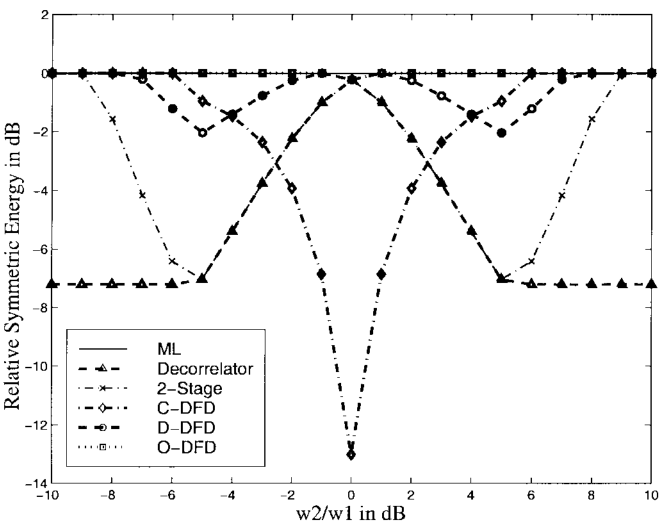

Fig. 9. RSE comparison of linear and decision feedback detectors with respect to the ML detector for a two-user channel with $|\rho| = 0.9$ and $w_1 + w_2$ held constant. Decision feedback detectors are far superior to linear detectors when the order of detection is chosen properly.

DFD and the D-DFD are the obvious choices at the base station with or without power control.

Finally, Fig. 9 depicts the relative symmetric energies (RSE) of the DFD's for the two-user channel with $|\rho| = 0.9$ and $w_1 + w_2$ held constant (the RSE's of all detectors are invariant to the choice of this constant) and the order of detection determined according to decreasing energies. For the sake of easy comparison, the RSE's of the decorrelator and the two-stage detector are also included. This figure therefore displays the information contained in Figs. 2 and 7 but in a form that makes it easy to assess the precise degree of suboptimality of the suboptimal detectors relative to the optimum detector. As expected, the performance of the O-DFD is virtually indistinguishable from that of the ML detector over the entire range of relative signal strengths and it uniformly dominates the performances of the C-DFD and the D-DFD. The O-DFD realizes, for example, an SNR gain of over 2 dB relative to the D-DFD when the second user is weaker or stronger than the first by 5 dB. The C-DFD performs poorly when the signal strengths are not very dissimilar with an SNR loss of as much as 13 dB relative to the optimum detector when powers are perfectly balanced. All nonlinear detectors (except the C-DFD) perform uniformly better than linear detectors, by as much as over 7 dB in near–far situations.

I. O-DFD Can User-Wise Outperform the Linear Optimum Detector

The BER performance analysis of the two-stage detector in [37] showed that feeding back a decorrelating (or conventional) first-stage decision from a weak user can be detrimental. Motivated by this, the simple and intuitive order of decreasing signal strengths was proposed in [7]. The benefit of this was clearly demonstrated in the symmetric energy analysis for the two-user channel in Figs. 7 and 9. In the general K-user channel, however, it is not just the energies but also the correlations that determine good ordering rules. Such rules were proposed for decision feedback group detection (and hence for the D-DFD) in [43]. In the case of the D-DFD, we showed the existence and fast computation of an order that ensured that the D-DFD dominates the decorrelator in AEE for all users (see Theorem 1). We have a similar result for the O-DFD and the C-DFD as well. Instead of the lower bound

$$E_k(\phi_{\text{D-DF}}) \geq \min_{i=1,2,\cdots,k} L_{ii}^2$$

the proof uses the more general version of the lower bound on the AEE performance of Corollary 1, i.e.,

$$E_k(\boldsymbol{F}, \boldsymbol{B}) \geq \Psi_k(\boldsymbol{F}, \boldsymbol{B}) = \min_{i=1,2,\cdots,k} E_i^g(\boldsymbol{F}, \boldsymbol{B})$$

that is applicable for arbitrary DFD's.

Theorem 6: Order users for the O-DFD (C-DFD) as follows: select the first user of the new order (denote this users index as i_1) as one that has the highest AEE among all users if each one of them were to be detected by the linear optimum (conventional) detector. For $k = 2, \cdots, K$ select the kth user of the new order (denote this user's index as i_k) as the user that has the highest AEE among the remaining $K - k + 1$ users, when each of them is detected by the linear optimum (conventional) detector for the user-expurgated channel consisting of just those remaining users (i.e., those indexed by $\{1, \cdots, K\} - \{i_1, \cdots, i_{k-1}\}$). When users are ordered as $\{i_1, \cdots, i_K\}$, the O-DFD (C-DFD) outperforms the linear optimum (conventional) detector in AEE for every user.

Interesting as the above theorem is for the O-DFD from an existence standpoint, the problem with the determination of the permutation of users therein is in the complexity of computation of the AEE of the linear optimum detector for every user in a successively user-expurgated channel. Fortunately, as a quick alternative, one could order users according to the ordering algorithm described for the D-DFD, and use that order for the O-DFD. As one might expect, the

O-DFD will still outperform the decorrelator for every user. The proof is left to the reader.

Corollary 4: The O-DFD outperforms the decorrelator in asymptotic effective energy for every user provided the users are arranged according to the order defined in the ordering algorithm for the D-DFD in Theorem 1.

The guarantees of this section are not available for the ordering of users according to the nonincreasing order of energies. It is easy to find a counterexample.

J. AEE of the O-DFD for the Two-User Channel

Symmetric energy is the minimum AEE and is important when all users are to be detected as in a centralized receiver. However, in applications where only some of the users' performance matters (as in hand-held or laptop communication devices), it is important to know the performance of a particular user which can be obtained in terms of the AEE measure.

In this subsection, we give per-user performances in terms of the AEE's of the D-DFD and the O-DFD for both users when they are detected in the decreasing signal strength order. As in (13)–(16), we specify $E_k(\phi)$ for user k and denote the other user's index as j so that all formulas apply for the cases $(k, j) \in \{(1, 2), (2, 1)\}$.

The D-DFD for the two-user channel is the decorrelating detector for the stronger user and the two-stage detector for the weaker user so that the AEE is given as

$$E_k(\phi_{\text{D-DF}}) = \begin{cases} E_k(\phi_D), & \text{if } 0 \leq \sqrt{\frac{w_j}{w_k}} \leq 1 \\ E_k(\phi_{2S}), & \text{if } \sqrt{\frac{w_j}{w_k}} \geq 1. \end{cases} \tag{54}$$

where $E_k(\phi_D)$ and $E_k(\phi_{2S})$ are given in (15) and (18), respectively.

Proposition 2: The AEE of the O-DFD for the two-user channel where the stronger user is detected first and the weaker user last is given as

$$E_k(\phi_{\text{O-DF}}) = \begin{cases} E_k(\phi_{\text{ML}}), & \text{if } 0 \leq \sqrt{\frac{w_j}{w_k}} \leq |\rho| \\ E_k(\phi_D), & \text{if } |\rho| < \sqrt{\frac{w_j}{w_k}} \leq 1 \\ E_k(\phi_{2S}), & \text{if } 1 < \sqrt{\frac{w_j}{w_k}} \leq \dfrac{1}{|\rho|} \\ E_k(\phi_{\text{ML}}), & \text{if } \sqrt{\frac{w_j}{w_k}} \geq \frac{1}{|\rho|} \end{cases} \tag{55}$$

with $E_k(\phi_{\text{ML}})$ and $E_k(\phi_D)$ are given in (13) and (15). △

Proof: Consider the first two cases of (55) with $0 \leq \sqrt{w_j/w_k} \leq 1$ follow from the fact that since user k is stronger, it is detected first using the linear optimum detector. If $0 \leq \sqrt{w_j/w_k} \leq |\rho|$, the linear optimum detector achieves optimum performance and if $|\rho| < \sqrt{w_j/w_k} \leq 1$, it achieves decorrelator performance (see (15)). In the other two cases, $\sqrt{w_j/w_k} > 1$ so that user j is stronger and is detected first using the linear optimum detector and user k is detected second. If $1 < \sqrt{w_j/w_k} \leq 1/|\rho|$, the linear optimum detector for user j coincides with the decorrelator so that user k achieves the performance of the two-stage detector with decorrelating first stage. Finally, if $\sqrt{w_j/w_k} \geq 1/|\rho|$, user j achieves optimum detector performance. However, the residual interference that results from subtracting the interference from user j in the matched filter output of user k is not statistically independent of its additive Gaussian noise component. As a result, the AEE cannot be computed as in the case of ϕ_{2S}. A lower bound can, however, be obtained using (40). It can be verified that the lower bound indeed coincides with $E_k(\phi_{\text{ML}})$ which is also an upper bound by virtue of the fact that it is the highest achievable value. Consequently, the inequality holds with equality and we have the last part of (55). □

Example 4: An interesting counterexample results for low values of ρ. It is left to the reader to verify that when $|\rho| \leq 1/3$, $E_k(\phi_{\text{O-DF}})$ matches the two-stage detector performance when $\sqrt{w_j/w_k} > 1$ whose AEE in turn is equal to the actual energy w_k. Moreover, the AEE is discontinuous at $\sqrt{w_j/w_k} = 1$. It continues to exhibit a discontuinity at $\sqrt{w_j/w_k} = 1$ for all values of $|\rho| < 1/2$ beyond which the discontinuity vanishes. Consider, for instance, the equal-energy situation and $|\rho| \leq 1/3$. The AEE for either user for the two-stage detector is equal to w_k so that the two-stage detector hence achieves optimum AEE for both users (see also Fig. 6). The O-DFD (the O-DFD is identical to the D-DFD in this case), on the other hand, achieves, say for $\rho = 0.3$, $E_k(\phi_{\text{O-DF}}) = 0.91 w_k$ and $E_j(\phi_{\text{O-DF}}) = w_j$. This example answers "no" to both questions "does the O-DFD (with decreasing signal strength order of detection) always outperform the two-stage detector for every user?" and "does the O-DFD always outperform the two-stage detector in symmetric energy?" This does not, of course, contradict Theorem 1 since the two-stage detector does not belong to the class of DFD's.

Proposition 3: When

$$\sqrt{w_2/w_1} \in [0, |\rho|] \cup [1/|\rho|, \infty)$$

the O-DFD that detects the stronger user first achieves the AEE of both users simultaneously, i.e., $E_1(\phi_{\text{O-DF}}) = E_1(\phi_{\text{ML}})$ and $E_2(\phi_{\text{O-DF}}) = E_2(\phi_{\text{ML}})$. △

Proof: That $E_1(\phi_{\text{O-DF}}) = E_1(\phi_{\text{ML}})$ follows directly from Proposition 2 by considering $(k, j) = (1, 2)$ in (55). Given the hypothesis of this proposition, it is also true that

$$\sqrt{w_1/w_2} \in [0, |\rho|] \cup [1/|\rho|, \infty)$$

so that $E_2(\phi_{\text{O-DF}}) = E_2(\phi_{\text{ML}})$ follows from (55) with $(k, j) = (2, 1)$. □

It is of interest to extend the result of the above proposition for the general K-user problem. Sufficient conditions under which the O-DFD achieves maximum-likelihood performance for the K-user channel are derived in Section IV-L. The results in the next subsection are a prerequisite.

K. Achieving Genie-Aided Performance with the O-DFD

Notation: For a given subset $S \subseteq \{1, \cdots, K\}$, and for some $k \in S$, let η_{k, l_o}^S denote the asymptotic efficiency of the linear optimum detector for user k when designed for, and employed over, a fictitious user-expurgated channel with only users in the set S active. An algorithm that computes this quantity can be found in [20] for $S = \{1, \cdots, K\}$ and can hence be applied to the case of nontrivial S. Define the amplitude ratios $\gamma_k = \sqrt{w_k/w_{k+1}}$.

Corollary 5: The O-DFD will achieve its genie-aided upper bounds on AEE's for every user if the energy ratios satisfy the following inequalities for $1 \le k \le K-1$:

$$\gamma_k^2 \ge \frac{\eta_{k+1,lo}^{\{k+1,\cdots,K\}}(\gamma_{k+1}, \cdots, \gamma_{K-1})}{\tilde{L}_{kk}^2} \tag{56}$$

where we explicitly indicate the dependence of $\eta_{k,lo}^{\{k+1,\cdots,K\}}$ on the amplitude ratios $\gamma_{k+1}, \cdots, \gamma_{K-1}$.

Proof: Sufficient conditions under which the O-DFD achieves the AEE performance of its genie-aided version result from an application of Theorem 3 which requires that the AEE's of the genie-aided O-DFD be in nonincreasing order, i.e., for each k in order from $K-1$ to 1 we need

$$w_k \eta_{k,lo}^{\{k,\cdots,K\}}(\gamma_k, \cdots, \gamma_K) \ge w_{k+1}\eta_{k+1,lo}^{\{k+1,\cdots,K\}}(\gamma_{k+1}, \cdots, \gamma_K) \tag{57}$$

for which the conditions in (56) are sufficient because

$$\eta_{k,lo}^{\{k,\cdots,K\}}(\gamma_k, \cdots, \gamma_K) \ge \tilde{L}_{kk}^2$$

the two being the asymptotic efficiencies of the linear optimum and decorrelating detectors for the user-expurgated channel with active users $k, \cdots, K$. □

The weakening of the inequalities in (57) allows us to obtain a set of conditions that can be easily verified. This feature emerges from the fact that unlike $\eta_{k,lo}^{\{k,\cdots,K\}}(\gamma_k, \cdots, \gamma_K)$, its lower bound $\tilde{L}_{kk}^2$ is independent of the energy ratios $\gamma_k, \cdots, \gamma_{K-1}$ (the independence on γ_k being the critical property). Note that (56) represents a set of sufficient conditions that require that the user energies be sufficiently disparate. They are more stringent for the O-DFD than for the D-DFD in Corollary 2.

Example 5: Consider a three-user channel with correlation matrix $\boldsymbol{R}$ with $R_{kl} = \rho^{|k-l|}$ as in Example 1. The sufficient conditions of Corollary 5 can be shown to be

$$\gamma_2^2 \ge \frac{1}{1-\rho^2} \quad \text{and} \quad \gamma_1^2 \ge \frac{1+\gamma_2^2 - 2|\rho|\gamma_2}{(1-\rho^2)\gamma_2^2}, \quad \text{for} \quad |\rho| > \frac{1}{\sqrt{2}} \tag{58}$$

and the corresponding AEE's are

$$E_1(\phi_{\text{O-DF}}) = E_1(\phi_{LO})$$
$$E_2(\phi_{\text{O-DF}}) = w_2 \frac{1+\gamma_2^2 - 2|\rho|\gamma_2}{(1-\rho^2)\gamma_2^2}$$

and

$$E_3(\phi_{\text{O-DF}}) = w_3. \tag{59}$$

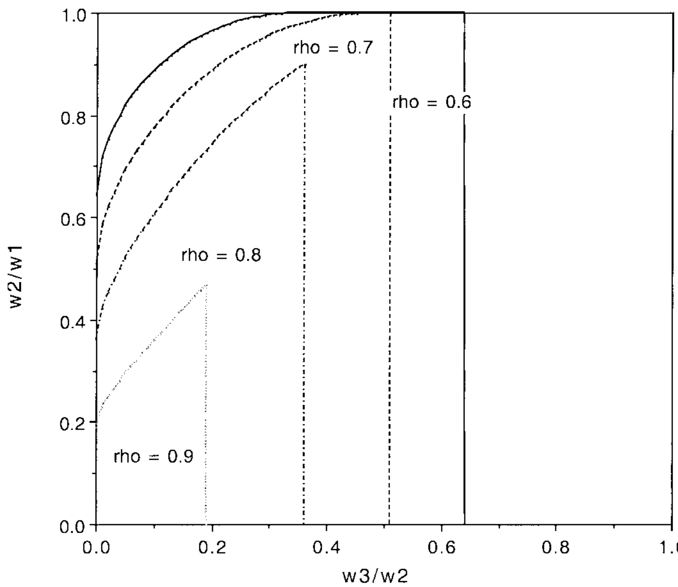

Fig. 10. The boundaries of the energy ratios within which the O-DFD achieves genie-aided performance.

When $|\rho| \le \frac{1}{\sqrt{2}}$ and

$$\gamma_2^2 \ge \frac{1}{1-\rho^2}$$
$$\text{and } \gamma_1^2 \ge \begin{cases} 1, & \text{if } \frac{1}{1-\rho^2} \le \gamma_2^2 \le \frac{1}{\rho^2} \\ \frac{1+\gamma_2^2-2|\rho|\gamma_2}{(1-\rho^2)\gamma_2^2}, & \text{if } \gamma_2^2 \ge \frac{1}{\rho^2} \end{cases} \tag{60}$$

the AEE's are given as

$$E_1(\phi_{\text{O-DF}}) = E_1(\phi_{LO})$$
$$E_2(\phi_{\text{O-DF}}) = \begin{cases} w_1(1-\rho^2), & \text{if } \frac{1}{1-\rho^2} \le \gamma_2^2 \le \frac{1}{\rho^2} \\ w_2\frac{1+\gamma_2^2-2|\rho|\gamma_2}{(1-\rho^2)\gamma_2^2}, & \text{if } \gamma_2^2 \ge \frac{1}{\rho^2} \end{cases}$$

and

$$E_3(\phi_{\text{O-DF}}) = w_3. \tag{61}$$

The regions described by (58) and (60) are depicted in Fig. 10 for various values of ρ.

The subset of $\mathbb{R}^{K-1}$ that contains energy ratios that satisfy (56) is no longer a hyperquadrant (as it was in the case of the D-DFD) because of the complicated dependence (cf. [20]) of the linear optimum detector asymptotic efficiency $\eta_{k+1,lo}^{\{k+1,\cdots,K\}}$ on the energy ratios $\gamma_{k+1}, \cdots, \gamma_{K-1}$. Notice, however, that in spite of the complicated structure of the set of energy ratios that satisfy (56), it is still possible to compute any set of energy ratios that will satisfy the sufficient conditions. The key is to first determine γ_{K-1} that satisfies the last condition (corresponding to $k = K-1$) of (56) and then proceed to pick γ_{K-2} that satisfies the second-last condition (corresponding to $k = K-2$) of (56), and so on, down to γ_1. The difficulty is a computational one and comes from having to compute the asymptotic efficiency of linear optimum detectors as in [20]. It can be easily circumvented if we are willing to only access a subset of the energy ratios characterized by (56) with the conditions

$$\gamma_k^2 \geq \frac{\eta_{k+1,\,opt}^{\{k+1,\,\cdots,\,K\}}(\gamma_{k+1},\,\cdots,\,\gamma_{K-1})}{\tilde{L}_{kk}^2}. \quad (62)$$

The term in the numerator is the optimum asymptotic efficiency of user $k+1$ in the fictitious user-expurgated channel consisting of users $\{k+1,\,\cdots,\,K\}$ and is given as

$$\eta_{k+1,\,opt}^{\{k+1,\,\cdots,\,K\}}(\gamma_{k+1},\,\cdots,\,\gamma_{K-1}) = \frac{1}{w_{k+1}} \min_{e \in E_{k+1,\{k+1,\,\cdots,\,K\}}} \boldsymbol{e}^T \boldsymbol{H} \boldsymbol{e} \quad (63)$$

where, for a subset $S \subseteq \{1,\,\cdots,\,K\}$, and with $j \in S$, we define

$$E_{j,\,S} \triangleq \{\boldsymbol{e};\, e_j \neq 0,\, e_l = 0\,\forall\, l \in \overline{S},\ \text{and}\ e_l \in \{-1,\,0,\,+1\}\,\forall\, l \in S\}. \quad (64)$$

L. Achieving Optimum AEE with the O-DFD for All Users

The O-DFD can, under certain conditions, achieve optimum AEE for all users simultaneously if the user energies are sufficiently disparate. This is in contrast to the linear optimum detector, whose symmetric energy often coincides with that of the decorrelator and can therefore do poorly relative to the optimum detector.

Theorem 7: Sufficient conditions for the O-DFD to achieve the AEE's of the maximum-likelihood detector for every user are that for k ranging from $K-1$ down to 1

$$\gamma_k^2 \geq \max\left\{ \frac{\min_{e \in E_{k+1,\,\{k+1,\,\cdots,\,K\}}} \boldsymbol{e}^T\boldsymbol{H}\boldsymbol{e}}{w_{k+1}\tilde{L}_{kk}^2},\ \max_{j=k,\,\cdots,\,K}{}^2 \left(\frac{1}{|R_{kj}|} \sum_{i=k+1}^{K} \left(\prod_{l=1}^{i-k-1} \gamma_{k+l}^{-1} \right) |R_{ij}| \right) \right\}. \quad (65)$$

Proof: The conditions in (65) imply that for $1 \leq k \leq K-1$

$$\gamma_k \geq \max_{j=k,\,\cdots,\,K} \left(\frac{1}{|R_{kj}|} \sum_{i=k+1}^{K} \left(\prod_{l=1}^{i-k-1} \gamma_{k+l}^{-1} \right) |R_{ij}| \right). \quad (66)$$

Applying the result in (11), it is easily seen that for $1 \leq k \leq K$, the above conditions ensure that the kth user achieves optimum AEE by using the linear optimum detector in the user-expurgated channel consisting of users $\{k,\,\cdots,\,K\}$. This implies that the genie-aided O-DFD achieves these optimum AEE's of successively user-expurgated channels. Next, the conditions in (65) imply those in (62) so that the O-DFD AEE's are equal to those of the genie-aided O-DFD. Since the genie-aided O-DFD AEE's are in turn optimal for successively user-expurgated channels, they are at least as large as the optimal AEE's achievable for the K-user channel. The latter are upper bounds on the performance of any detector and hence on the O-DFD in particular. As a result, the O-DFD achieves optimum AEE's for all users provided the conditions in (65) are satisfied. □

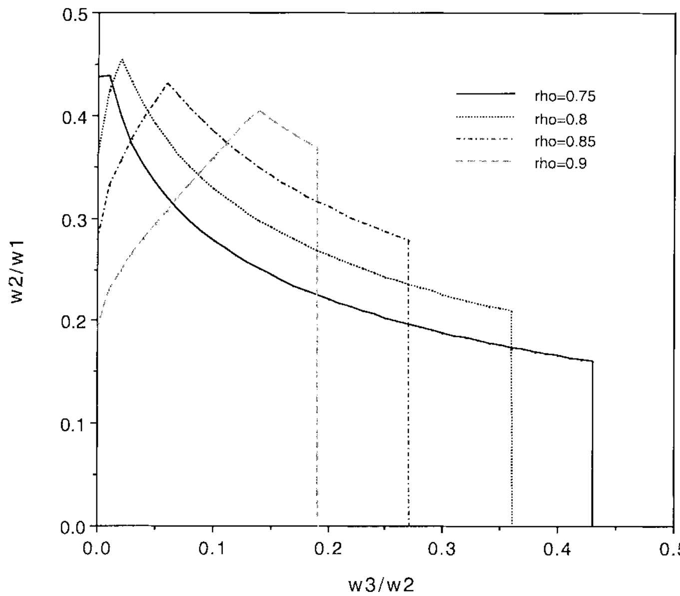

Fig. 11. The boundaries of the energy ratios within which the O-DFD achieves the AEE performance of the optimum detector for all users simultaneously.

Note that the above sufficient conditions are not only easy to verify for a given set of energy ratios, but they can also be used to successively compute any set of energy ratios from γ_{K-1} down to γ_1 that will satisfy them.

Example 6: Let us first consider a two-user example with normalized correlation denoted as ρ. It is easily seen that the above theorem yields the sufficient condition

$$\gamma_1^2 \geq \max\left\{ \frac{1}{1-\rho^2},\, \frac{1}{\rho^2} \right\}. \quad (67)$$

Consider the 3-user channel of Example 5 and restrict attention to the interesting case of high correlation where $\rho \geq 1/\sqrt{2}$. The sufficient conditions for all users to achieve optimum AEE as given by the above theorem and when applied to the problem at hand yield

$$\gamma_2^2 \geq \frac{1}{1-\rho^2}$$
$$\text{and } \gamma_1^2 \geq \max\left\{ \frac{1+\gamma_2^2 - 2|\rho|\gamma_2}{(1-\rho^2)\gamma_2^2},\, \frac{\rho^2\gamma_2^2 + 1 + 2|\rho|\gamma_2}{\rho^4\gamma_2^2} \right\}. \quad (68)$$

The region described by (68) is depicted in Fig. 11 for various values of ρ.

It is possible to obtain a tighter set of sufficient conditions than those stated in the above theorem. The price paid is in the ease of verification. From k starting at $K-1$ down to 1, let

$$\gamma_k^2 = \max_{j=k,\,\cdots,\,K}{}^2 \left(\frac{1}{|R_{kj}|} \sum_{i=k+1}^{K} \left(\prod_{l=1}^{i-k-1} \gamma_{k+l}^{-1} \right) |R_{ij}| \right) \triangleq X_k \quad (69)$$

and let

$$Y_k = \frac{\min_{e \in E_{k+1,\,\{k+1,\,\cdots,\,K\}}} \boldsymbol{e}^T\boldsymbol{H}\boldsymbol{e}}{\min_{e \in E_{k,\,\{k,\,\cdots,\,K\}}} \boldsymbol{e}^T\boldsymbol{H}\boldsymbol{e}}. \quad (70)$$

Note that the denominator in the above expression depends on γ_k. If $Y_k \le X_k$, then proceed to evaluate γ_{k-1} as in (69). Else, define Z_k as

$$Z_k \triangleq \frac{\min_{\boldsymbol{e} \in E_{k+1, \{k+1, \cdots, K\}}} \boldsymbol{e}^T \boldsymbol{H} \boldsymbol{e}}{\tilde{L}_{kk}^2} \tag{71}$$

and find the smallest value of γ_k^2 in the interval $[X_k, Z_k]$ such that $Y_k \le X_k$. This is guaranteed to happen somewhere in the interval $[X_k, Z_k]$ because it certainly does happen when $\gamma_k^2 = Z_k$. The proof that the above algorithm will yield a set of energy ratios that will imply that the O-DFD will achieve optimum AEE is left to the reader.

Example 7: Let us consider the two-user channel again. The above procedure yields the sufficient condition

$$\gamma_1^2 \ge \max\left\{\frac{1}{\rho^2}, 4\rho^2\right\}. \tag{72}$$

Note that the sufficient condition (72) coincides with that in (67) for $|\rho| \le 1/\sqrt{2}$. However, for higher value of the correlation coefficient, we have that (72) is tighter than the condition in (67) with the difference between the two conditions becoming more significant as $|\rho|$ increases towards unity.

This example must be compared with the result of Proposition 3. We see that the condition obtained therein for both users to achieve optimum AEE with the O-DFD is the same as the region defined by (72) for $|\rho| \le 1/\sqrt{2}$. However, the region defined by Proposition 3 strictly contains that defined in (72) for $|\rho| > 1/\sqrt{2}$. It is not clear how the analysis leading to that proposition must be better generalized to the K-user problem to obtain a tighter set of sufficient conditions than those given by the algorithm at the end of Theorem 7.

Example 8: Consider again the three-user channel described before. In this problem, it is no longer possible to characterize the set of energy ratios in closed form. It is possible to improve on the previous result, however. It can be shown that the sufficient condition for γ_2 is given by

$$\gamma_2^2 \ge \begin{cases} \frac{1}{\rho^2}, & \text{if } |\rho| \le \frac{1}{\sqrt{2}} \\ 4\rho^2, , & \text{if } |\rho| > \frac{1}{\sqrt{2}}. \end{cases} \tag{73}$$

For any γ_2 that satisfies the above equality, the smallest γ_1 is chosen as follows: define X and Z and Y as

$$X \triangleq \left(\frac{1}{|\rho|} + \frac{1}{\rho^2\gamma_2}\right)^2 \quad \text{and} \quad Z \triangleq \frac{1 + \gamma_2^2 - 2|\rho|\gamma_2}{(1-\rho^2)\gamma_2^2}$$
$$\text{and} \quad Y \triangleq \frac{1 + \gamma_2^2 - 2|\rho|\gamma_2}{\gamma_2^2 \eta_{1, opt}}. \tag{74}$$

For any given γ_2 that satisfies (73), if $X > Y$, then we have $\gamma_1^2 \ge X$. Else, find the smallest γ_1 in the interval $[X, Z]$ for which $\gamma_1^2 \ge Y$.

The region described by the above algorithm is shown in Fig. 12. The region described by the less stringent set of sufficient conditions of Theorem 7 given in (68) for this example are also shown for comparison. Note that the tighter sufficient conditions are significantly better for higher values of correlation.

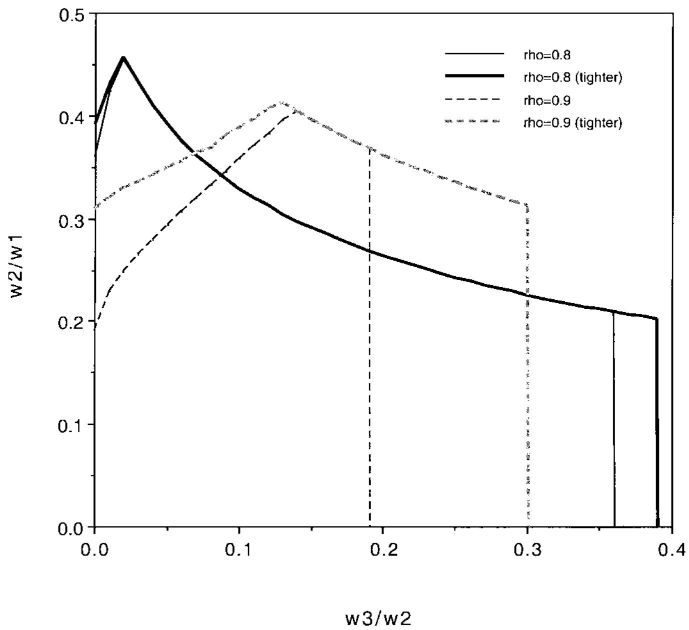

Fig. 12. The boundaries of the energy ratios according to the tighter sufficient conditions within which the O-DFD will achieve the AEE performance of the optimum detector for all users simultaneously.

V. Conclusions

This paper adopts a rigorous approach to both analysis and design of decision-feedback multiuser detectors without relying on the perfect feedback assumption. A new performance measure, called symmetric energy, is introduced that is an indicator of the joint error rate in high-SNR regions of a multiuser detector that at least one user is detected erroneously. The symmetric energy of an arbitrary decision feedback detector is shown to be equal to the symmetric energy of the genie-aided version of that detector where the genie ensures perfect feedback. Using this property, the O-DFD is found as a solution to the problem of optimizing symmetric energy among decision feedback detectors. A comparison of several well-known linear and decision feedback detectors including the O-DFD is given in terms of symmetric energy which reveals that decision feedback detectors achieve at least the performance of their linear counterparts independently of the order in which they are detected and much higher performance than their linear counterparts when users are detected in well-chosen orders. The implications of our work on power control for multiuser detection are detailed through the two-user channel. Ordering algorithms are given whereby decision feedback detectors can user-wise outperform their linear counterparts. The usual ordering of users according to the nonincreasing order of energies does not guarantee this result. Several results on the performance analysis of the decorrelating decision feedback detector are given including exact formulas for the bit-error rate and the asymptotic effective energy. It is also shown that the O-DFD can, under certain conditions, achieve the asymptotic effective energy performance of the maximum-likelihood detector for every user. Moreover, there are several ideas in this paper that apply to more general settings than to linear modulation and the synchronous CWMA channel model.

ACKNOWLEDGMENT

The author would like to thank Deepak Das and Dara Parsavand for writing the computer programs to generate the figures of this paper. He would also like to thank the anonymous reviewers for their helpful comments. The more detailed proofs of some results and the occasional tutorial exposition are in response to one reviewer's suggestion.

REFERENCES

[1] S. A. Altekar and N. C. Beaulieu, "Upper bounds to the error probability of decision feedback equalizers," *IEEE Trans. Inform. Theory*, vol. 37, pp. 145–156, Jan. 1991.

[2] R. M. Buehrer and B. D. Woerner, "Analysis of adaptive multistage interference cancellation for CDMA using an improved Gaussian approximation," *IEEE Trans. Commun.*, vol. 44, pp. 1308–1321, Oct. 1996.

[3] G. Caire, G. Taricco, J. Ventura-Traveset, and E. Biglieri, "A multiuser approach to narrowband cellular communication," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1503–1517, Sept. 1997.

[4] T. M. Cover, "Some advances in broadcast channels," in *Advances in Communication Systems.* New York: Academic, 1975, pp. 229–260.

[5] M. C. Davis, "Factoring the spectral matrix," *IEEE Trans. Automat. Contr.*, vol. AC-8, pp. 296–305, Oct. 1963.

[6] D. Divsalar, M. Simon, and D. Raphaeli, "Improved parallel interference cancellation for CDMA," *IEEE Trans. Commun.*, vol. 46, pp. 258–268, Feb. 1998.

[7] A. Duel-Hallen, "Decorrelating decision feedback multiuser detector for synchronous code-division multiple-access channel," *IEEE Trans. Commun.*, vol. 41, pp. 285–290, Feb. 1993.

[8] A. Duel-Hallen, J. Holtzman, and Z. Zvonar, "Multiuser detection for CDMA systems," *IEEE Personal Commun.*, pp. 46–58, Apr. 1995.

[9] A. Duel-Hallen, "A family of decision feedback multiuser detetcors for asynchronous code-division multiple-access channels," *IEEE Trans. Commun.*, vol. 43, nos. 2/3/4, pp. 421–434, 1995.

[10] D. L. Duttweiler, J. E. Mazo, and D. G. Messerschmitt, "An upper bound on the error probability in decision-feedback equalization," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 490–496, July 1974.

[11] W. Van Etten, "Maximum likelihood receiver for multiple channel transmission systems," *IEEE Trans. Commun.*, vol. 44, pp. 276–283, Feb. 1996.

[12] E. Fain and M. K. Varanasi, "Cyclic decision feedback multiuser sequence detection," in *Proc. 32nd Annu. Allerton Conf. Communation, Control, and Computing*, Oct. 1994, pp. 372–381.

[13] V. Ghazi-Moghadam, L. B. Nelson, and M. Kaveh, "Parallel interference cancellation for CDMA systems," in *Proc. 33nd Annu. Allerton Conf. Communication, Control, and Computing*, Oct. 1995, pp. 216–224.

[14] S. D. Gray, M. Kocic, and D. Brady, "Multiuser detection in mismatched multiple-access channels," *IEEE Trans. Commun.*, vol. 43, pp. 3080–3089, Dec. 1995.

[15] R. Haas and J. C. Belfiore, "Spectrum efficiency limits in mobile cellular systems," *IEEE Trans. Veh. Technol.*, vol. 45, pp. 33–40, Feb. 1996.

[16] C. Hegarty and B. Vojcic, "Two-stage multiuser detection for noncoherent CDMA," in *Proc. 33rd Allerton Conf. Communication, Control and Computing*, Monticello, IL, Oct. 1995, pp. 1063–1072.

[17] R. Iltis and L. Mailaender, "Multiuser detection of quasi-synchronous CDMA signals using linear decorrelators," *IEEE Trans. Commun.*, vol. 44, pp. 1561–1571, Nov. 1996.

[18] M. Juntti and B. Aazhang, "Finite memory-length linear multiuser detection for asynchronous CDMA communications," *IEEE Trans. Commun.*, vol. 45, pp. 611–622, May 1997.

[19] R. Kohno, H. Imai, and M. Hatori, "Cancellation techniques of co-channel interference in asynchronous spread spectrum multiple access systems," *Trans. IECE Japan* , vol. J66-A, no. 5, pp. 416–423, May 1983, in Japanese. Also, *Electron. Commun. in Japan.*, vol. 66, no. 5, pp. 20–29, May 1983, English version.

[20] R. Lupas and S. Verdú, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 123–136, Jan. 1989.

[21] ——, "Near–far resistance of multiuser detectors in asynchronous channels," *IEEE Trans. Commun.*, vol. 38, pp. 496–508, Apr. 1990.

[22] U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.*, vol. 32, pp. 3178–3188, Dec. 1994.

[23] S. Moshavi, "Multi-user detection for DS-CDMA communications," *IEEE Commun. Mag.*, pp. 124–136, Oct. 1996.

[24] P. Patel and J. Holtzman, "Analysis of a DS/CDMA interference cancellation scheme using correlations," in *Proc. Globecom 1993,*, Nov./Dec. 1993, pp. 76–80.

[25] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Spinger-Verlag, 1989.

[26] H. V. Poor and S. Verdú, "Probability of error in MMSE multiuser detection," *IEEE Trans. Inform. Theory*, vol. 43, pp. 858–871, May 1997.

[27] K. S. Schneider, "Optimum detection of code-division multiplexed signals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-15, pp. 181–185, Jan. 1979.

[28] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Processing*, vol. 42, pp. 2146–2157, Aug. 1994.

[29] C. Schegel, S. Roy, P. Alexander, and Z. Xiang, "Multi-user projection receivers," *IEEE J. Select. Areas Commun.*, vol. 14, pp. 1610–1618, Oct. 1996.

[30] Z. L. Shi, W. Du, and P. F. Driessen, "A new multistage detector for synchronous CDMA communications," *Trans. Commun.*, vol. 44, pp. 538–541, May 1996.

[31] M. Tsatsanis, "Inverse filtering criteria for CDMA systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 102–112, Jan. 1997.

[32] Telecommunications Industry Association, "Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread-Spectrum Cellular System," Std. IS-95A, TIA/EIA, Washington, DC, 1995.

[33] S. Ulukus and R. Yates, "Optimum multiuser detection is tractable for synchronous CDMA systems using M-sequences," *IEEE Commun. Lett.*, vol. 2, no. 4, pp. 89–91, 1998.

[34] M. K. Varanasi and B. Aazhang, "Near-optimum demodulation for coherent communications in asynchronous Gaussian CDMA channels," in *Proc. CISS*, Mar. 1988, pp. 832–839.

[35] ——, "Probability of error comparison of linear and iterative multiuser detectors," in *Proc. ComCon 88* (Baton Rouge, LA, Oct. 1988).

[36] ——, "Multistage detection for asynchronous code-division multiple-access communications," *IEEE Trans. Commun.*, vol. 38, pp. 509–519, Apr. 1990.

[37] ——, "Near-optimum detection in synchronous code-division multiple-access systems," *IEEE Trans. Commun.*, vol. 39, pp. 725–736, May 1991.

[38] ——, "Optimally near–far resistant multiuser detection in differentially coherent synchronous channels," *IEEE Trans. Inform. Theory*, vol. I37, pp. 1006–1018, July 1991.

[39] M. K. Varanasi, "Group detection in QAM synchronous CDMA channels," in *Proc. 1992 Conf. Information Sciences and Systems* (Princeton Univ., Princeton, NJ, Mar. 1992), pp. 820–826.

[40] ——, "Noncoherent detection in asynchronous multiuser channels," *IEEE Trans. Inform. Theory*, vol. 39, pp. 157–176, Jan. 1993.

[41] M. K. Varanasi and S. Vasudevan, "Multiuser detectors for synchronous CDMA communication over nonselective Rician fading channels," *IEEE Trans. Commun.*, vol. 42, pp. 711–722, Feb. 1994.

[42] M. K. Varanasi, "Group sequence detection: Bridging the yawning gap between linear and optimum multiuser detection," in *Proc. Workshop Information Theory, Multiple-Access and Queueing*, Apr. 1995, p. 26.

[43] ——, "Group detection for synchronous Gaussian code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1083–1096, July 1995.

[44] ——, "Power control for multiuser detection," in *1996 Conf. Information Sciences and Systems* (Princeton Univ., Pronceton, NJ, Mar. 1996), pp. 866–873.

[45] ——, "Optimizing symmetric energy and permuting users for decision feedback multiuser detection to user-wise outperform linear multiuser detection," in *Proc. Conf. Information Sciences and Systems* (The Johns Hopkins Univ., Baltimore, MD), Mar. 1997, pp. 492–497.

[46] ——, "Multiuser receivers with power control," a National Science Foundation Proposal NCR-9714679, submitted to the Networking and Communications Research Program in April 1997 and as "Power Control for Multiuser detectors in CDMA Systems," to the Office of Naval Research in October 1996, CU Proposal 1096.03.1364B.

[47] M. K. Varanasi and T. Guess, "Bandwidth efficient multiple-access via signal design for decision feedback receivers: Towards an optimal spreading-coding trade-off," in *Proc. Communications Theory Mini-Conf., IEEE Global Telecomm Conf.* (Phoenix, AZ, Nov. 1997), pp.

159–165.
[48] ——, "Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple-access channel," in *Proc. 31st Asilomar Conf. Signals, Systems, and Computers* (Pacific Grove, CA, Nov. 1997), pp. 1405–1409.
[49] S. Verdú, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 85–96, Jan. 1986.
[50] ——, "Optimum multiuser asymptotic efficiency," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 890–897, Sept. 1986.
[51] ——, "Multiuser detection," in *Advances in Statistical Signal Processing: Signal Detection*, H. V. Poor and J. B. Thomas, Eds. Greenwich, CT: JAI Press, 1993, pp. 369–410.
[52] A. J. Viterbi, "Very low rate convolutional codes for maximum theoretical performance of spread-spectrum multiple-access channels," *IEEE J. Select. Areas Commun.*, vol. 8, no. 4, pp. 641–649, May 1990.
[53] ——, *CDMA: Principles of Spread-Spectrum Communications.* Reading, MA: Addison-Wesley, 1995.
[54] X. Wang and H. V. Poor, "Blind multiuser detection: A subspace approach," *IEEE Trans. Inform. Theory*, vol. 44, pp. 91–103, Jan. 1998.
[55] N. Wiener and P. Masani, "The prediction theory of multivariate stochastic processes I and II," *Acta Math.*, vols. 98/99, pp. 112–150 and 93–137, Nov. 1957, Apr. 1958.
[56] A. D. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 2–10, Jan. 1974.
[57] Z. Xie, C. T. Rushforth, and R. T. Short, "Multi-user signal detection using sequential decoding," *IEEE Trans. Commun.*, vol. 38, pp. 578–583, May 1990.
[58] Z. Xie, R. T. Short, and C. T. Rushforth, "A family of suboptimum detectors for coherent multiuser communications," *IEEE J. Select. Areas Commun.*, vol. 8, pp. 683–690, May 1990.
[59] J. Yang and S. Roy, "Joint transmitter–receiver optimization for multi-input multi-output systems with decision feedback," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1334–1347, Sept. 1994.
[60] ——, "On joint transmitter and receiver optimization for multiple-input–multiple-output (MIMO) transmission," *IEEE Trans. Commun.*, vol. 42, Dec. 1994.
[61] Y. C. Yoon, R. Kohno, and H. Imai, "Combination of an adaptive array antenna and a canceller of interference for direct-sequence spread-spectrum multiple-access system," *IEEE J. Select. Areas Commun.*, vol. 8, May 1990.
[62] X. Zhang and D. Brady, "Asymptotic multiuser efficiency for decision-directed multiuser detection," *IEEE Trans. Inform. Theory*, vol. 44, pp. 502–515, Mar. 1998.