RECENT PROGRESS IN MULTIUSER DETECTION

Sergio Verdú

Princeton University

I. Introduction

Multiuser Detection is an active research area in Detection Theory. Its objective is the study of strategies to demodulate the digital information sent simultaneously by several transmitters who share a multiple-access channel. Common channels that are encompassed by this general model include up-link satellite channels, local-area networks and radio networks.

Static strategies, such as Frequency-Division Multiple-Access (FDMA) and Time-Division Multiple-Acess (TDMA), whereby the multiple-access channel is effectively partitioned into independent single-user sub-channels are common in practice. The analysis of those approaches lies fully whithin the domain of classical single-user communication, and offers few, if any, conceptual challenges. Moreover, from the practical viewpoint, static strategies tend to be wasteful in applications where most users actively send information during a very small percentage of time. Current dynamic channel-sharing strategies, which allow the active users a larger share of the channel while they are transmitting, fit into two categories: random-access communication and simultaneous transmission systems. In random-access communication it is assumed that the receiver is not capable of demodulating more than one simultaneous transmission, and so the problem is to design protocols to schedule channel access at nonoverlapping times, and if collisions between messages occur (as they do when the protocol operates in a decentralized fashion) to ensure that those messages are eventually retransmitted successfully. Simultaneous transmission systems differ from static strategies and random-access protocols in that users are allowed to transmit simultaneously, asynchronously and through the same channel. The destination receives a noisy version of the superposition of all the transmitted waveforms and it is the task of the receiver to demodulate all (as in the case of the satellite channel) or a subset (as in multipoint-to-multipoint topologies) of the transmitted messages. This is the communication problem that is the subject of multiuser detection and multiuser information theory. In practice, the major multiaccess strategy using the simultaneous transmission philosophy is Code-Division Multiple-Access (CDMA). In this technique, each transmitter is assigned a fixed, distinct signature waveform which he uses to modulate his message as in single-user communication. Then the information sent by each user can be demodulated by correlating the received signal with each of the signature waveforms. This demodulator, whose use is widespread in practice, is referred to as the conventional single-user detector. As is well-known, when the channel output is corrupted by additive white Gaussian noise, the conventional single-user detector minimizes the probability of error in a single-user channel, i.e., in the absence of interfering users. The fact that this is no longer true in the multiple-access channel is the raison d'être of the area of multiuser detection.

The performance of the conventional single-user detector is acceptable provided that the energies of the received signals are not too dissimilar and that the signature waveforms are designed so that their crosscorrelations are low enough (this depends on the desired maximum number of simultaneous users). In practice, low crosscorrelations are usually achieved employing Spread-Spectrum Pseudonoise sequences of long periodicity. If the received signal energies are indeed dissimilar, i.e., some users are very weak in comparison to others, then the conventional single-user detector is unable to recover the messages of the weak users reliably, even if the signature waveforms have very low crosscorrelations. This is known as the *near-far problem* and is the main shortcoming of

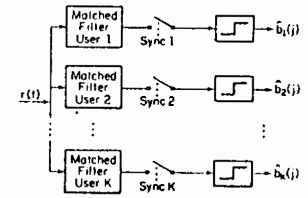

Figure 1. Conventional detector.

currently operational Direct-Sequence Spread-Spectrum Multiple-Access systems.

Due to the reduced multiple-access capability and the increased vulnerability to hostile sources caused by the near-far problem, its solution or alleviation had been a target of researchers in the area for several years. Before the emergence of solutions based on multiuser detection, success had been very limited and, essentially, the only remedies available were power control and the design of signals with even more stringent crosscorrelation properties. Unfortunately, power control (i.e., the adaptive adjustment of transmitter power depending on its location and on the received powers of the other users) dictates significant reductions in the transmitted powers of the strong users in order for the weaker users to achieve reliable communication. Thus, power control can become self-defeating since it actually decreases the overall multiple-access and antijamming capabilities of the system. Furthermore, more and more complex signature waveforms lead to rapid increases in system cost and bandwidth, and, as we have noted, do not eliminate the near-far problem. For these reasons, it can be seen why the solution to the near-far problem has been highlighted as the main achievement of multiuser detection.

The chief reason why multiuser detection did not develop until relatively recently was the belief shared by many a worker in Spread Spectrum that multiuser interference is accurately modeled as a white Gaussian random process, and thus the conventional detector is essentially optimum. This serves as a good illustration of A. N. Kolomogorov's assertion that "the Central Limit Theorem is a dangerous tool in the hands of amateurs," and it ranks along with the stability of the original ALOHA algorithm, and the independence assumption in queueing models of store-and-forward communication networks as one of the maximum exponents of the area of "voodoo multiuser communication." While it is not difficult to build an infinite population multiuser signal model which can be rigorously shown to be asymptotically Gaussian as the individual amplitudes go to zero with the appropriate speed, the number of transmitters, signature waveforms, and power levels encountered in many practical situations (e.g. in near-far environments) render the Gaussian approximation completely useless. This was recognized by H. V. Poor who proposed in 1980 [9] the use of techniques from both minimax robustness and non-Gaussian signal detection to improve the performance of the conventional single-user detector in multiuser channels. Several earlier works had already investigated receivers that used the knowledge of the interfering signature waveforms to improve the performance of Code-Division Multiplexing systems; Timor [15, 16] showed that it was possible to double the maximum number of simultaneous users achievable with the conventional noncoherent demodulator of Frequency-Hopped FSK systems, Horwood and Gagliardi [3] and Schneider [12, 13] considered multiuser receivers for synchronous systems. In particular, [12] claimed that an appropriately chosen linear transformation of a bank of matched filter outputs results in optimum decisions. Although that claim turned out to be erroneous, such a receiver and its generalization to the asynchronous case do have very desirable properties as we discuss in the sequel. In the same work, Schneider briefly considered the asynchronous channel arguing that it could be modeled as a finite state machine and conjectured that Viterbi's convolutional decoding algorithm could be used to demodulate the data. Those ideas were closely related to earlier work to combat crosstalk in single-user multichannel communication systems by Shnidman [14], Kaye and George [4] and Van Etten [1, 2] among others. Those works dealt with PAM systems through multi-input multi-output dispersive channels, which are no more than a vector generalization of the conventional scalar intersymbol interference model. In contrast, in the asynchronous multiuser channel each signal overlaps with two consecutive signals of each of the interferers; thereby introducing memory in the channel in a way which (as we will see in Section II) is akin to a scalar periodically time-varying generalization of the conventional intersymbol interference model.

Section II reviews the main strategies proposed so far in multiuser detection, and Section III summarizes the main results on the performance of the various receivers. For the sake of clarity and in order to highlight the fundamental features of the approaches reviewed in this paper, it is convenient to circumscribe the discussion to the two-user case, and the reader is referred to the corresponding treatment with an arbitrary number of users in each of the referenced sources.

II. Multiuser Detectors

The two-user white Gaussian asynchronous multiple-access channel considered in this paper is

$$y(t) = \sum_{i=-M}^{M} b_1(i)\, s_1(t-iT-\tau_1) + \sum_{i=-M}^{M} b_2(i)\, s_2(t-iT-\tau_2) + n(t), \tag{1}$$

where $n(t)$ is white Gaussian noise with power spectral density equal to σ^2. The unit-energy signature waveforms assigned to both users are $s_1(t)$ and $s_2(t)$ and have duration T. The symbol streams $\{b_1(i)\}$ and $\{b_2(i)\}$ take values on $\{-\sqrt{w_1},+\sqrt{w_1}\}$ and $\{-\sqrt{w_2},+\sqrt{w_2}\}$, respectively, where w_k denotes the received energy per bit of user k. The delays $\tau_1 \in [0,T]$ and $\tau_2 \in [0,T]$ account for the asynchronism between both transmitted streams.

The objective is to find the optimum demodulator for the streams of data transmitted by both users. This would appear to apply only to multipoint-to-point topologies where detection is centralized (e.g. the up-link satellite channel). However, in situations where each location is interested in demodulating only one of the transmitted streams, the same strategy can be used as long as the receiver knows the signature waveforms of the interfering users. (The derivation of optimum detectors without knowledge of the interfering signature waveforms is discussed later in this section.) A key point in the derivation of the optimum receiver [20] for the asynchronous channel (1) is that in order to make optimum decisions on any particular bit it is necessary to use the observation of the entire received waveform rather than just the received waveform on the interval of that particular bit. In other words, due to the asynchronism between the users, we face a problem of sequence detection rather than one-shot detection. It is easy to show [20] that the sequence of outputs of matched filters for $\{s_1(t)\}$ and $\{s_2(t)\}$ is a set of sufficent statistics to demodulate the data streams $\{b_1(i)\}$ and $\{b_2(i)\}$. Those sufficient statistics can be obtained by a receiver that knows the signature waveforms of both users upon acquisition of the timing of the bit-epochs of each stream. Assuming without loss of generality that $\tau_1 \le \tau_2$, the sufficient statistics admit the following expression in terms of the data

$$\begin{bmatrix} y_1(i) \\ y_2(i) \end{bmatrix} = \begin{bmatrix} 0 & \rho_{21} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} b_1(i-1) \\ b_2(i-1) \end{bmatrix} + \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix} \begin{bmatrix} b_1(i) \\ b_2(i) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \rho_{21} & 0 \end{bmatrix} \begin{bmatrix} b_1(i+1) \\ b_2(i+1) \end{bmatrix} + \begin{bmatrix} n_1(i) \\ n_2(i) \end{bmatrix} \tag{2}$$

for $-M \le i \le M$ and where

$$\rho_{12} = \int_0^T s_1(t)\, s_2(t+\tau_1-\tau_2)\, dt, \qquad \rho_{21} = \int_0^T s_1(t)\, s_2(t+T+\tau_1-\tau_2)\, dt \tag{3}$$

are the crosscorrelations between the signature waveforms. The discrete-time random process $\{[n_1(i)\ n_2(i)]^T\}$ in (2) is Gaussian with zero-mean and covariance matrix:

$$E\left[\begin{bmatrix} n_1(i) \\ n_2(i) \end{bmatrix} [n_1(j)\ \ n_2(j)] \right] = \sigma^2\, H(i-j)$$

where $H(i) = 0$ if $|i| > 1$, and $H(1)$, $H(0)$ and $H(-1)$ are the matrices appearing in (2), i.e.,

$$H(0) = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix}, \qquad H(1) = H^T(-1) = \begin{bmatrix} 0 & \rho_{21} \\ 0 & 0 \end{bmatrix}.$$

Since the noise in (1) is white and Gaussian, maximum likelihood decisions are obtained by selecting the symbol sequences $\{b_1(i)\}$ and $\{b_2(i)\}$ that best explain the observations in a mean-square sense, i.e., assuming that the optimum detector knows the received energies, it should select the argument achieving

$$\min_{\substack{b_1(i) \in \{-\sqrt{w_1}, +\sqrt{w_1}\} \\ b_2(i) \in \{-\sqrt{w_2}, +\sqrt{w_2}\} \\ i = -M, \cdots M}} \int [y(t) - \sum_{k=1}^{2} \sum_{i=-M}^{M} b_k(i)\, s_k(t-iT-\tau_k)]^2\, dt. \tag{4}$$

Before considering how to obtain a detector that solves for the minimum in (4) it is instructive to consider the synchronous ($\tau_1 = \tau_2$) version of the problem for which a one-shot approach suffices:

$$\min_{\substack{b_1(0) \in \{-\sqrt{w_1}, +\sqrt{w_1}\} \\ b_2(0) \in \{-\sqrt{w_2}, +\sqrt{w_2}\}}} \int_0^T [y(t) - \sum_{k=1}^{2} b_k(0)\, s_k(t)]^2\, dt = \max_{\mathbf{b} \in \{-\sqrt{w_1}, \sqrt{w_1}\} \times \{-\sqrt{w_2}, \sqrt{w_2}\}} 2\mathbf{b}^T \mathbf{y} - \mathbf{b}^T \mathbf{H} \mathbf{b} \tag{5}$$

where $\mathbf{y} = [y_1(0), y_2(0)]^T$, $\mathbf{b} = [b_1(0), b_2(0)]^T$ and $\mathbf{H} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, $\rho = \int_0^T s_1(t) s_2(t)\, dt$.

The solution to the right-hand side of (5) is easy; the function therein can be computed for each of the four possible values of $\mathbf{b}$ and the maximizing vector gives the solution to the optimum demodulation problem. In the K-user case, there are 2^K possible values of $\mathbf{b}$ and no algorithm is known to solve (5) in a number of steps which is polynomial in the number of users. In fact, it has been shown that (5) is an NP-hard combinatorial optimization problem [25], and, therefore, it can admit a polynomial solution only if the same is true for longstanding combinatorial problems such as the *travelling salesman* and *integer linear programming* problems.

In the asynchronous case, it is possible to follow the same approach by viewing each symbol as transmitted by a different fictitious user over the interval $[-MT, MT+2T]$. In the K-user case, this is equivalent to having a synchronous channel with $(2M+1)K$ fictitious users. Thus, the complexity of the above brute-force method is out of the question because of the large values of M one expects in applications. Fortunately, the asynchronous problem admits a much more efficient algorithm whose complexity per demodulated bit is very similar to that of the synchronous case (exponential in K but independent of M), making it practical for situations where the number of simultaneous active users does not exceed the range 10-15. The reason for the existence of such an algorithm is that the $(2M+1)K \times (2M+1)K$ matrix $\mathbf{H}$ of crosscorrelations between each possible pair of symbols has a great deal of structure, as each symbol only overlaps with two consecutive symbols of the other users. For example, in the two-user case it is equal to

$$\mathbf{H} = \begin{bmatrix} 1 & \rho_{12} & & & & & \\ \rho_{12} & 1 & \rho_{21} & & & & \\ & \rho_{21} & 1 & \rho_{12} & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & \cdot & \\ & & & & \rho_{21} & 1 & \rho_{12} \\ & & & & & \rho_{12} & 1 \end{bmatrix}. \tag{6}$$

This implies that the objective function in the right-hand side of (5) admits the following additive decomposition

$$
\begin{aligned}
2\mathbf{b}^T\mathbf{y} - \mathbf{b}^T\mathrm{H}\mathbf{b} = {} & \lambda_{-M}^{(1)}[0,\, b_1(-M)] + \lambda_{-M}^{(2)}[b_1(-M),\, b_2(-M)] + \lambda_{-M+1}^{(1)}[b_2(-M),\, b_1(-M+1)] + \cdots \\
& + \lambda_0^{(1)}[b_2(-1),\, b_1(0)] + \lambda_0^{(2)}[b_1(0),\, b_2(0)] + \lambda_1^{(1)}[b_2(0),\, b_1(1)] + \cdots \\
& + \lambda_{M-1}^{(2)}[b_1(M-1),\, b_2(M-1)] + \lambda_M^{(1)}[b_2(M-1),\, b_1(M)] + \lambda_M^{(2)}[b_1(M),\, b_2(M)],
\end{aligned} \tag{7}
$$

where

$$\lambda_i^{(k)}[a,\, b] = 2by_k(i) - w_k - 2ab\,\rho_{jk}, \qquad j \neq k.$$

The interesting thing to note on (7) is that the dependence on the components of $\mathbf{b}$ is sequential; each of the functions $\lambda_i^{(k)}$ depends on only two symbols and every symbol appears only in two consecutive functions. Therefore, the maximization of (7) is equivalent to finding the longest path in a layered directed graph (trellis) where there are two nodes in each layer (Fig. 2). (The K-user algorithm in [20] has 2^{K-1} nodes per layer and each node is connected to two nodes in the following layer.) The natural solution to this combinatorial problem is dynamic programming, which can be implemented in real time by running it in forward fashion and with decisions that look ahead over a finite window of stages, rather than waiting for all the data to be received (the Viterbi algorithm). Hence, the structure of the optimum multiuser receiver is now complete: a matched filter front-end followed by a Viterbi algorithm (Fig. 3). Due to the interference between the users, maximum likelihood is not the unique optimality criterion even if the transmitted streams are assumed independent and identically distributed. Minimum probability of error decisions can be accomplished by a backward-forward dynamic programming algorithm [19, 23]; however, the bit-error-rate of the maximum likelihood receiver turns out to be indistinguishable from the minimum achievable bit-error-rate in the region of signal-to-noise ratios of interest, so for most applications it is not worth incurring the additional implementation cost required by the backward-forward algorithm over the Viterbi algorithm.

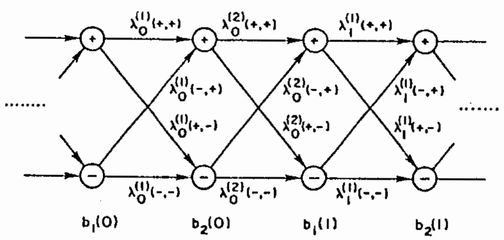

Figure 2. Trellis for two-user asynchronous demodulation.

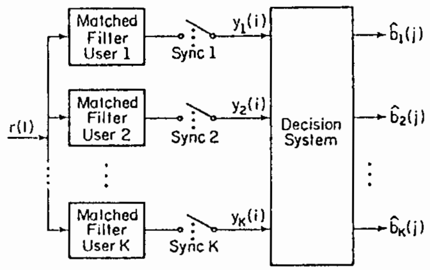

Figure 3. Optimum multiuser detector.

In the special case when the energies of the users are the same and $\rho_{12} = \rho_{21}$ (i.e. H is Toeplitz), the functions $\lambda_i^{(k)}$ in (7) no longer depend on k and i (except through the data $y_k(i)$) and they become identical to the metric functions obtained by Ungerboeck [17] for maximum likelihood sequence detection in the single-user intersymbol interference channel. This is because if the energies and signature waveforms of both users are identical and $\tau_2 - \tau_1 = T/2$, then the multiuser channel (1) is entirely equivalent to an intersymbol interference channel where each symbol overlaps with two symbols. Thus, the decomposition in (7) is akin to a periodically time-varying generalization of Ungerboeck's decomposition [17]. Similar decompositions [21] of the likelihood function are also feasible in asynchronous multiple-access channels which are much less structured than (1).

As we discuss in Section III the optimum detector affords important performance gains over the conventional single-user detector, and, in particular, it solves the near-far problem. However, the price for this is exponential complexity in the number of users, since the combinatorial optimization problem of selecting the most likely transmitted bits given the matched filter outputs is inherently hard [25]. This has spurred several works that propose lower complexity multiuser detectors whose performance lies somewhere in the big gap between optimum and conventional bit-error-rate. A natural alternative to explore is the replacement of the Viterbi algorithm by a sequential decoding algorithm. This is done in [11] where a metric is proposed for a stack sequential decoder based on the additive decomposition of the loglikelihood function in (7). At each point in the search in the binary tree of the sequential decoding algorithm the metric is the sum of two parts: one that the depends on "past" symbols exclusively and the other that depends on "future" unexplored symbols. The metric proposed in [11] takes the decisions of the conventional detector in lieu of those unexplored symbols. As usual, results on the computational complexity of the resulting sequential decoding algorithm are hard since the number of steps depends on the noise realization and there is the possibility of buffer overflow. But it appears that unless the signal-to-noise ratios are extremely poor such a strategy does indeed afford important computational savings over the Viterbi algorithm when the number of users is large.

The approach taken in [5-8] to find low-complexity multiuser detectors that exhibit good performance, and, in particular, are near-far resistant, is to restrict attention to linear transformations of the bank of matched filter outputs. As in the discussion of the maximum likelihood detector, it is beneficial to consider first the synchronous version of the problem. The maximization in (5) is over a four-element set because the receiver is assumed to know the energies of both users. If those energies were unknown and we were interested in obtaining maximum likelihood estimates of the energies in addition to the maximum likelihood decisions on the transmitted streams, then the maximization in (5) would be over the whole plane, and the solution would be $\mathbf{H}^{-1}\mathbf{y}$, i.e., the sign of the components of $\mathbf{H}^{-1}\mathbf{y}$ gives the most likely transmitted bit and their absolute values give the maximum-likelihood amplitude estimates. (Note that better amplitude estimates would be possible with a sequence approach rather than a one-shot approach.) This boils down to using a simple modification of the conventional detector: instead of correlating the received signal with $s_1(t)$ and comparing the output to zero to determine the bit transmitted by user 1, we correlate the received signal with $s_1(t) - \rho\, s_2(t)$. In the hypothetical case when there is no background Gaussian noise ($\sigma^2 = 0$), we have $\mathbf{y} = \mathbf{H}\mathbf{b}$, and the bits selected by this receiver are $sgn(\mathbf{H}^{-1}\mathbf{y}) = sgn(\mathbf{b})$, i.e., the true data, regardless of the values of the received energies. A nice property, which augurs well for the performance of this receiver (at least in the high SNR region). Indeed, it is the lack of this feature that makes the conventional receiver so sensitive to the near-far problem. In the general K-user case, the essence of the $sgn(\mathbf{H}^{-1}\mathbf{y})$ receiver is that instead of correlating with $s_k(t)$ it correlates with the projection of $s_k(t)$ on the subspace orthogonal to the subspace spanned by the other signature waveforms $\{s_j(t),\ j \neq k\}$. Thus, it effectively tunes out the multiuser interference, and hence its name: *decorrelating detector*.

The foregoing derivation of the decorrelating detector has highlighted that if no prior information on the transmitted energies is available, maximum likelihood decisions are obtained by a linear receiver. Note that Schneider [12] sought to minimize error probability in a synchronous multiuser channel with equal-energy signals and, erronously, arrived at the same receiver $sgn(\mathbf{H}^{-1}\mathbf{y})$. At this point, it is natural to ask whether it is possible to achieve better performance by choosing a different linear transformation when the received energies are known. The answer is affirmative; for example, the optimum linear receiver (for high SNR's) for user 1 correlates the received signal with [6]:

$$\begin{cases} s_1(t) - \sqrt{\dfrac{w_2}{w_1}}\, sgn(\rho)\, s_2(t) & \text{if } \dfrac{w_2}{w_1} \leq \rho^2 \\ s_1(t) - \rho\, s_2(t) & \text{otherwise} \end{cases}$$

However, as we discuss in Section III, the decorrelating receiver provides the same degree of near-far resistance as the optimum multiuser receiver [20] (which uses knowledge of the received energies) and it is much easier to compute than the optimum linear receiver in the K-user case.

How do we generalize the decorrelating receiver to the asynchronous case? We can take the same approach as in the derivation of the optimum receiver: each of the $(2M+1)K$ symbols can be viewed as transmitted by a different user. However, the resulting detector would be impossible to compute as it would involve the inversion of a crosscorrelation matrix of dimension $(2M+1)K$. Fortunately, as $M \to \infty$ the rule $sgn(\mathbf{H}^{-1}\mathbf{y})$ is equivalent to passing the sufficient statistics $\mathbf{y}(i) = [y_1(i)\ y_2(i)]^T$ through a two-input two-output discrete-time linear time invariant system with transfer function [7]

$$[\mathrm{H}^T(1)\, z + \mathrm{H}(0) + \mathrm{H}(1)\, z^{-1}]^{-1} = \frac{1}{1 - \rho_{12}^2 - \rho_{21}^2 - \rho_{12}\rho_{21} z - \rho_{12}\rho_{21} z^{-1}} \begin{pmatrix} 1 & -\rho_{12} - \rho_{21} z^{-1} \\ -\rho_{12} - \rho_{21} z & 1 \end{pmatrix} \quad (8)$$

Notice that the matrix in (8) performs the same function as the decorrelating detector in the synchronous case, namely, it substracts from each matched filter output (say, the one corresponding to the i^{th} bit of user 1) the matched filter outputs corresponding to the two overlapping bits of the other user (bits $i-1$ and i of user 2). However, now this is not enough because this operation introduces intersymbol interference between the symbols of user 1 as the matched filter outputs of the overlapping bits ($i-1$ and i of user 2) are contaminated by bits $i-1$ and $i+1$ of user 1. This is the origin of the recursive part of the filter (8) which acts as a linear equalizer for the intersymbol interference introduced by the nonrecursive part. Note that the impulse response of the recursive part is of the form $[\alpha(\rho_{12},\rho_{21})]^{|n|}$, i.e., it is non-causal, and therefore, in practice, it is necessary to approximate it by truncation.

An alternative to the decorrelating detector for asynchronous channels, which is presented here for the first time, can be obtained by taking a one-shot approach where each symbol interval is considered separately. Let us focus attention on bit 0 of user 1, which occupies the interval $[0, T]$ (assuming without loss of generality $\tau_1 = 0$). This bit overlaps with bit -1 of user 2, over the interval $[0, \tau_2]$ and with bit 0 of user 2 over the interval $[\tau_2, T]$. We can view this situation as a 3-user synchronous channel (see Fig. 4) with unit-energy signature waveforms $\tilde{s}_1(t) = s_1(t)$; $\tilde{s}_2(t) = e_2^{-1/2} s_2^L(t)$; $\tilde{s}_3(t) = (1-e_2)^{-1/2}\, s_2^R(t)$, where

$$s_2^L(t) = \begin{cases} s_2(t+T-\tau_2) & 0 \le t \le \tau_2 \\ 0 & \tau_2 \le t \le T \end{cases} \qquad s_2^R(t) = \begin{cases} 0 & 0 \le t \le \tau_2 \\ s_2(t-\tau_2) & \tau_2 \le t \le T \end{cases}$$

and $e_2 = \int_0^{\tau_2} s_2^2(t+T-\tau_2)dt$ $(0 < e_2 < 1)$ is the partial energy of the interfering signal over the left overlapping interval. We can now solve for a 3-user decorrelating detector with

$$\mathrm{H} = \begin{bmatrix} 1 & e_2^{-1/2}\rho_{12} & (1-e_2)^{-1/2}\rho_{21} \\ e_2^{-1/2}\rho_{12} & 1 & 0 \\ (1-e_2)^{-1/2}\rho_{21} & 0 & 1 \end{bmatrix}, \quad (9)$$

resulting in a detector for user 1 that correlates the received signal with

$$[s_1(t) - \frac{\rho_{12}}{e_2}\, s_2^L(t) - \frac{\rho_{21}}{1-e_2}\, s_2^R(t)].$$

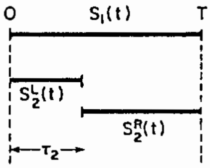

Figure 4. Intervals for one-shot decorrelating detector.

This one-shot decorrelating detector has lower complexity than the asynchronous decorrelating detector described before at the expense of some performance degradation.

Another approach to obtain suboptimum low-complexity multiuser signal detectors is proposed in [18] by Varanasi and Aazhang. Again, it is best to consider first the synchronous case. If we had an estimate, say $\hat{b}_2$, of the symbol transmitted by user 2 in the interval $[0,T]$, then we could subtract $\hat{b}_2\, s_2(t)$ from the received signal and process the remaining signal as if $b_2 = \hat{b}_2$, i.e., correlate it with $s_1(t)$ and find the polarity of the output. The detector proposed in [18] is assumed to know the received energies, and the tentative decisions are obtained with the conventional single-user detector, i.e., $\hat{b}_2$ is $+\sqrt{w_2}$ or $-\sqrt{w_2}$ depending on the polarity of the output of the matched filter of the signature waveform of user 2 (Fig. 5). If the signal of user 2 is comparatively strong, then the conventional receiver will select $\hat{b}_2 = b_2$ with high probability, in which case the decision on the bit transmitted by user 1 will noticeably improve the (poor) decision made by the conventional receiver. On the other hand, the decision on the bit of user 2

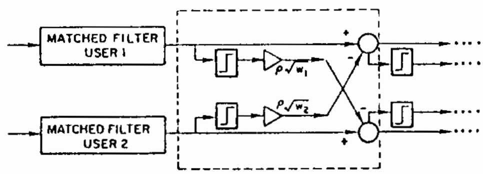

Figure 5. Multi-stage detector for two synchronous users.

will not improve the one obtained with the conventional detector since $b_1 = \hat{b}_1$ does not occur with high enough probability. However, we can now repeat the same procedure as many times as desired using the tentative decisions of the previous stage in lieu of those obtained by the conventional detector. For example, at the output of the second stage, the decisions on user 2 will be near-optimal because the available estimate of the bit of user 1 is excellent, whereas the decisions on user 1 will not improve those obtained in the first stage. Roughly speaking, the decisions put out by even [resp. odd]-numbered stages will be good for the powerful [resp. weak]-transmitters. When the energies are similar, the prerequisite for adequate operation is that the crosscorrelations and noise level be such that the initial decisions put out by the conventional detector have moderate bit-error-rate. In the asynchronous case, the principle of operation is very similar except that now we need to subtract two terms corresponding to the two symbols that overlap with each symbol. Therefore, at the output of each stage we need to store a sliding window of decisions. For example, if there is only one stage we need to store two consecutive outputs of each conventional single-user detector, whereas if there are two stages, we need two consecutive outputs of the first stage and hence, a window of three consecutive conventional decisions at the beginning. In general, if there are m stages, the k^{th} stage needs to store $m - k + 1$ consecutive outputs for each user. As in the case of the one-shot decorrelating detector, this detector can be simplified using (partial) correlations with $s_2^L(t)$ and $s_2^R(t)$.

We now turn our attention to the derivation of demodulators for decentralized (multipoint-to-multipoint) topologies, where each receiver is interested in demodulating the information transmitted by only one of the active users. Obviously, all the foregoing detectors can still be used provided the signature waveforms of the interferers are known and provided that synchronism with each of the transmitted streams is acquired. The decorrelating detector is especially suited for this, as we can implement the linear transformation dictated by each row of H^{-1} or $[\mathrm{H}^T(1)\, z + \mathrm{H}(0) + \mathrm{H}(1)\, z^{-1}]^{-1}$ separately, thus resulting in a simple decentralized version of the detector. The rest of the detectors discussed above do not lend themselves to such decentralized implementations, but in many cases it is possible to lower their complexity substantially, without compromising performance, by ignoring the comparatively weak interferers. In any case, it is important to investigate which detection structures result when the signature waveforms of the interferers are not known or when the receiver acquires synchronization only with the user of interest. The problem becomes one of single-user signal detection in additive colored non-Gaussian noise, since the equivalent noise seen by each signal is the sum of the multiple-access interference and the additive background Gaussian noise.

The solution to this signal detection problem is found in [10] and [19, Chap. 5] both in the case where the interfering signature waveforms are known (but no synchronization is available) and in the case where the modulation format is Direct Sequence Spread Spectrum and the interfering signature sequences are not known. Even though the unavailability of the timing epochs of the interfering users results in very complex nonlinear detectors, it is shown in [10] that important reductions in complexity are possible in several asymptotic cases such as when the number of chips per symbol is large and the background Gaussian noise level is low. Also, when the interfering users are weak the locally optimum detector reduces to a modification of the conventional single-user detector, whereby the received signal is correlated with a smoothed replica of the signature waveform of the user of interest. This is because for the purposes of locally optimum detection it is necessary to take care only of the nonwhiteness of the multiple-access noise.

III. Performance Analysis

We focus attention now on the bit-error-rate achieved by the detectors reviewed in Section II. We denote by $P_k(\sigma)$ the bit-error-rate of the k^{th} user when the white Gaussian noise level is σ^2. Of particular interest is the bit-error-rate in the region of low σ^2, because that quantifies the inherent performance degradation due to the presence of other users in the channel. A convenient way to quantify bit-error-rate in the region of high signal to noise ratios is the *asymptotic efficiency* which is defined as follows [20, 22]. Define the *effective energy* of user k, $e_k(\sigma)$, as the energy that that user would require to achieve bit-error-rate $P_k(\sigma)$ in the same Gaussian channel but without interfering users, i.e., $P_k(\sigma) = Q(e_k^{1/2}(\sigma)/\sigma)$. Thus, the *efficiency* or ratio between the effective and actual energies $e_k(\sigma)/w_k$ is an alternative way to characterize $P_k(\sigma)$. The *asymptotic efficiency* is defined as $\eta_k = \lim_{\sigma \to 0} e_k(\sigma)/w_k$ and measures the slope with which $P_k(\sigma)$ goes to 0 in the high signal-to-noise ratio region. A third useful performance measure is the *near-far resistance*, which is defined as the minimum asymptotic efficiency over the relative energies of all the other users. This measure quantifies the robustness of a receiver against the near-far problem, which as we argued in the introduction is the major shortcoming of coherent CDMA systems.

The minimum probability of error achievable in a Gaussian asynchronous multiple-access channel was obtained in [20]. Prior to the appearance of that paper, works devoted to the analysis of error probability in multiuser channels had focused attention exclusively on the bit-error-rate of the conventional single-user receiver. For fixed, arbitrary offset between the signals, it is easy to see that the two-user asynchronous error probability of a conventional receiver for user 1 is

$$\begin{aligned} P_1^c(\sigma) = {} & \frac{1}{4}\, Q\left((\sqrt{w_1} + \sqrt{w_2}\,\rho_{12} + \sqrt{w_2}\,\rho_{21})/\sigma^2\right) + \frac{1}{4}\, Q\left((\sqrt{w_1} + \sqrt{w_2}\,\rho_{12} - \sqrt{w_2}\,\rho_{21})/\sigma^2\right) \\ & + \frac{1}{4}\, Q\left((\sqrt{w_1} - \sqrt{w_2}\,\rho_{12} + \sqrt{w_2}\,\rho_{21})/\sigma^2\right) + \frac{1}{4}\, Q\left((\sqrt{w_1} - \sqrt{w_2}\,\rho_{12} - \sqrt{w_2}\,\rho_{21})/\sigma^2\right) \end{aligned} \tag{10}$$

In the region of low background Gaussian noise the expression in (10) is dominated by the Q-function with the smallest argument. Hence, according to the foregoing definition, the asymptotic efficiency is

$$\eta_1^c = \max{}^2\{0,\ 1 - \sqrt{\frac{w_2}{w_1}}\,(\,|\rho_{12}| + |\rho_{21}|)\} \tag{11}$$

If the interferer is very weak $w_2 << w_1$, then, η_1^c is close to unity (the conventional receiver is quasi-optimal because very little penalty is paid for ignoring user 2). As w_2/w_1 grows, η_1^c decreases monotonically, until it becomes zero for

$$\frac{w_2}{w_1} \geq \frac{1}{[\,|\rho_{12}| + |\rho_{21}|]^2},$$

in which case, the probability of error is worse than 0.25 regardless of the background noise level. Therefore, the near-far resistance is equal to zero and there is always an interference level that renders the conventional receiver useless no matter how good the signal design.

The minimum probability of error derived in [20] exhibits a very different behavior. The main contribution of that paper was the introduction of the so-called *method of indecomposable sequences* to obtain upper bounds on the probability of error of m-ary hypothesis testing problems in terms of the error probability of binary problems. This method has been successfully used in a wide variety of channels: both single-user [24] and multiple-access channels with both synchronous and asynchronous users, and with both Gaussian and Poisson [21] observation models. Space limitations prevent us from an exposition of the method of indecomposable sequences and its application to code division multiple-access channels. The main conclusion from that analysis is that the probability of error of the maximum likelihood receiver is optimum in the region of moderate to large signal-to-noise ratios and it behaves as that of a single transmitter with reduced energy. This reduction in effective energy is quantified by the optimum asymptotic efficiency [22]

$$\eta_1 = 1 - \frac{\sqrt{w_2}}{\sqrt{w_1}} \left[\left(2\,|\rho_{12}| - \frac{\sqrt{w_2}}{\sqrt{w_1}} \right)^+ + \left(2\,|\rho_{21}| - \frac{\sqrt{w_2}}{\sqrt{w_1}} \right)^+ \right] \tag{12}$$

As in the case of the conventional receiver, η_1 is close to unity if $w_2/w_1 << 1$. However, unlike (11), η_1 is not monotonic in w_2/w_1. Actually, if

$$\frac{w_2}{w_1} \geq 4\,\max\{\rho_{12}^2, \rho_{21}^2\}, \tag{13}$$

then $\eta_1 = 1$. Therefore, as long as the energy of user 2 exceeds the threshold given by (13) the bit-error-rate of user 1 is equivalent to the single-user case where user 2 is not active. The explanation of this behavior of the optimum receiver is that if the interfering user is sufficiently powerful, then the primary source of errors committed in the optimum demodulation of user 1 is the background Gaussian noise, rather than the randomness of the information carried by the interfering signal. Note that according to the threshold in (13), an interferer who is 3 dB weaker than the user of interest has no appreciable effect on the bit-error-rate as long as the maximum crosscorrelation is below 0.35 (a mild condition on the signal design). Interestingly, the same is true if the relative energy of the interferer is *higher* than -3 dB. The minimization of (12) with respect to w_2/w_1 yields

$$1 - \max^2\{\,|\rho_{12}|,\ |\rho_{21}|,\ (\,|\rho_{12}| + |\rho_{21}|)/\sqrt{2}\},$$

which implies that there is a nonzero level of guaranteed asymptotic efficiency regardless of the received energies (unless both signature signatures are identical and there is no offset between them). A more stringent measure of near-far resistance is obtained by minimizing the asymptotic efficiency over the energies of all the users letting those energies be time-varying; the rationale for this being that the near-far problem occurs primarily in networks with dynamically changing topologies. In such case, the near-far resistance is given by

$$\overline{\eta}_1 = \sqrt{[1 - (\rho_{12} + \rho_{21})^2][1 - (\rho_{12} - \rho_{21})^2]}. \tag{14}$$

In the general K-user case, it can be shown that $\overline{\eta}_1$ is equal to the smallest energy of the signals in Ξ_1, which is the set of multiuser signals such that $b_1(0) = 1$, i.e., $\Xi_1 = \{\sum_i \sum_{k=1}^{K} b_k(i) s_k(t - iT - \tau_k)$, $b_1(0) = 1,\ b_k(i) \in R,\ (k,i) \neq (1,0)\}$. In the synchronous case, $\overline{\eta}_k$ can be shown [6] to equal $1/(\mathbf{H}^{-1})_{kk}$, whereas in the asynchronous case [7] we have

$$\overline{\eta}_k = \left[\frac{1}{\pi} \int_0^{\pi} [\mathbf{H}(1)^T\, e^{j\omega} + \mathbf{H}(0) + \mathbf{H}(1)\, e^{-j\omega}]_{kk}^{-1}\, d\omega \right]^{-1}.$$

The decorrelating detector eliminates the multiuser interference from the decision statistic. Therefore, its bit-error-rate has the very desirable property that it is independent of the energy of the interfering users. It can be shown that it is equal to (in the K-user case) [7]:

$$P_k^d(\omega) = Q\left(\frac{\sqrt{w_k \overline{\eta}_k}}{\sigma} \right) \tag{15}$$

which implies that the effective energy is $e_k(\omega) = w_k \overline{\eta}_k$, and therefore, the efficiency is independent of the noise level as well as of the energies and is equal to the near-far resistance of the optimum receiver. Hence, we arrive at the interesting conclusion that the decorrelating detector achieves optimum near-far resistance. Figure 6 illustrates the near-far behavior of the conventional, optimum and decorrelating detectors, comparing their asymptotic efficiencies as a function of the relative energies. Note that there is a gap between the minimum efficiency achieved by the optimum receiver and the efficiency of the conventional detector. This is because the asymptotic efficiency of the optimum detector is the function in (12) where the energies are assumed time-invariant. The minimum of the optimum asymptotic efficiency over the amplitudes of each of the symbols does indeed coincide with the decorrelating asymptotic efficiency.

The bit-error-rate of the one-shot decorrelating detector for asynchronous channels introduced in Section II is also independent of the energy of the interferers. Its efficiency (and near-far resistance) is given by the inverse of the first diagonal element of the inverse of (9), i.e.,

$$\eta_1^{od} = 1 - \frac{\rho_{12}^2}{e_2} - \frac{\rho_{21}^2}{1 - e_2}.$$

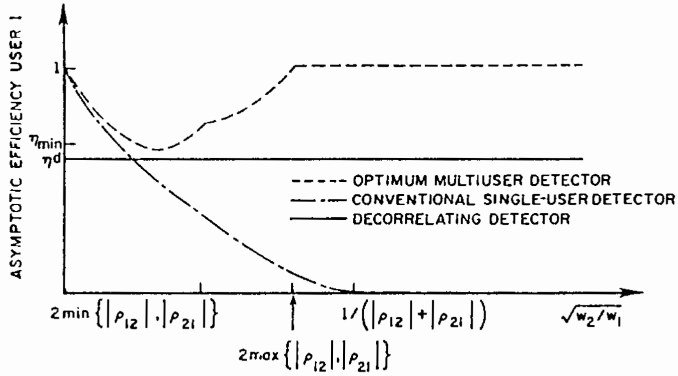

Figure 6. Asymptotic efficiencies for $\rho_{12} = 0.3$, $\rho_{21} = 0.5$. $\eta_{min} = 0.68$ and $\eta^d = 0.59$.

The performance analysis of the remaining detectors reviewed in Section II does not lend itself to similar analytical results due to the nonlinear nature of those detectors. A numerical computation of the receiver proposed in [18] shows that in the two-user case a single stage detector achieves noticeable improvements over the conventional receiver especially when the interferer is very strong and the signature waveforms have good crosscorrelation properties. No bit-error-rate analysis of the nonlinear single-user detectors obtained in [10] has been undertaken to date. Nevertheless, [10] shows the interesting result that in the absence of background Gaussian noise the transmitted bits can be demodulated perfectly even if the signature waveforms of the interferers are not known to the receiver, thus hinting that some degree of robustness to the near-far effect is also possible in that case. However, no practically implementable receiver that ignores the interfering signature waveforms is known to be near-far resistant.

References

1. W. Van Etten, "An Optimum Linear Receiver for Multiple Channel Digital Transmission Systems," *IEEE Trans. on Communications*, pp. 828-834, Aug 1975.

2. W. Van Etten, "Maximum Likelihood Receiver for Multiple Channel Transmission Systems," *IEEE Trans. Commun.*, vol. COM-24, no. 2, pp. 276-283, February 1976.

3. D. Horwood and R. Gagliardi, "Signal Design for Digital Multiple Access Communications," *IEEE Trans. Commun.*, vol. COM-23, pp. 378-383, March 1975.

4. A.R. Kaye and D.A. George, "Transmission of multiplexed PAM signals over multiple channel and diversity systems," *IEEE Trans. on Communication Technology*, vol. COM-18, pp. 520-525, Oct 1970.

5. R. Lupas-Golaszewski and S. Verdu, "Asymptotic Efficiency of Linear Multiuser Detectors," *Proc. 25th Conf. on Decision and Control*, pp. 2094-2100, Athens, Dec. 1986.

6. R. Lupas and S. Verdu, "Linear Multiuser Detectors for Synchronous Code-Divsion Multiple-Access Channels," *IEEE Trans. Information Theory*, vol. IT-34. to appear.

7. R. Lupas and S. Verdu, "Near-Far Resistance of Multiuser Detectors in Asynchronous Channels," *IEEE Trans. Communications*, vol. COM-37. to appear

8. R. Lupas and S. Verdu, "Optimum Near-Far Resistance of Linear Detectors for Code-Division Multiple-Access Channels," *Abs. 1988 IEEE Int. Symp. Information Theory*, p. 14, Kobe, Japan, June 1988.

9. H.V. Poor, "Signal detection in multiple access channels," U.S. Army Research Office Proposal (Contract DAAG29-81-K-0062 to Coordinated Science Laboratory, University of Illinois),, 1980.

10. H. V. Poor and S. Verdu, "Single-User Detectors for Multiuser Channels," *IEEE Trans. Communications*, vol. 36, pp. 50-60, Jan. 1988.

11. C. Rushforth and Z. Xie, "Multiuser Signal Detection using Sequential Decoding," *1988 IEEE Military Communications Conf.*, San Diego, Oct. 1988.

12. K.S. Schneider, "Optimum Detection of Code Division Multiplexed Signals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-15, no. 1, pp. 181-185, January 1979.

13. K.S. Schneider, "Crosstalk Resistant Receiver for M-ary Multiplexed Communications," *IEEE Trans. on Aerospace and Electronic Systems*, vol. AES-16, pp. 426-433, Jul 1980.

14. D.A. Shnidman, "A Generalized Nyquist Criterion and an Optimum Linear Receiver for a Pulse Modulation System," *The Bell System Technical Journal*, pp. 2163-2177, Nov 1967.

15. U. Timor, "Improved Decoding Scheme for Frequency-Hopped Multilevel FSK System," *The Bell System Technical Journal*, vol. 59, pp. 1839-1855, December 1980.

16. U. Timor, "Multistage Decoding of Frequency-Hopped FSK System," *The Bell System Technical Journal*, vol. 60, pp. 471-483, April 1981.

17. G. Ungerboeck, "Adaptive Maximum Likelihood Receiver for Carrier-Modulated Data Transmission Systems," *IEEE Trans. Commun.*, vol. COM-22, no. 5, pp. 624-636, May 1974.

18. M. K. Varanasi and B. Aazhang, "Near Optimum demodulation for Coherent Communications in Asynchronous Gaussian CDMA Channels," *Proc. 22nd Princeton Conf. on Information Sciences and Systems*, pp. 832-839, Mar. 1988.

19. S. Verdu, "Optimum multi-user signal detection," Ph. D. Dissertation, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign. Report T-151 Coordinated Science Laboratory, Urbana, IL, Aug. 1984.

20. S. Verdu, "Minimum Probability of Error for Asynchronous Gaussian Multiple-Access Channels," *IEEE Trans. on Information Theory*, vol. IT-32, pp. 85-96, Jan. 1986.

21. S. Verdu, "Multiple-Access Channels with point-process observations: Optimum demodulation," *IEEE Trans. Information Theory*, vol. IT-32, pp. 642-651, September 1986.

22. S. Verdu, "Optimum multiuser asymptotic efficiency," *IEEE Trans. Communications*, vol. COM-34, pp. 890-897, Sept. 1986.

23. S. Verdu and H. Vincent Poor, "Abstract Dynamic Programming Models under Commutativity Conditions," *SIAM J. Control and Optimization*, vol. 24, pp. 990-1006, Jul. 1987.

24. S. Verdu, "Maximum Likelihood Sequence Detection for Intersymbol Interference Channels: A New Upper Bound on Error Probability," *IEEE Trans. Information Theory*, vol. IT-33, pp. 62-68, Jan. 1987.

25. S. Verdu, "Computational Complexity of Optimum Multiuser Detection," *Algorithmica*, vol. 4, to appear.