# 1

# OVERVIEW OF MULTIUSER DETECTION

Michael L. Honig

... But what is the use of counterpoint when, if played, one imagines that four different orchestras are playing at the same time four different tunes in four different keys and measures? A veritable nightmare!

—Music critic Arthur Bird, writing about R. Strauss's tone poem *Ein Heldenleben*, May 1, 1899 [69, p. 186]



## 1.1 INTRODUCTION

One of the key challenges in designing multiuser communications systems is mitigating interference. This challenge is apparent for modern wireless networks, such as mobile cellular, and wireless local and metropolitan area networks, where achieving high spectral efficiencies requires aggressive frequency reuse. Interference also limits the performance of many wired channels, such as the digital subscriber line (DSL). Although each DSL is typically associated with a single user, capacitive coupling between pairs of DSLs in close physical proximity causes cross-talk interference, which degrades performance (e.g., see [72]).

Interference encompasses *self*-interference, due to reflections of the same signal, in addition to *multi-user*-interference associated with other signals sharing the same bandwidth. Self-interference arises from multipath in a wireless channel, a bridge tap in a Digital Subscriber Line (DSL), and bandwidth constraints, which cause

*Advances in Multiuser Detection*. Edited by Michael L. Honig

intersymbol interference. Throughput this book, interference typically refers to signals from other users (or more generally, data streams) associated with the *same* system. Clearly, techniques for reducing, or mitigating interference lead directly to improved performance, either in terms of reduced error rate, increased data rate, or number of users that can be served.

Effective techniques for mitigating interference must depend, of course, on what is known at the transmitter and receiver. For example, in a multiple-access channel the receivers for all transmitters are co-located. Hence instead of detecting the transmitted bits from a particular user in isolation, treating the other signals as background noise, the receiver can *jointly* detect *all* of the transmitted information bits, or symbols. In that case, the structure of the multiple access interference can be used to aid the detection of the desired symbol. For example, it may be possible to use estimates of interfering symbols to cancel, or at least reduce, the level of the interfering signals. In contrast, if the receivers are not co-located, then interference cancellation, which relies on accurate interference estimates, may not be practical.

*Multiuser detection* refers to the scenario in which a single receiver jointly detects multiple simultaneous transmissions. Examples include the uplink of a single cell in a cellular system, and a group of twisted-pair copper subscriber lines that terminate at the same central switching office. More generally, multiuser detection techniques apply to the joint detection of different signals transmitted over any multi-input/multi-output (MIMO) channel. In addition to the preceding examples, others include channels in which multiple transmitted information streams are multiplexed over multiple transmit antennas. In that scenario, the multiple "users" refer to the multiple information streams, even though the transmitted signal may originate from a single user.

Closely related to multiuser detection is *interference suppression*. The key distinction is that a multiuser detector attempts to retrieve *multiple* (i.e., at least two) transmitted signals, or information streams, whereas interference suppression implies that the receiver is interested in only one signal among the received superposition of transmitted signals. For example, this is typically the case for the downlink of a cellular system, in which a mobile wishes to demodulate a single transmitted information stream in the presence of interfering signals from the associated base station and from nearby base stations. Of course, in general a receiver may wish to demodulate a subset of two or more signals from among a larger mix of signals. In that case, the receiver jointly detects the subset of desired signals while suppressing the interfering signals. An example of this is the uplink of a cellular system, in which the receiver is interested in demodulating the transmitted signals from users within the cell in the presence of interference from other cells.

### 1.1.1 Applications

Much of the work on multiuser detection and interference suppression over the past couple of decades has been motivated by the commercial success of Code-Division Multiple Access (CDMA) in mobile cellular systems. Namely, second-generation CDMA cellular systems were introduced in the early 1990s, and CDMA is currently used in third-generation cellular systems. The performance of CDMA is generally limited by multi-user interference. In particular, the performance of the uplink, which refers to the multiple access channel from users to base station, is sensitive to received power variations across users. The classic example of this is the near–far problem, in which a user close to the base station causes excessive interference to a user far from the base station. Commercial CDMA systems generally use closed-loop power control to minimize received power variations (both across time to mitigate fading and across users). It was recognized early on, however, that the sensitivity to interference power is not inherent to CDMA, but rather is a property of the conventional matched filter detector. That motivated studies of multiuser detection techniques for uplink CDMA, starting with the work of Verdú [86].[1]

Prior to the introduction of CDMA for mobile cellular systems, some specific multiuser detectors had been derived for linear MIMO channels. See, for example, [80], which discusses the Maximum Likelihood (ML) detector, and [1,45,46,63], which derive linear and decision feedback detectors.[2]

In recent years, the primary motivation for multiuser detection has shifted from CDMA to other applications (in particular, links with multiple antennas). There are several reasons for this. First, as shown in the initial studies on the achievable rates for wireless links with multiple antennas [19,77], substantial gains in spectral efficiency and reliability can be achieved by adding antennas to the transmitter and/or receiver of a single-user wireless channel. That stimulated an enormous amount of activity on coding and reception techniques for MIMO channels, so that nearly all evolving wireless systems and standards include provisions for multiple transmit antennas. In those scenarios, multiuser detection techniques are useful for mitigating interference among the different transmit antennas [6].

The second reason for the shift away from CDMA applications is that current designs for next-generation cellular and wide-area wireless networks are based primarily on Orthogonal Frequency Division Multiplexing (OFDM) and Orthogonal Frequency Division Multiple Access (OFDMA). That is due in part to the more substantial role played by data services, as opposed to voice, in evolving wireless networks. Namely, current CDMA cellular systems rely on "interference averaging" and power control to achieve robustness with respect to variations in the active user set and associated channels. That requires a relatively large number of low-rate (e.g., voice) users. In contrast, data traffic associated with internet services is bursty, and depends on the rates provided. Higher rates enable shorter bursts, but typically generate more interference due to higher transmit power. Hence without further coordination among users, the interference seen at a base station or mobile is likely to vary substantially as high-rate users enter and leave, degrading performance.[3]

[1] Although [86] stimulated much of the subsequent work on multiuser detection, the structure of the optimal detector for CDMA had been previously derived in [67].

[2] Although that work predates the development of mobile cellular and DSL, similar types of "multi-terminal" applications are mentioned, including multi-cable and diversity channels. The MIMO models in [1,45] were originally motivated by dual polarization radio transmission with crosstalk.

[3] In principle, fluctuations in interference in a CDMA system can be mitigated through the application of multiuser detection, although the challenges subsequently discussed still apply.

To mitigate interference associated with bursty users and high data rates, transmissions can be scheduled over different time slots (scheduling intervals). Because data is delay-tolerant, that also allows the scheduler to select users with favorable channel conditions, thereby exploiting multiuser diversity [3,5,49]. Hence current (third-generation) cellular data systems typically rely on scheduling, as opposed to spreading (as in CDMA), to reduce interference. Emerging systems based on OFDMA provide further flexibility in allocating both time *and* frequency resources among requests to mitigate interference and exploit multiuser diversity. Interference mitigation techniques are potentially useful for OFDMA, although the interference originates from other cells (or sectors), so that associated channel conditions may be more difficult to estimate.

### 1.1.2 Mobile Cellular Challenges

Cellular systems pose several challenges that prevent a straightforward application of most multiuser detection techniques. This is especially true for uplink asynchronous CDMA with full spreading, or frequency reuse. The main difficulty is that for interference averaging, it is desirable to extend spreading sequences over many symbols, so that interference from a particular user is averaged over many (randomly chosen) signatures. In that way, the performance is not limited by the possibility of choosing a particular set of signatures with undesirable correlation properties. These "long" signatures, however, greatly complicate the implementation of a multiuser detector, which exploits properties of the particular set of assigned signatures for each symbol. That is, the structure of the multiuser detector must change from symbol to symbol, which may require excessive computation. Long signatures also prevent the application of standard adaptive filtering methods (used to equalize single-user channels), which require linear modulation with short signatures (i.e., the signature for a particular user is repeated from symbol to symbol).[4]

Another challenge for multiuser detection posed by mobile cellular systems is fading. The multiuser detector must track or adapt to channel variations in time and frequency caused by mobility. That includes channel variations associated with the interferers as well as the desired user. This may be feasible when the channel variations are slowly varying (e.g., over a few hundred symbols), or when the number of interferers is relatively small, so that the number of channel parameters to track is manageable. However, in a CDMA cellular system there may be a large number of mobile users, several of which experience rapid fading. The inability to track all of these channel variations can significantly compromise performance. Furthermore, the complexity of the multiuser detector generally increases with the system size (number of users and processing gain). For this reason, application of multiuser detection to uplink CDMA has been mostly limited to relatively simple interference cancellation techniques.

[4] At a basic level, one might argue that interference averaging as a design objective for cellular systems runs contrary to a design based on multiuser detection. Namely, interference averaging is associated with a large number of weak interferers, whereas multiuser detection becomes most attractive when there are a small number of strong interferers.

One way to address the preceding challenges is to redesign the CDMA system with multiuser detection in mind. In addition to using short signatures, that generally means reducing the number of users, or transmitted information streams, which are jointly detected at the base station, and also slowing down the channel variations (fade rate). Reducing the number of users for joint detection can be accomplished by sub-dividing the channel resources in time or frequency. For example, the Time-Division Duplex (TDD) version of UMTS[5] is based on a combination of Time-Division Multiple Access (TDMA) and CDMA [30]. The uplink channel is divided into 10 msec frames with 15 time slots, and multiple users can be assigned to a particular time slot through the use of direct-sequence spread spectrum signaling (i.e., direct-sequence CDMA). For example, if the original CDMA system supports 75 users in a cell, then introducing 15 time slots means that the users can be divided into 15 groups, each containing five users, which are assigned to the different time slots. That reduces the required spreading (processing gain) by approximately a factor of 15, and eases the burden on a multiuser detector. Furthermore, the introduction of time slots increases the symbol rate, and hence decreases the fade rate (rate of channel variations normalized by the symbol rate). That simplifies the associated channel estimation.

As previously discussed, the desire to integrate voice services with different data services having variable Quality of Service (e.g., delay) requirements has motivated the trend towards OFDM and OFDMA. The channel is then divided into both time and frequency slots, and each slot is designated for a *single* user. Hence in general the transmitted signals are not spread. That obviates the need for multiuser detection on the uplink, although interference from other cells still limits performance. Hence interference suppression techniques may still be beneficial. Nevertheless, multiuser detection and interference suppression for OFDM systems are primarily focused on mitigating interference among multiple antennas, which creates MIMO channels across the sub-carriers.

### 1.1.3 Chapter Outline

In the next section we discuss the linear (matrix) channel model, which is the basis for the multiuser detection methods discussed in this book. We then give a brief overview of a few different multiuser detectors along with performance comparisons for a simple version of this channel model in which the mixing matrix has independent, identically distributed (*i.i.d.*) elements. This discussion is meant to provide some background and motivation for the topics discussed in subsequent chapters. We also discuss and compare some additional multiuser detection techniques. A comprehensive treatment of some of the detectors presented here (e.g., optimal and linear) is given in the book by Verdú [87].

[5] Universal Mobile Telecommunications System is an air interface standard for third generation cellular systems. The standard was approved in 1998, and consists of two modes corresponding to Frequency-Division Duplex (FDD) and TDD operation [30,57].

## 1.2 MATRIX CHANNEL MODEL

In its simplest and most general form the matrix channel model is given by:

$$\mathbf{y}(i) = \mathbf{M}(i)\mathbf{b}(i) + \mathbf{n}(i) \tag{1.1}$$

where $\mathbf{y}$ is an $N \times 1$ vector of received samples, $\mathbf{b}$ is a $K \times 1$ vector of transmitted symbols, $\mathbf{M}$ is an $N \times K$ *mixing* matrix, and $\mathbf{n}$ is an $N \times 1$ vector of additive noise samples. Also, $i$ is the discrete time index. Hence this discrete-time model includes the combination of any filtering (analog or digital) at the transmitters and receivers. Assuming quadrature modulation, all of the variables in (1.1) are complex valued.

For synchronous CDMA the $k$th entry of $\mathbf{b}(i)$, denoted $b_k(i)$, is the $i$th symbol transmitted by user $k$, and the $k$th column of $\mathbf{M}$ is the signature for user $k$. In that case the elements of $\mathbf{M}$ are typically chosen from a discrete set, e.g., $\pm 1$ for real codes, or $\pm 1 \pm j$ for complex signatures. In practice, the signature elements are often chosen randomly, although it is also desirable if the signatures (columns of $\mathbf{M}$) are orthogonal. To represent different powers across the users, we can write $\mathbf{M} = \mathbf{SA}$, where $\mathbf{S}$ is the signature matrix, and $\mathbf{A}$ is a diagonal matrix of amplitudes.

For asynchronous CDMA, the model (1.1) still applies, but with different interpretations for $\mathbf{b}(i)$ and $\mathbf{M}$. Namely, assuming each signature spans a single symbol interval, we can write:

$$\mathbf{y}(i) = \mathbf{M}_1\mathbf{b}(i-1) + \mathbf{M}_2\mathbf{b}(i) + \mathbf{M}_3\mathbf{b}(i+1) + \mathbf{n} = \mathbf{M}\underline{\mathbf{b}}(i) + \mathbf{n} \tag{1.2}$$

where the $k$th columns of $\mathbf{M}_1$, $\mathbf{M}_2$, and $\mathbf{M}_3$ are the corresponding segments of signatures (padded with zeros) associated with $b_k(i-1)$, $b_k(i)$, and $b_k(i+1)$, respectively, $\underline{\mathbf{b}}(i)$ contains $\mathbf{b}(i-1)$, $\mathbf{b}(i)$, and $\mathbf{b}(i+1)$ stacked on top of each other, and $\mathbf{M}$ is the corresponding block-diagonal matrix.

For a single-user link with $K$ transmit antennas and $N$ receive antennas, if $\mathbf{b}(i)$ contains $K$ symbols (i.e., is $K \times 1$), then $b_k$ in (1.1) is transmitted by the $k$th antenna, and the $(n, k)$th entry of $\mathbf{M}$ is the complex channel gain from the $k$th transmit antenna to the $n$th receive antenna. More generally, the number of transmitted symbols $K'$ [dimension of $\mathbf{b}(i)$] may be less than the number of transmit antennas $K$, in which case an additional $K \times K'$ *precoding matrix* $\mathbf{V}$ is needed to map the symbols to antennas. That is, $\mathbf{M} = \mathbf{HV}$, where $\mathbf{H}$ is the $N \times K$ channel matrix. The channel gains are often modeled as *i.i.d.* random variables (typically complex Gaussian corresponding to flat Rayleigh fading). Other statistical models for the matrix $\mathbf{M}$, corresponding to correlated fading across antennas, are discussed in Chapter 6.

The *interference channel* can also be modeled by (1.1), where $\mathbf{y}$ corresponds to the signal at a particular receiver, which estimates a *subset* of $\mathbf{b}$. The remaining symbols are presumably transmitted to other receivers at different locations, and are treated as interference. That applies, for example, to the uplink where the estimated symbols are transmitted by users within the cell and the interfering symbols are transmitted by users in other cells.

The model (1.1) also applies to a single-input/single-output channel with intersymbol interference (ISI). Namely, suppose a single transmitter transmits the sequence of symbols $\{b(i)\}$ through a linear, dispersive channel with impulse response:

$$h(-m),\, h(-m+1), \ldots, h(0),\, h(1), \ldots, h(m),$$

where the length of the impulse response is assumed to be $2m+1$, so that the output:

$$y(i) = \sum_{k=-m}^{m} h(k)b(i-k) \tag{1.3}$$

We can represent this in the form (1.1) by defining the vector of $N = 2n + 1$ received samples, corresponding to time $i$, as:

$$\mathbf{y}(i) = [\, y(i+n), \ldots, y(i), \ldots, y(i-n)]^T. \tag{1.4}$$

where $[\cdot]^T$ denotes transpose, and the vector of $K = 2(n+m) + 1$ transmitted symbols as $\mathbf{b}(i) = [b(i+n+m), b(i+n+m-1), \ldots, b(i-n-m)]^T$. The channel matrix $\mathbf{M}$ is then the $N \times K$ Toeplitz matrix:

$$\mathbf{M} = \begin{bmatrix} h(-m) & \ldots & h(m) & 0 & \ldots & 0 \\ 0 & h(-m) & \ldots & h(m) & \ldots & 0 \\ & & \ddots & & \ddots & \\ 0 & \ldots & 0 & h(-m) & \ldots & h(m) \end{bmatrix}. \tag{1.5}$$

In practice, the receiver would presumably use $\mathbf{y}(i)$ to detect the subset of symbols $b(i+n-m), \ldots, b(i-n+m)$, which forms the reduced symbol vector $\bar{\mathbf{b}}(i)$. That is, the receiver estimates only the subset of transmitted symbols, which correspond to columns of $\mathbf{M}$ containing the *entire* impulse response. (That maximizes the received energy associated with each detected symbol.) Assuming that symbols are transmitted continuously, the trailing $2m$ symbols of $\mathbf{b}$, namely, $b(i-n+m-1), \ldots, b(i-n-m)$, therefore experience intersymbol interference from the preceding vector $\mathbf{b}(i-1)$, and the leading $2m$ symbols $b(i+n+m), \ldots, b(i+n-m+1)$ interfere with $\mathbf{b}(i+1)$.

To model an OFDM system, the channel matrix $\mathbf{M}$ is first converted to a *circulant* matrix by setting the first $m$ symbols of $\mathbf{b}$ equal to the trailing $m$ symbols of $\bar{\mathbf{b}}(i)$, and the trailing $m$ symbols of $\mathbf{b}$ equal to the first $m$ symbols of $\bar{\mathbf{b}}(i)$ [i.e., $b(i+n+m) = b(i-n+2m-1), \ldots, b(i+n+1) = b(i-n+m)$ and $b(i-n-1) = b(i+n-m), \ldots, b(i-n-m) = b(i+n-2m+1)$]. If we rewrite (1.1) in terms of $\bar{\mathbf{b}}(i)$, i.e.:

$$\mathbf{y}(i) = \bar{\mathbf{M}}\bar{\mathbf{b}}(i) + \mathbf{n}(i), \tag{1.6}$$

then $\bar{\mathbf{M}}$ is circulant, i.e., its rows are cyclic shifts of the first row[6]

$$h(0), \ldots, h(m), 0, \ldots, 0, h(-m), \ldots, h(-1).$$

The matrix $\bar{\mathbf{M}}$ can be diagonalized by pre-multiplying $\bar{\mathbf{b}}$ by the inverse DFT matrix $\mathbf{W}$, and post-multiplying $\mathbf{y}(i)$ by the DFT matrix $\mathbf{W}^{\dagger}$. Hence the conclusion is that (1.1) models a single-user OFDM channel, where the mixing matrix $\mathbf{M}$ can be assumed to be diagonal. The diagonal entries are the complex channel gains across sub-channels.

Other interpretations of the model (1.1) that pertain to uplink and downlink CDMA with dispersive channels are discussed in Chapter 4. In those scenarios, the mixing matrix $\mathbf{M}$ is the product of a Toeplitz, circulant, or diagonal channel matrix, depending on the implementation, and a signature matrix. It becomes apparent that other network configurations with multiple users, multiple antennas, dispersive channels, and with or without spreading can be modeled by (1.1), where $\mathbf{M}$ and $\mathbf{b}(i)$ take on the appropriate forms. For purposes of the following overview, we will focus on the preceding interpretations of the model (1.1), which apply to synchronous CDMA, multi-antenna links (MIMO channels), and dispersive Single-Input/Single-Output (SISO) channels.

## 1.3 OPTIMAL MULTIUSER DETECTION

We now provide a brief overview of multiuser detection techniques that have been proposed for the matrix model (1.1), starting in this section with optimal detectors. The intent is to provide some background and points of reference for the advances and performance results presented in subsequent chapters. We also discuss some additional related techniques.

### 1.3.1 Maximum Likelihood (ML)

Referring to (1.1), the ML detector chooses the vector of estimated symbols as:

$$\hat{\mathbf{b}} = \arg\max_{\mathbf{b}} \Pr(\mathbf{y} \text{ received} \mid \mathbf{b} \text{ transmitted}) \tag{1.7}$$

where the dependence on $i$ has been dropped for convenience. If the additive noise $\mathbf{n}$ is Gaussian, then this is equivalent to selecting:

$$\hat{\mathbf{b}} = \arg\min_{\mathbf{b}} \|\mathbf{y} - \mathbf{M}\mathbf{b}\| \tag{1.8}$$

where the norm is the regular Euclidean norm and the elements of $\mathbf{b}$ are constrained to be constellation points.

[6]Alternatively, the transmitted packet can be rearranged so that the first $2m$ symbols are the same as the trailing $2m$ symbols. The first $2m$ symbols are then the "cyclic prefix" [24, Ch. 12].

The introduction of the ML detector for CDMA in [86] essentially launched the field of multiuser detection. The main reason for this is that it was shown that the performance of the ML detector is insensitive to power variations among the users, unlike the matched filter detector. Hence this suggested that improving the CDMA receiver could relax the requirements on closed-loop power control, which was one of the most complex aspects of CDMA system design.

Although the ML detector can nearly eliminate the degradation in performance due to multiuser interference for low to moderate loads $K/N$ (users per degree of freedom), it has two main drawbacks: complexity and required side information. Because $\mathbf{b}$ takes on discrete values, computing the estimate $\hat{\mathbf{b}}$ is an optimization over a discrete set, which is known to be computationally difficult. Specifically, the computation associated with known algorithms for determining the minimum in (1.1), assuming that $\mathbf{b}$ and $\mathbf{y}$ can be chosen arbitrarily, grows exponentially with $K$ (the size of the vector $\mathbf{b}$), corresponding to the exponential growth in the size of the set over which the minimization is taken. This is not a major issue if the dimension of $\mathbf{b}$ is small (say, $K < 10$ with binary signaling), so that ML detection may be appropriate for single multi-antenna links, or for a TDMA system in which there are relatively few co-channel users per time slot. However, the ML search complexity clearly becomes impractical for a CDMA system with much more than ten users per sector.

The second issue with the ML detector is that (1.1) implicitly assumes that $\mathbf{M}$ is known. Again, this is probably not a major issue for multi-antenna links, which experience slow fading, so that the channel can be accurately estimated. Even so, for some wireless applications, ML multiuser detection may not be as attractive as other simpler decision feedback techniques, to be discussed. The reason is that when combined with error control coding, the combined ML detector for the error control code concatenated with the channel becomes prohibitively complex. In that case, it is desirable to produce soft estimates of the transmitted symbols, i.e., with reliability information. This is further discussed later in this chapter and in Chapter 2.

### 1.3.2 Optimal (Maximum *a Posteriori*) Detection

The Maximum *a Posteriori* (MAP) detector selects:

$$\hat{\mathbf{b}} = \arg\max_{\mathbf{b}} \Pr(\mathbf{b} \text{ transmitted} \mid \mathbf{y} \text{ received}), \tag{1.9}$$

which minimizes the probability of error. This is the same as the ML estimate if the symbols are equally likely. However, when combined with error control coding and iterative soft decoding, the decoder can pass reliability information to the multiuser detector in the form of the *a priori* distribution, or likelihood ratio for each transmitted symbol. Hence, in that scenario the MAP estimate generally differs from the ML estimate. Furthermore, the MAP detector itself computes soft estimates of each symbol (e.g., likelihood ratios), although the final (hard) estimates are obtained from the soft estimates by thresholding.

If the receiver detects a subset of the vector **b**, then the MAP estimate maximizes the corresponding marginal distribution [e.g., $\Pr(b_k \mid \mathbf{y})$ for a particular user $k$]. In general, this differs from the estimate in (1.9) and requires less computation.

For asynchronous CDMA, the multiuser MAP detector can be implemented using the standard *forward-backward algorithm* (e.g., see [87, Sec. 4.2]). The MAP detector suffers from the same drawbacks as the ML detector, namely, the complexity grows exponentially with the size of **b**, and it requires knowledge of **M**. However, in some applications where the system size is relatively small, the complexity may be manageable.

### 1.3.3 Sphere Decoder

Because of the high complexity of the optimal detector, it is generally desirable to incur some performance loss in order to simplify the receiver. That trade off motivates the linear and decision feedback detectors to be discussed. Other reduced-complexity detectors have been proposed that essentially approximate the optimal detector (e.g., see [51,70,75]), or rely on specific properties of the signatures (e.g., [64,66]).

Although the *worst-case* complexity of the ML search increases exponentially with $K$(e.g., see [87, Ch. 4]),[7] the *average* complexity, taking into account the model (1.1), may be much less. An example of a search algorithm with relatively low average complexity, assuming the Gaussian noise model, is the *sphere decoder*, discussed in [10,33].

To describe the sphere decoder, we first observe that if each component of **b** is selected from a rectangular (e.g., QAM) constellation, then each **b** corresponds to a point in a rectangular lattice. The ML estimate in (1.8) can then be computed by performing an exhaustive search over that lattice. To reduce the search complexity, the sphere decoder restricts **b** to lie in a hypersphere of radius $r$ centered at **y**. If the hypersphere contains at least one lattice point, then this restricted search still gives the ML estimate.

Given $r$, an algorithm for finding all points in a lattice within the hypersphere defined by:

$$(\mathbf{b} - \hat{\mathbf{b}})^\dagger \mathbf{M}^\dagger \mathbf{M} (\mathbf{b} - \hat{\mathbf{b}}) \leq r^2, \tag{1.10}$$

where $\hat{\mathbf{b}}$ is the center of the sphere, was presented in [16]. Application of this sphere decoding algorithm to decoding a lattice code was presented in [11], and was subsequently proposed for MIMO channels in [10,33,90]. (See also [76], which considers a somewhat more general search constraint.) For convenience, we assume that all variables are real-valued (hence Hermitian transpose † will be replaced by transpose $T$). The extension to rectangular QAM constellations is obtained by stacking the real and imaginary components of each complex vector (see [10]). The extension to circular (PSK) constellations is discussed in [33].

[7] "Worst-case" refers to the scenario in which **y** and **b** can be chosen to maximize the search time, given a particular search algorithm.

We first note that if $\mathbf{b}$ is not constrained to take on discrete values (i.e., lie in a lattice), then the ML solution in (1.8) is given by $\hat{\mathbf{b}} = (\mathbf{M}^{\dagger}\mathbf{M})^{-1}\mathbf{M}^{\dagger}\mathbf{y}$. It is then easy to show that minimizing $(\mathbf{b} - \hat{\mathbf{b}})^T \mathbf{M}^T\mathbf{M}(\mathbf{b} - \hat{\mathbf{b}})$ over the discrete set of $\mathbf{b}$'s is equivalent to minimizing $\|\mathbf{y} - \mathbf{M}\mathbf{b}\|^2$. The sphere decoder first performs the Cholesky factorization $\mathbf{M}^T\mathbf{M} = \mathbf{U}^T\mathbf{U}$, where $\mathbf{U}$ is upper triangular. Substituting for $\mathbf{M}^T\mathbf{M}$ in (1.10) then gives the equivalent condition:

$$\sum_{k=1}^{K} u_{kk}^2 \left[ b_k - \hat{b}_k + \sum_{j=k+1}^{K} \frac{u_{kj}}{u_{kk}} (b_j - \hat{b}_j) \right]^2 \leq r^2. \tag{1.11}$$

This inequality is then used to generate upper and lower bounds for each component $b_k$, starting with $b_K$ and repeating for $b_{K-1}$, $b_{K-2}$, and so forth. The lattice points within the sphere can then be enumerated by searching through all combinations of component values that lie within the bounding intervals.

More specifically, because all terms in (1.11) are positive, we can discard the first $K-1$ terms in the sum to write:

$$u_{KK}^2 (b_K - \hat{b}_K)^2 \leq r^2, \tag{1.12}$$

which implies:

$$\underline{Q}\left(\hat{b}_K - \frac{r}{u_{KK}}\right) \leq b_K \leq \bar{Q}\left(\hat{b}_K + \frac{r}{u_{KK}}\right), \tag{1.13}$$

where $\bar{Q}(\cdot)$ and $\underline{Q}(\cdot)$ denote, respectively, the largest and smallest constellation values closest to the argument. Once a candidate value for $b_K$ is selected, this is used in a similar way to compute upper and lower bounds on $b_{K-1}$. Namely, discarding the first $K-2$ terms in the sum in (1.10) gives bounds on $b_{K-1}$ in terms of the candidate value $b_K$. A candidate value for $b_{K-1}$ is then selected, which is used with the candidate value $b_K$ to compute bounds for $b_{K-2}$. A candidate value for $b_{K-2}$ is then selected, and the algorithm continues in this way to select candidate values for $b_{K-3}$, $b_{K-4}$, and so forth.

If the bounding intervals computed for each $b_k$, $k = K, K-1, \ldots, 1$, contain at least one constellation value, then the preceding procedure finds a candidate value for $\mathbf{b}$, i.e., the corresponding $\mathbf{b}$ satisifies (1.10). However, it can also happen that for some $k'$, the bounding interval does not contain a constellation point, which implies that any $\mathbf{b}$ with the corresponding candidate values $b_{k'+1}, \ldots, b_K$ must lie outside the sphere defined by (1.10). In that case, the algorithm must backtrack to find alternative candidate values that lead to a candidate $b_{k'}$. That is, the algorithm chooses another candidate value $b_{k'+1}$, if available, and continues as before. If such a $b_{k'+1}$ is not available, or still leads to an empty bounding interval for $b_{k'}$, then the algorithm backtracks to index $k' + 2$. All candidate points $\mathbf{b}$ are then obtained by searching through all possible candidate values at each index $k$, either including the resulting

**b** if the procedure continues to $k = 1$, or terminating the procedure and backtracking when the bounding interval is empty.[8]

To reduce the search time for the ML estimate, each time a candidate point **b** is found, the radius $r$ can be reduced to the corresponding value. That way, the preceding procedure terminates (i.e., continues to $k = 1$) only if a better estimate is found. The complexity of the algorithm is determined by the initial choice of $r$. As $r$ increases, so does the number of candidate values, and hence the search time. As $r$ decreases, the probability that the sphere does not contain any lattice points increases. In that case, the search would likely start over with a larger value of $r$. For the model (1.1) in which the noise is Gaussian, $r$ can be chosen to ensure that this probability is sufficiently small (e.g., see [33,90]).

The expected complexity of the sphere decoder has been shown in [31] to be roughly cubic in the number of data streams $K$ for constellations of interest.[9] This is a big improvement over the worst-case exponential complexity. This cubic complexity order is comparable to that of the decision feedback detectors to be discussed. Still, decision feedback generally has somewhat lower complexity overall, since it can make intermediate decisions for a subset of symbols (e.g., $b_{k+1}, \ldots, b_K$) before estimating $b_k$, thereby avoiding a search through all possible candidates.

The sphere decoder can be extended to MAP as well as ML detection with similar benefits. This extension and the application to iterative detection are discussed in [33,90]. Other related applications are discussed in [89].

## 1.4 LINEAR DETECTORS

Linear detectors are generally much less complex than the optimal detector, making them practical for most applications. A linear multiuser detector takes a linear combination of channel outputs to obtain an estimate of $\mathbf{b}(i)$. It is therefore represented by the $N \times K$ matrix C which forms the (soft) estimate:

$$\tilde{\mathbf{b}}(i) = \mathbf{C}^{\dagger}(i)\mathbf{y}(i) \tag{1.14}$$

As indicated, **C** may depend on the time index $i$, e.g., when estimated in an adaptive mode. In principle, **C** can be selected to minimize the probability of error; however, it is more convenient to select **C** to minimize the Mean Squared Error (MSE):

$$\xi = E\left[\|\mathbf{b}(i) - \tilde{\mathbf{b}}(i)\|^2\right], \tag{1.15}$$

[8]This algorithm has been described in terms of a tree in which the nodes at level $k$ correspond to the candidate values of $b_k$ [10]. The sphere decoder finds all leaves in the tree by tracing paths from different starting points.

[9]Strictly speaking, the expected complexity still grows exponentially with $K$ when $K$ becomes large [40]. This is because the radius $r$ must grow with $K$ to maintain a target detection probability. However, the complexity is more accurately described as being polynomial with $K$ when $K$ is sufficiently small, or the SNR is sufficiently large.

in which case the solution is:

$$\mathbf{C} = \mathbf{M}(\mathbf{M}^{\dagger}\mathbf{M} + \sigma^2\mathbf{I}_K)^{-1} \tag{1.16}$$

where the noise samples in (1.1) are assumed to be uncorrelated with variance $\sigma^2$, and $\mathbf{I}_{\mathrm{K}}$ is the $K \times K$ identity matrix.

It is apparent that a linear detector can detect any subset of the elements of $\mathbf{b}$ without having to estimate the remaining symbols. In other words, we have:

$$\tilde{b}_k(i) = \mathbf{c}_k^{\dagger}(i)\mathbf{y}(i) \tag{1.17}$$

where $\mathbf{c}_k$ is the $k$th column of $\mathbf{C}$. The linear multiuser detector is therefore a bank of uncoupled single-user (or SISO) filters, each of which suppresses interference. This makes the linear multiuser detector suitable for interference suppression applications, such as downlink cellular, where the objective is to detect a single transmitted data stream in the presence of multiuser interference.

Selecting $\mathbf{c}_k$ to minimize the MSE $\xi_k = E[|b_k - \mathbf{c}_k^{\dagger}\mathbf{y}|^2]$ gives:

$$\mathbf{c}_k = \mathbf{R}^{-1}\mathbf{m}_k \tag{1.18}$$

where $\mathbf{R} = E[\mathbf{y}\mathbf{y}^{\dagger}] = \mathbf{M}\mathbf{M}^{\dagger} + \sigma^2\,\mathbf{I}_N$ is the received signal covariance matrix and $\mathbf{m}_k$ is the $k$th column of $\mathbf{M}$. Hence we can write $\mathbf{C} = \mathbf{R}^{-1}\mathbf{M}$, which is an alternate expression to that given in (1.16). The signal power is $E[|\mathbf{c}_k^{\dagger}\mathbf{m}_k|^2]$ and the noise plus interference power is obtained by subtracting this from the total power $E[|\mathbf{c}_k^{\dagger}\mathbf{y}|^2]$. Evaluating these for the optimal $\mathbf{c}_k$ gives the Signal-to-Interference Plus Noise Ratio (SINR):

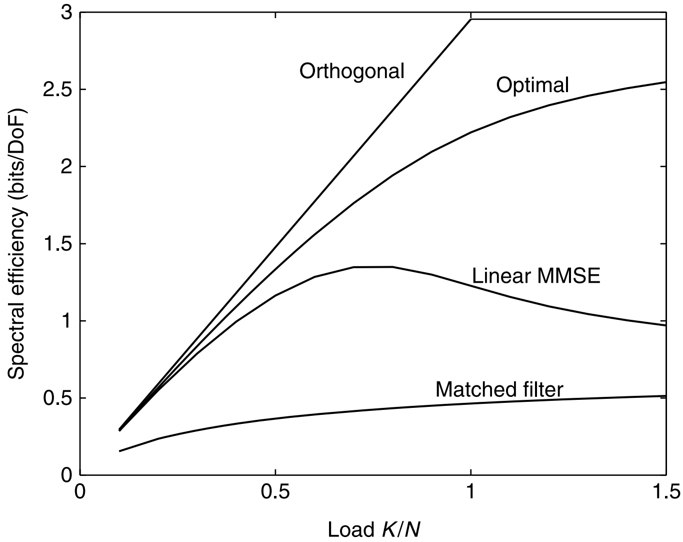
$$\rho_k = \mathbf{m}_k^{\dagger}\mathbf{R}_k^{-1}\mathbf{m}_k \tag{1.19}$$

where $\mathbf{R}_k = \mathbf{R} - \mathbf{m}_k\mathbf{m}_k^{\dagger}$ is the interference-plus-noise covariance matrix for user $k$.

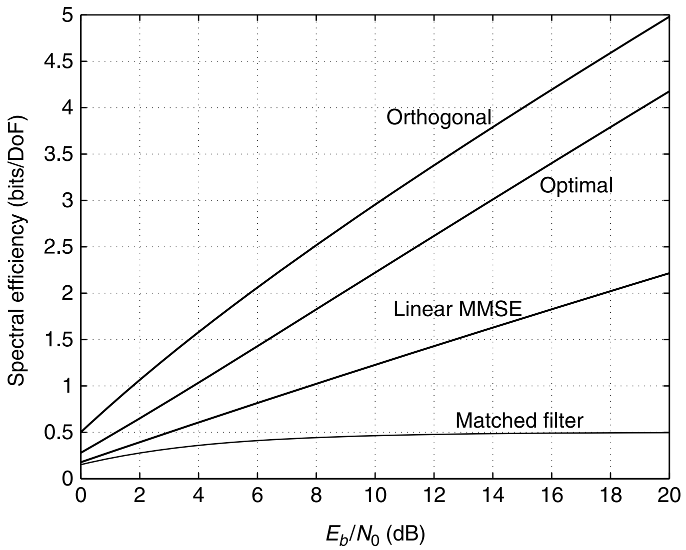
### 1.4.1 Comparison with Optimal Detection

For a particular mixing matrix $\mathbf{M}$ the SINR can be computed from (1.19). To gain additional insight, we can consider a random model for $\mathbf{M}$, and evaluate the performance averaged over $\mathbf{M}$. This approach is discussed in Chapter 4, where performance results are presented for MMSE filters with random inputs and different assumptions concerning the form of $\mathbf{M}$. Here we assume that $\mathbf{M}$ contains *i.i.d.* elements, which applies to CDMA with randomly assigned signatures and equal received powers, and to a MIMO channel with random channel gains. The results in this section are taken primarily from [88].

Figure 1.1 shows plots of sum spectral efficiency versus load $\beta = K/N$ (users per dimension), and versus energy per bit over noise density ($E_b/N_0$). In Figure 1.1a, $E_b/N_0 = 5$ dB and in Figure 1.1b the load $\beta = 0.5$. The sum spectral efficiency is the Shannon capacity summed over all users, and normalized by the degrees of




(**a**) Spectral Efficiency vs. Load ($E_b/N_0$ = 5 dB)




(**b**) Spectrum Efficiency vs. $E_b/N_0$ (load $\beta = 1$)

**Figure 1.1.** Spectral Efficiency vs. Load $E_b/N_0$ for the matched filter, linear **MMSE**, and optimal (**ML**) receivers with a random matrix model (1.1) (i.e., **M** has *i.i.d.* elements). The results are asymptotic as $K$ and $N$ tend to infinity with fixed load $\beta = K/N$. Also shown is the spectral efficiency with orthogonal signatures.

freedom $N$. With the optimal (ML) receiver the sum spectral efficiency is:

$$C = \frac{1}{N} \log \det(\mathbf{I}_N + \sigma^{-2}\mathbf{M}\mathbf{M}^\dagger). \tag{1.20}$$

in bits per channel use per degree of freedom. The Signal-to-Noise Ratio (SNR) is $1/\sigma^2$, so that $E_b/N_0 = 1/(C\sigma^2)$.

With the optimal linear receiver, treating the residual interference plus noise as Gaussian, the spectral efficiency is:

$$C_{\text{lin}} = \frac{1}{N} \sum_{k=1}^{K} \log(1 + \rho_k) \tag{1.21}$$

where $\rho_k$ is the SINR defined in (1.19). Also shown in the figure is the spectral efficiency for the matched filter $\mathbf{c}_k = \mathbf{m}_k$.

The curves shown in Figure 1.1 correspond to the large system limit in which $K$ and $N$ tend to infinity with fixed $K/N$. Explicit expressions for $C$ and $C_{\text{lin}}$ in this limit are given in Chapter 4,[10] and it is shown there that this limit accurately predicts the performance of small finite-size systems averaged over $\mathbf{M}$. These results show that the spectral efficiency achieved by the optimal linear receiver is close to the maximum (achieved with the ML detector) at relatively small loads; however, there is a significant performance gap at large loads (e.g., $\beta > 1/2$), which increases with SNR. This is because as the load increases, the linear receiver has fewer degrees of freedom with which to suppress the interference, and becomes more closely aligned with the matched filter.

### 1.4.2 Properties of Linear Multiuser Detection

Linear detection is discussed in more detail in subsequent chapters. Here we state some of the important properties and limitations of linear multiuser detection.

1. It can be implemented as an adaptive digital filter, which requires little side information. Specifically, either a training sequence is needed at the start of each transmission, or the receiver must know the desired user's signature, channel, and associated timing. Amplitudes, phases, and signatures of interferers are not required for adaptation.
2. As the noise variance $\sigma^2 \to 0$, the MMSE detector completely suppresses a limited number of interferers. Specifically, an $N \times 1$ MMSE filter can completely suppress $N-1$ synchronous interfering streams. If the streams are asynchronous, then a digital filter spanning a single symbol interval can suppress $N/2$ interferers. By increasing the observation window, the filter can suppress up to $N - 1$ interferers [47]; however, adaptation becomes more difficult.

[10]Both $C$ and $C_{\text{lin}}$ converge to *deterministic* limits.

3. The MMSE solution coherently combines all multipath within the window spanned by the filter. This is evident from (1.16), which shows that the MMSE filter is the concatenation of the matched filter matrix $\mathbf{M}$ with an interference suppression matrix. (The matched filter $\mathbf{M}$ combines the multipath.)
4. The performance of the MMSE detector degrades gracefully with the number of (equal power) interferers, although for $K/N$ close to one, the performance of the MMSE solution approaches that of the matched filter.

These properties make linear MMSE estimation attractive for interference suppression applications where little or no information is available about the interferers. Still, mobile cellular applications present major challenges for implementation, as discussed in Section 1.1 Namely, current CDMA systems typically use long signatures, which span a large number of symbols. Although this precludes the use of adaptive filtering for interference suppression, an adaptive filter can still be used for *equalization*. This is important for downlink CDMA, where transmission of orthogonal signatures at the base station does not generally produce an orthogonal set at each mobile receiver. That is, a frequency-selective channel introduces inter-chip interference, which alters the received signature from what was transmitted. Hence, equalizing the channel at the mobile can restore orthogonality, thereby eliminating intra-cell interference [20].

A second challenge to adaptive interference suppression is filter estimation in the presence of time variations due to fading.[11] Various noncoherent techniques have been proposed to enhance performance in the presence of fading, including differential detection [36,39,52,82], generalized likelihood ratios [62,85,91,92], and subspace, or reduced-rank techniques, to be discussed in the next section and in Chapter 3. Those techniques help to make performance more robust with respect to the desired user's channel variations. However, fast time- and frequency-selective fading can substantially degrade the performance of adaptive interference suppression. This is primarily due to the fact that frequency-selective fading changes the interference subspace, making it difficult to track.[12]

The preceding discussion implies that the linear multiuser detector is most appropriate with a relatively small number of users in slow fading. In that case, it is insensitive to power variations across interferers. However, to accommodate more users and provide higher spectral efficiencies, it is necessary to use nonlinear techniques such as ML and decision-feedback, to be discussed.

## 1.5 REDUCED-RANK ESTIMATION

Estimation of a linear MMSE interference suppression filter generally requires a training sequence, which increases linearly with the filter length. This may present

[11] Bursty interference is also a challenge with asynchronous transmissions, e.g., see [34].
[12] This becomes important as the number of interferers increases. Namely, with a small number of interferers, the filter may be able to suppress individual multipath components from all interferers. However, that is no longer possible once the total number of multipath components summed over all users becomes comparable with the filter length.

a problem with long filters (e.g., associated with a large processing gain) due to the associated overhead, or due to channel variations, which may occur within the training period. In those scenarios, to reduce training it may be desirable to *approximate* the linear MMSE detector with another detector, which is suboptimal, but which can be rapidly estimated. An extreme example of this is the matched filter ($\mathbf{c}_k = \mathbf{m}_k$), which may require little or no training, but is not robust with respect to power variations among interferers.

A *reduced-rank* estimate of the linear MMSE filter selects $\mathbf{c}_k$ to minimize the MSE $\xi_k$ subject to the constraint that $\mathbf{c}_k$ lies within a pre-defined subspace. Here we assume that the subspace is spanned by the columns of the $N \times D$ matrix $\mathbf{S}_k$, where the subspace dimension $D < N$. We can therefore write $\mathbf{c}_k = \mathbf{S}_k \mathbf{v}_k$, where $\mathbf{v}_k$ is the $D \times 1$ vector of combining coefficients. Selecting $\mathbf{v}_k$ to minimize the MSE $\xi_k$, given $\mathbf{S}_k$, gives:

$$\mathbf{v}_k = (\mathbf{S}_k^{\dagger}\mathbf{R}\mathbf{S}_k)^{-1}(\mathbf{S}_k^{\dagger}\mathbf{m}_k) \tag{1.22}$$

and the associated SINR for the rank-$D$ filter is:

$$\rho_{k,D} = \mathbf{m}_k^{\dagger}\mathbf{S}_k^{\dagger}(\mathbf{S}_k^{\dagger}\mathbf{R}_k\mathbf{S}_k)^{-1}\mathbf{S}_k\mathbf{m}_k. \tag{1.23}$$

The advantage of reduced-rank estimation is that given $\mathbf{S}_k$, the number of parameters to estimate is reduced from $N$ to $D$. The amount of training overhead needed for filter estimation is therefore reduced in proportion. This also reduces the complexity associated with filter estimation. Of course, this generally comes at the cost of an increase in MSE.

Assuming that $\mathbf{S}_k$ has rank $D$ (i.e., is full column rank), as $D$ increases, the MSE decreases, and when $D = N$ the MSE becomes the MMSE. That is because $\mathbf{c}_k$ is no longer constrained to lie in a lower dimensional subspace.

Of course, the performance (MSE) depends on the choice of $\mathbf{S}_k$. A few different methods for constructing $\mathbf{S}_k$ have been proposed in the literature, and offer different tradeoffs between performance and complexity. Some of these methods are described next. Here we discuss MMSE versions of the reduced-rank filters in which the subspace $\mathbf{S}_k$ is given, and the reduced-rank filter $\mathbf{v}_k$ is optimized. Adaptive versions of these filters can also be developed in which the subspace is estimated along with $\mathbf{v}_k$ (i.e., from the sample covariance matrix).

### 1.5.1 Subspaces from the Matched Filter

Some of the initial reduced-rank schemes for CDMA interference suppression were based on simple manipulations of the desired user's signature $\mathbf{m}_k$. Namely, the columns of $\mathbf{S}_k$ can be taken to be cyclic shifts of the matched filter [53]. Another possibility, proposed in [68] is to let the columns of $\mathbf{S}_k$ be nonoverlapping segments of $\mathbf{m}_k$ padded with zeros. That corresponds to taking linear combinations of the outputs of "partial" matched filters, which are matched to different segments of the matched filter (so-called "partial despreading").

The main advantage of these schemes is that $\mathbf{S}_k$ is easily computed, and a modest performance improvement over the matched filter can be obtained with relatively small values of $D$. Furthermore, $D$ can be varied between $D = 1$, corresponding to the matched filter, and $D = N$, corresponding to the MMSE filter. For partial despreading the corresponding MSE in dB varies almost linearly with $D$.

### 1.5.2 Eigen-Space Methods

Some of the initial work on reduced-rank filtering was motivated by the array (spatial) processing application [44]. The vector $\mathbf{c}_k$ in that scenario again consists of weights across antenna elements, as for the MIMO (multi-antenna) interpretation of the model (1.1), although the weights are generally adjusted to steer the spatial array in a given "look" direction (as opposed to being optimized for data detection). In those schemes, the columns of $\mathbf{S}_k$ (or some subset) are taken to be eigen-vectors of the covariance matrix $\mathbf{R}$, or the interference-plus-noise covariance matrix $\mathbf{R}_k = \mathbf{R} - \mathbf{m}_k\mathbf{m}_k^\dagger$. In some situations (specifically, when the dimension of the signal space $K$ is small relative to the number of dimensions $N$), these methods can achieve MMSE performance with a small rank $D$. Some different eigen-space methods are briefly described below.

***1.5.2.1 Principal Components (PC) [65]*** In this scheme, the columns of $\mathbf{S}_k$ are the $D$ eigen-vectors of $\mathbf{R}$ corresponding to the largest $D$ eigen-values. In other words, the filter $\mathbf{c}_k$ is constrained to lie in the $D$-dimensional subspace, which contains the maximum signal energy. The main motivation for this approach is that the MMSE filter $\mathbf{c}_k$ must lie in the $K$-dimensional subspace spanned by the columns of $\mathbf{M}$. Hence the PC reduced-rank filter is the MMSE filter when $D \geq K$. This can be advantageous when $K$ is small relative to $N$, since a low-rank filter gives optimal performance.

For larger values of $K$, a low-rank PC filter can perform quite poorly (worse than the matched filter). This is especially true when the interfering signals are strong relative to the desired signal. In that case, the filter $\mathbf{c}_k$ is constrained to lie in a subspace that is primarily aligned with the interfering signatures, as opposed to the desired signature. That reduces the signal component at the filter output.

***1.5.2.2 Generalized Side-lobe Canceller (GSC) [81]*** A block diagram of the GSC is shown in Figure 1.2. The received vector $\mathbf{y}$ is passed through a matched

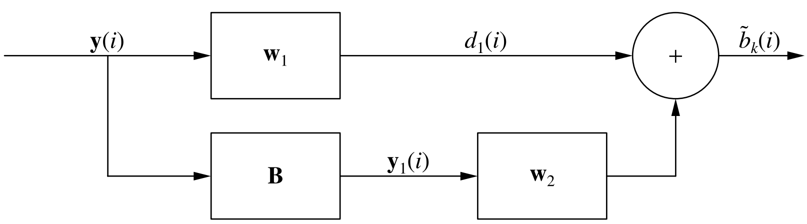



**Figure 1.2.** Block diagram of the Generalized Sidelobe Canceller (GSC).

filter $\mathbf{w}_1 = \mathbf{m}_k$, and also a *blocking matrix* $\mathbf{B}$, which *blocks* the desired signature, i.e., $\mathbf{B}\mathbf{w}_1 = 0$. For example, we can take $\mathbf{B} = \mathbf{I} - \mathbf{w}_1\mathbf{w}_1^\dagger$. The filter output is then $\tilde{b}_k = (\mathbf{w}_1 + \mathbf{B}\mathbf{w}_2)^\dagger\mathbf{y}$. In the full-rank version of the GSC the filter $\mathbf{w}_2$ is selected to minimize the MSE $E[|d_1 - \mathbf{w}_2^\dagger\mathbf{y}_1|^2]$, where $\mathbf{y}_1 = \mathbf{B}\mathbf{y}$ is the output of the blocking matrix. (Equivalently, we can choose $\mathbf{w}_2$ to minimize the output variance $E[|\tilde{b}_k|^2]$.) With this choice, the GSC is equivalent to a scaled MMSE filter. To obtain the MMSE filter, we simply multiply the output of the GSC by an appropriate constant. (Note that this scaling does not change the output SINR, but does affect the MSE.)

In the reduced-rank version of the GSC, $\mathbf{w}_2$ is constrained to lie in a lower-dimensional subspace. Specifically, we can constrain $\mathbf{w}_2$ to lie in an *interference* subspace, spanned by eigen-vectors of the input covariance matrix $\tilde{\mathbf{R}}_1 = E[\mathbf{y}_1\mathbf{y}_1^\dagger] = \mathbf{B}\mathbf{R}\mathbf{B}^\dagger$. To suppress the most interference, and thereby maximize the SINR, we choose the subspace associated with the largest eigen-values of $\tilde{\mathbf{R}}_1$.

The filter subspace for the rank-$D$ GSC is therefore spanned by the desired signature $\mathbf{s}_k$ along with $D-1$ eigen-vectors of $\tilde{\mathbf{R}}_1$. (Those are the columns of the projection matrix $\mathbf{S}_k$.) Note that this subspace is equivalent to that spanned by $\mathbf{m}_k$ and eigen-vectors of the interference-plus-noise covariance matrix $\mathbf{R}_k = \mathbf{R} - \mathbf{m}_k\mathbf{m}_k^\dagger$. Consequently, the reduced-rank GSC can perform no worse than the matched filter, and is optimal in the MMSE sense when $D$ exceeds the dimension of the *interference* subspace. Substantial performance gains, relative to the matched filter, can be achieved with small values of $D$.

***1.5.2.3 Cross-Spectral Method [25]*** In this method, the columns of $\mathbf{S}_k$ are the eigen-vectors of $\mathbf{R}$, which minimize the MSE objective:

$$\xi_k = 1 - \mathbf{s}_k^\dagger\mathbf{R}^{-1}\mathbf{s}_k = 1 - \sum_{i=1}^{D}\frac{|\mathbf{s}_k^\dagger\mathbf{v}_i|^2}{\lambda_i} \tag{1.24}$$

where $\mathbf{v}_1, \ldots, \mathbf{v}_N$ and $\lambda_1, \ldots, \lambda_N$ are the eigen-vectors and corresponding eigen-values of $\mathbf{R}$, and the eigen-values are sorted in decreasing order. Hence, to minimize the MSE we choose the $D$ eigen-vectors associated with the largest values of $|\mathbf{s}_k^\dagger\mathbf{v}_i|^2/\lambda_i$. This metric clearly takes into account how closely $\mathbf{v}_i$ is aligned with the desired signature $\mathbf{s}_k$. It is also interesting that this metric decreases with $\lambda_i$. Hence, unlike PC, a larger value of $\lambda_i$ *decreases* the likelihood that the corresponding eigen-vector will be chosen. Although this method clearly must perform better than PC (with respect to MSE), it generally performs worse than the GSC for small $D$.

***1.5.2.4 Comparison*** All of the preceding methods achieve the MMSE if the filter rank $D$ exceeds the dimension of the signal subspace $K$. This can be beneficial when $K \ll N$. Although for given $D$, principal components performs worse than the other eigen-space methods, the projection matrix $\mathbf{S}_k$ does not depend on the signature $\mathbf{m}_k$. This saves computation when computing multiple reduced-rank filters (e.g., for the multiple-access channel). Also, the performance of the other methods can be substantially degraded when $\mathbf{m}_k$ is not precisely known (e.g., due to channel impairments).

Computing $\mathbf{S}_k$ for eigen-space methods is generally more complex than for the other reduced-rank methods described here. This may be especially challenging in a scenario where the mixing matrix $\mathbf{M}$ is time-varying due to fading and changing interference. Eigen-space methods will be discussed further in Chapter 3 with more elaborate channel models that explicitly account for frequency-selective fading and multiple antennas.

### 1.5.3 Krylov Subspace Methods

Given the covariance matrix $\mathbf{R}$ and the signature $\mathbf{m}_k$, the associated $D$-dimensional Krylov subspace is defined as:[13]

$$\mathcal{K}_k = \mathrm{span}\{\mathbf{m}_k, \mathbf{R}\mathbf{m}_k, \mathbf{R}^2\mathbf{m}_k, \ldots, \mathbf{R}^{D-1}\mathbf{m}_k\}. \tag{1.25}$$

Constraining $\mathbf{c}_k$ to lie in this subspace gives the optimal reduced-rank filter in (1.22), which can be rewritten as:

$$\mathbf{v}_k = \boldsymbol{\Gamma}^{-1}\underline{\gamma} \tag{1.26}$$

where $\Gamma$ is a $D \times D$ matrix with $(l,\ m)$th element $\gamma_{l+m-1} = \mathbf{m}_k^\dagger \mathbf{R}^{l+m-1}\mathbf{m}_k$, and $\underline{\gamma} = [\gamma_0,\ \gamma_1, \ldots, \gamma_{D-1}]^T$.

Attractive properties of this reduced-rank technique are: (i) There are efficient (low complexity) algorithms for computing the reduced-rank filter $\mathbf{v}_k$ given by (1.26), and (ii) the rank needed to achieve near-MMSE (full-rank) performance is typically quite small. In fact, full-rank performance can often be achieved with a $D$ much smaller than the dimension of the signal space.

This type of reduced-rank filter has taken a few different forms in the literature[14], and is closely related to the Lancosz and conjugate gradient algorithms for solving sets of linear equations. An overview of those techniques is given in [12]. Here we emphasize the connection to the Multi-Stage Wiener Filter (MSWF), presented in [26]. We also present a rank-recursive algorithm, which recursively generates filters of increasing rank. That is useful for determining an appropriate filter rank in an adaptive mode. A further discussion of iterative methods for linear interference suppression is given in Chapter 2.

#### *1.5.3.1 Multi-Stage Wiener Filter (MSWF)*

To explain the MSWF, we refer to the block diagram of the GSC in Figure 1.2. We observe that $\mathbf{w}_2$ in this block diagram can be viewed as an estimation filter for $d_1$ given the input $\mathbf{y}_1$. Replacing $\mathbf{w}_2$ by the corresponding MMSE filter gives the full-rank (MMSE) solution for the overall filter $\mathbf{c}_k$.

[13]This subspace does not change if $\mathbf{R}$ is replaced by the interference-plus-noise covariance matrix $\mathbf{R}_k$.

[14]For example, see [7,26,27,29,54,55,58,93] and the discussion of iterative linear techniques in Chapter 2.

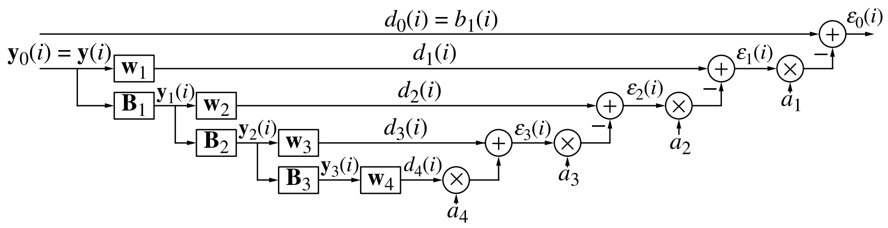



**Figure 1.3.** Block diagram of the Multi-Stage Wiener Filter (MSWF).

We can decompose $\mathbf{w}_2$ in Figure 1.2 in the same way as the GSC decomposition of $\mathbf{c}_k$. This is illustrated in Figure 1.3, where $\mathbf{w}_1$ and $\mathbf{B}_1$ are the same as in Figure 1.2. Also, Figure 1.3 shows the filter output error $\varepsilon_0(i) = b_k(i) - \mathbf{c}_k^\dagger \mathbf{y}(i)$. (The scale factors $a_k$, $k = 1, \ldots, D$, are needed to obtain the MMSE estimate, as discussed in Section 1.5.2.2.) Since $\mathbf{w}_2$ estimates $d_1$ given $\mathbf{y}_1$, the associated "matched filter" is defined as $\mathbf{w}_2 = E[d_1^* \mathbf{y}_1]$, and the blocking matrix $\mathbf{B}_2$ in Figure 1.3 is chosen so that $\mathbf{B}_2 \mathbf{w}_2 = \mathbf{0}$.

Similarly, the filter $\mathbf{w}_3$, which follows $\mathbf{B}_2$ in Figure 1.3, can be viewed as estimating $d_2(i)$ from $\mathbf{y}_2(i)$. Selecting $\mathbf{w}_3$ as the corresponding MMSE filter gives the full-rank MMSE filter. However, Figure 1.3 again shows a GSC decomposition of the filter after $\mathbf{B}_2$. Namely, we take $\mathbf{w}_3 = E[d_2^* \mathbf{y}_2(i)]$, which is the corresponding matched filter, and $\mathbf{B}_3 \mathbf{w}_3 = \mathbf{0}$. Iterating in this fashion, replacing each filter $\mathbf{w}_k$ with the corresponding GSC structure, expands the MMSE filter in terms of the $N$ "matched filters" $\mathbf{w}_1, \ldots, \mathbf{w}_N$. Truncating this process after $D < N$ iterations, setting $\mathbf{w}_D = E[d_{D-1}^* \mathbf{y}_{D-1}]$, gives a MSWF with rank $D$. Figure 1.3 shows an example with $D = 4$. (If $\mathbf{w}_4$ is instead the MMSE filter for estimating $d_3(i)$ given $\mathbf{y}_3(i)$, then we obtain the full-rank MMSE filter.)

From this description it is straightforward to show that the reduced-rank MSWF lies in the subspace spanned by the basis vectors $\mathbf{w}_1$, $\mathbf{B}_1\mathbf{w}_2$, $\mathbf{B}_1\mathbf{B}_2\mathbf{w}_3, \ldots,$ $\left(\prod_{i=1}^{D-1} \mathbf{B}_i\right)\mathbf{w}_D$. (These are the columns of the projection matrix $\mathbf{S}_k$.) Furthermore, it can be shown that this basis is an orthogonal basis for the $D$-dimensional Krylov subspace (1.25) [38]. The combining coefficients in the MSWF are determined by the scale factors $a_0, \ldots, a_{D-1}$ shown in Figure 1.3. The MSWF recursions are given by (1.27)–(1.33).

**ALGORITHM 1: Recursions for Multi-stage Wiener Filter (MSWF)**

Initialization:

$$d_0(i) = b_k(i), \qquad \mathbf{y}_0(i) = \mathbf{y}(i) \tag{1.27}$$

For $n = 1, \ldots, D$ (*Forward Recursion*):

$$\mathbf{w}_n = \boldsymbol{E}[d_{n-1}^* \mathbf{y}_{n-1}(i)] / \|\boldsymbol{E}[d_{n-1}^* \mathbf{y}_{n-1}]\| \tag{1.28}$$

$$d_n(i) = \mathbf{w}_n^\dagger \mathbf{y}_{n-1}(i) \tag{1.29}$$

$$\mathbf{B}_n = \text{null}(\mathbf{w}_n) \qquad (\text{if } n < D) \tag{1.30}$$

$$\mathbf{y}_n = \mathbf{B}_n^\dagger \mathbf{y}_{n-1} \qquad (\text{if } n < D) \tag{1.31}$$

Decrement $n = D, \ldots, 1$ (*Backward Recursion*):

$$a_n = E[d_{n-1}^*(i)\varepsilon_n(i)] / E[|\varepsilon_n(i)|^2] \tag{1.32}$$

$$\varepsilon_{n-1}(i) = d_{n-1}(i) - a_n^* \varepsilon_n(i) \tag{1.33}$$

Where $\varepsilon_D(i) = d_D(i)$. The estimate of $d_0$ is $a_1^* \varepsilon_1$.

***1.5.3.2 Rank-Recursive (Conjugate Gradient) Algorithm*** An alternative method for computing the Krylov reduced-rank filter can be obtained by expressing the rank $D$ filter in terms of the rank $D-1$ filter. To state the algorithm, it will be convenient to denote the basis vectors for the subspace associated with the rank $D$ MSWF as the columns of $\mathbf{V}_D = [\mathbf{v}_1, \ldots, \mathbf{v}_D]$. (That is, $\mathbf{V}_D = \mathbf{S}_k$ for user $k$.) The MSWF has the property that the projected input $\tilde{\mathbf{y}}(i) = \mathbf{V}_D^\dagger \mathbf{y}(i)$ has a real-valued, *tri-diagonal* covariance matrix $\tilde{\mathbf{R}}_D = \mathbf{V}_D^\dagger \mathbf{R} \mathbf{V}_D$ [26]. Let $\tilde{\mathbf{R}}_D$ have diagonal elements $\alpha_1, \ldots, \alpha_D$, and off-diagonal elements $\delta_2, \ldots, \delta_D$. That is, $\delta_n$ is the $(n, n-1)$st, or $(n-1, n)$th element of $\tilde{\mathbf{R}}_D$. (All other elements are zero.)

The conjugate gradient (or Lanczos) version of the Krylov reduced-rank filter is given by the recursions (1.35)–(1.45) (for derivations see [43] or [12]). The reduced-rank filter is $\mathbf{c} = \mathbf{V}_D \tilde{\mathbf{c}}_D$, where $\tilde{\mathbf{c}}_D$ is the $D \times 1$ vector of combining coefficients, and $\mathbf{v}_{D,n}$ denotes the $n$th element of $\mathbf{v}_D$. This algorithm can be used to increment the filter rank to any desired rank starting with the rank-one (matched) filter. The MMSE for the rank-$D$ filter is given by:

$$\xi_D = 1 - \mathbf{m}_k^\dagger \mathbf{V}_D \tilde{\mathbf{R}}_D^{-1} \mathbf{V}_D^\dagger \mathbf{m}_k = 1 - \beta_1 \tilde{\mathbf{c}}_{D,1} \tag{1.34}$$

The recursions (1.35)–(1.45) are an efficient way to compute the reduced-rank filter. What's more, as will be discussed further, this method is easily combined with a stopping rule, which determines when to terminate the recursions (i.e., selects the desired filter rank). Of course, the algorithm must terminate if $\beta_D = 0$, in which case the filter rank is taken to be $D-1$ (assuming $\beta_{D-1} \neq 0$).

***1.5.3.3 Performance*** For a particular mixing matrix $\mathbf{M}$ the reduced-rank SINR as a function of $D$ can be computed from (1.23). To gain additional insight, we again consider a random model for $\mathbf{M}$ (i.e., $\mathbf{M}$ has *i.i.d.* elements). The SINR can be explicitly evaluated in the large system limit as $K \to \infty$ and $N \to \infty$ with fixed ratio $\beta = K/N$ [38]. In this limit the SINR for the rank-$D$ filter converges to a *deterministic* value $\bar{\rho}_D$ (independent of the user), which is given by:

$$\bar{\rho}_D = \frac{1}{\sigma^2 + \dfrac{\beta}{1 + \bar{\rho}_{D-1}}} \tag{1.46}$$

with $\bar{\rho}_0 = 0$. Hence, the SINR in this limit converges to the full-rank MMSE as a *continued fraction*. It takes only a few iterations for $\rho_D$ to converge to the full-rank MMSE (e.g., $D \leq 8$), independent of the load $\beta$ and the noise level $\sigma^2$.

---

**ALGORITHM 2: Rank-Recursive (Conjugate Gradient) Version of a Krylov Subspace Reduced-Rank Filter**

---

Initialization (rank-one filter):

$$\beta_1 = \|\mathbf{m}_k\|, \quad \mathbf{v}_1 = \mathbf{m}_k/\beta_1 \tag{1.35}$$
$$\mathbf{t}_1 = \mathbf{R}\mathbf{v}_1 \tag{1.36}$$
$$\alpha_1 = \mathbf{v}_1^\dagger \mathbf{t}_1 \tag{1.37}$$
$$\tilde{\mathbf{c}}_1 = \beta_1/\alpha_1, \quad \mathbf{q}_1 = [\,]\,(null) \tag{1.38}$$

Increment rank $D$ and perform the following recursions until the stopping rule is satisfied:

$$\mathbf{u}_D = \mathbf{t}_{D-1} - \alpha_{D-1}\mathbf{v}_{D-1} - \beta_{D-1}\mathbf{v}_{D-2} \tag{1.39}$$
$$\beta_D = \|\mathbf{u}_D\| \tag{1.40}$$
$$\mathbf{v}_D = \mathbf{u}_D/\beta_D \tag{1.41}$$
$$\mathbf{t}_D = \mathbf{R}\mathbf{v}_D \tag{1.42}$$
$$\alpha_D = \mathbf{v}_D^\dagger \mathbf{t}_i \tag{1.43}$$
$$\mathbf{q}_D = \begin{bmatrix} \mathbf{q}_{D-1} \\ -1 \end{bmatrix} \frac{\beta_D}{\beta_{D-1}\mathbf{q}_{D-1,D-2} - \alpha_{D-1}} \tag{1.44}$$
$$\tilde{\mathbf{c}}_D = \begin{bmatrix} \tilde{\mathbf{c}}_{D-1} \\ 0 \end{bmatrix} + \frac{\beta_D \mathbf{c}_{D-1,D-1}}{\alpha_D - \beta_D \mathbf{q}_{D,D-1}} \begin{bmatrix} \mathbf{q}_D \\ -1 \end{bmatrix} \tag{1.45}$$

---

We emphasize that to obtain the result in (1.46), the rank $D$ is *fixed* while both $K$ and $N$ become large. Hence the $D$ needed to achieve full-rank performance[15] does

[15]More precisely, "full-rank" performance refers to a target MSE, which can be arbitary close but not equal to the full-rank MSE.

not depend on the dimension of the signal subspace, unlike the eigen-space methods. The extension of this analysis to non-uniform power distributions over the transmitted symbols is presented in [48,50]. Related results, which apply to iterative methods for computing the linear MMSE filter, are discussed in [78]. (See also Chapter 2.)

If we instead consider the single-user intersymbol interference model where $\mathbf{M}$ is given by (1.5), then the performance of the reduced-rank filter depends on the channel impulse response. Although there is currently no expression for the SINR analogous to (1.46) in this case, the SINR can be computed from the moments of the squared magnitude of the channel frequency response [74]. Again the full-rank performance can typically be achieved with a relatively low-rank filter.

***1.5.3.4 Adaptive Rank Selection*** In practice, the parameters of the reduced-rank filter can be estimated with an adaptive algorithm. For example, we can directly estimate the matrix $\mathbf{S}_k$ associated with the Krylov space by replacing $\mathbf{R}$ with the sample covariance matrix $\hat{\mathbf{R}} = \sum_{i=0}^{N} \mathbf{y}(i)\mathbf{y}^{\dagger}(i)$. An estimate for the signature $\mathbf{m}_k$ can also be obtained given a training sequence. For the MSWF, the filters $\mathbf{w}_0, \ldots, \mathbf{w}_{D-1}$ and the combining coefficients $a_0, \ldots, a_{D-1}$ can be estimated by similar types of time-averages (e.g., see [35]).

With limited training or data samples for filter estimation, the choice of filter rank becomes important. Decreasing the rank means fewer parameters must be estimated, which increases estimation accuracy (assuming a fixed training length), but reduces the degrees of freedom available to suppress interference. Choosing the rank is also complicated by the fact that as $D$ increases, the basis vectors spanning $\mathcal{K}_k$ in (1.25) become nearly linearly dependent, even when $D$ is relatively small (i.e., $<K$). Hence computing $\mathbf{v}_k$ via (1.22) becomes ill-conditioned.

As mentioned earlier, the rank-recursive method for computing the reduced-rank filter presented in Section 1.5.3.2 is easily combined with a rank selection, or stopping criterion. For a nonadaptive filter in which $\mathbf{R}$ and the steering vector $\mathbf{m}_k$ are known *a priori*, the MSE can be computed directly from (1.34). That could be used to select $D$ according to a target performance objective. For an adaptive reduced-rank filter in which $\mathbf{R}$ and $\mathbf{m}_k$ are unknown *a priori*, the following stopping criteria can be used.

*Estimated MSE.* Given a training sequence $\{b_k(i)\}$, which is known to the receiver, the filter error for a rank $D$ filter is $e_{k,D}(i) = b_k(i) - \tilde{\mathbf{c}}_D^{\dagger}\mathbf{V}_D^{\dagger}\mathbf{y}(i)$, and the MSE can be estimated as $\hat{\xi}_D = \sum_i |e_{k,D}(i)|^2$, where the summation index corresponds to the training symbols. This estimate, however, cannot be used to select the rank $D$, since $\tilde{\mathbf{c}}_D$ is computed from the data in the sum, so that the corresponding estimated MSE decreases monotonically with $D$. The minimizing rank would therefore be the largest, which can give relatively poor performance. One way to circumvent this problem is to compute the error over non-training symbols, for example, using an estimate for the symbol (e.g., the sign of the filter output if $b_k(i)$ is $\pm 1$). To determine the optimal rank, the cost function must be computed for all possible ranks.

A more efficient method for selecting the termination rank is to use the *relative incremental* estimated MSE as a metric [73, Ch. 3]:

$$f(D) = \frac{\hat{\xi}_{D-1} - \hat{\xi}_D}{\hat{\xi}_{D-1}}. \tag{1.47}$$

That is, we terminate the rank-update algorithm when $f(D) < \varepsilon$, where $\varepsilon$ is a stopping threshold (e.g., between 0.01 and 0.1). The metric $f(D)$ can be efficiently computed in terms of the algorithm parameters, namely:

$$\hat{\xi}_{D-1} - \hat{\xi}_D = \frac{1}{\Delta_D \Delta_{D-1}} \prod_{i=1}^{D} \beta_i^2 \tag{1.48}$$

where $\Delta_D = \det \tilde{\mathbf{R}}_D$, and satisfies:

$$\Delta_D = \alpha_D \Delta_{D-1} - \beta_D^2 \Delta_{D-2}. \tag{1.49}$$

The preceding recursions are initialized as:

$$\hat{\xi}_1 = \frac{1}{T} \sum_{i=1}^{T} |b_k(i)|^2 - \frac{\beta_1^2}{\alpha_1} \tag{1.50}$$

$$\Delta_0 = 1 \qquad \Delta_1 = \alpha_1. \tag{1.51}$$

*Projection Norm.* As the dimension of the Krylov subspace increases, the basis vectors become nearly linearly dependent. As a consequence, the matrix $\tilde{\mathbf{R}} = \mathbf{V}_D^\dagger \mathbf{R} \mathbf{V}_D$ becomes ill-conditioned. One way to prevent this is to increment the rank from $D$ to $D+1$ only if the new basis vector $\mathbf{R}^D \mathbf{m}_k$ is far enough outside the space spanned by $\mathcal{K}_D$. This is equivalent to ensuring that the norm of the orthogonal projection of $\mathbf{R}^D \mathbf{m}_k$ onto $\mathcal{K}_D$ is sufficiently large (i.e., greater than some stopping threshold $\varepsilon$). This norm can be interpreted as a measure of the "new information" provided by the new basis vector. An efficient method for computing this norm in terms of the $\beta_i$ parameters is given in [73, Ch. 3]. This method tends to overestimate the rank with the best performance.

### 1.5.4 Performance Comparison

To illustrate some of the previous discussion, Figure 1.4 (taken from [35]) shows performance (error rate) versus filter rank for the GSC, cross-spectral filter, the MSWF, and a matched filter. Here a CDMA scenario is assumed in which asynchronous users transmit streams of binary symbols with randomly chosen signatures. The filter is estimated from 200 training symbols (i.e., using the sample covariance and an estimate for $\mathbf{m}_k$), the signature length is 128, and there are 42 users with a log-normal

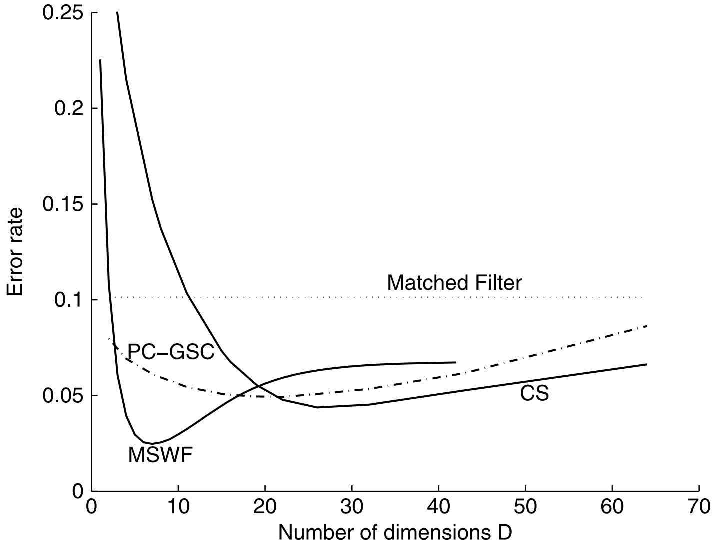



**Figure 1.4.** Error rate versus filter rank for reduced-rank algorithms after training with 200 symbols (taken from [35]). The signature length is 128, there are 42 asynchronous users, and the received powers are log-normal with standard deviation 6 dB. The desired user's SNR is 10 dB.

distribution over the received power (variance 6 dB). The results show that the optimal rank for the MSWF is quite small (around eight), and that the associated performance is better than that of the other (more complicated) eigen-space methods.

## 1.6 DECISION-FEEDBACK DETECTION

One of the main drawbacks of linear multiuser detection is that the performance degrades significantly when the dimension of the space containing strong interference approaches the filter length $N$. This is illustrated in Figure 1.1 which shows that a significantly larger spectral efficiency can be achieved with an optimal (nonlinear) receiver when the load $\beta$ becomes large. However, the optimal receiver may be prohibitively complex when $K$ and $N$ are moderate to large. This motivates the study of relatively simple nonlinear decision feedback (DF), or interference cancellation schemes, which can perform better than the linear receiver.

A block diagram of a DF multiuser detector (DFD) is shown in Figure 1.5. It consists of an $N \times K$ linear feedforward filter (matrix) $\mathbf{F}$ followed by a feedback loop with a $K \times K$ feedback filter (matrix) $\mathbf{B}$. The input to the decision device is:

$$\mathbf{x}(i) = \mathbf{F}^{\dagger}\mathbf{y}(i) - \mathbf{B}^{\dagger}\hat{\mathbf{b}}(i) \tag{1.52}$$

where $\hat{\mathbf{b}}(i)$ is the $K \times 1$ estimate of the transmitted symbol vector $\mathbf{b}(i)$. The purpose of the feedback filter is therefore to *cancel* interference associated with other transmitted symbols. Typically the decision device is nonlinear (i.e., $\hat{\mathbf{b}}$ is a nonlinear

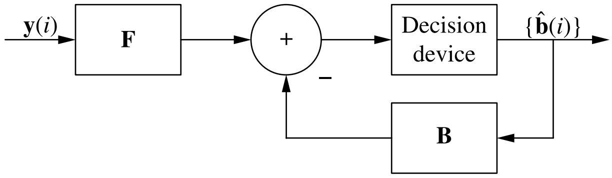



**Figure 1.5.** Block diagram of a decision feedback detector.

function of $\mathbf{b}$), which makes the receiver nonlinear.[16] We assume that $\mathbf{B}$ has zeros along the diagonal to avoid cancelling the desired symbol.

*Successive* DF means that the users are detected in successive order $k = 1, \ldots, K$, so that the estimate $\hat{b}_k(i)$ depends only upon the *previous* estimated symbols $\hat{b}_1(i), \ldots, \hat{b}_{k-1}(i)$. The feedback filter $\mathbf{B}$ in the Successive-DFD (S-DFD) is therefore lower diagonal. In contrast, *parallel* DF means that estimates of *all* other symbols $\hat{b}_m(i)$, $m \neq k$, are available for cancellation. In that case, $\mathbf{B}$ is generally full except for zeros along the diagonal. The P-DFD is generally associated with iterative cancellation methods, where the estimated vector $\hat{b}(i)$, which is fed back for cancellation, is computed in the preceding iteration.

Successive interference cancellation has its origins in earlier work on the Gaussian multiple access channel. Namely, successive decoding and interference cancellation achieves the associated capacity region [21], [9, Ch. 14]. The basic idea is that by coding a user's stream at a rate just below capacity, treating the other users as noise, that user's data can be reliably decoded and then used to cancel the associated interference to other users. The S-DFD was proposed for the CDMA application in [14,15], and is related to earlier work on decision-feedback equalization for MIMO channels [13,45]. It is shown in [28] that the S-DFD achieves the sum capacity at corner points of the CDMA capacity region. (Hence the sum spectral efficiency of the S-DFD receiver coincides with the curve in Figure 1.1 corresponding to the optimal detector.) The S-DFD has been proposed and analyzed for multi-antenna applications in [17,18].

Following the discussion in [94], we can jointly optimize the filters $\mathbf{F}$ and $\mathbf{B}$. Our performance objective is again MSE, where the error is defined as:

$$\mathbf{e}_{\mathrm{dfd}}(i) = \mathbf{b}(i) - \mathbf{x}(i), \tag{1.53}$$

where $\mathbf{x}$ is defined by (1.52). In general, the decision device either makes a hard decision on $\mathbf{b}$ (e.g., $\pm 1$ for binary signaling), or outputs a soft estimate of $\mathbf{b}$ (e.g., the conditional expectation $E[\mathbf{b} \mid \mathbf{y}]$), which uses knowledge about the prior distribution (likelihoods of the different hypotheses) and the noise statistics. Soft estimates are especially desirable when combined with error control coding, and will be discussed in more detail in Chapter 2. For purposes of optimizing the filters $\mathbf{F}$ and $\mathbf{B}$, here we make the simplifying assumption of perfect feedback, i.e., $\hat{\mathbf{b}} = \mathbf{b}$,

[16] A *linear* decision device [e.g., taking $\hat{\mathbf{b}}(i) = \mathbf{x}(i)$] leads to iterative solution methods for the optimal *linear* receiver, which are discussed in Chapter 2.

corresponding to high SNRs. Of course, in practice the feedback contains errors, which can propagate across users. That is much more difficult to analyze, and furthermore, the ideal feedback assumption gives significant insight into the roles of **F** and **B** and associated performance.

We divide the users into two sets:

$$\mathcal{D} = \{ j : \hat{b}_j \text{ is fed back} \} \tag{1.54}$$

$$\mathcal{U} = \{ j : j \notin \mathcal{D} \} \tag{1.55}$$

i.e., "detected" and "undetected" users. In general, these two sets depend on the particular user being detected. We also define the $N \times |\mathcal{D}|$ matrix of signatures for the detected users as $\mathbf{M}_{\mathcal{D}}$, and similarly, $\mathbf{M}_{\mathcal{U}}$ contains the signatures for $k \in \mathcal{U}$.

We observe that the MSE for user $k$, $\varepsilon_{\text{dfd},k} = E[|\mathbf{e}_{\text{dfd},k}|^2]$, depends only on $\mathbf{F}_k$ and $\mathbf{B}_k$, i.e., the $k$th column of **F** and **B**, respectively. Optimizing $\mathbf{F}_k$ and $\mathbf{B}_k$ gives:

$$\mathbf{F}_k = \mathbf{R}_{\mathcal{U}}^{-1}\mathbf{p}_k, \qquad \mathbf{B}_k = \mathbf{P}_{\mathcal{D}}^{\text{H}}\mathbf{F}_k \tag{1.56}$$

where:

$$\mathbf{R}_{\mathcal{U}} \triangleq \mathbf{M}_{\mathcal{U}}\mathbf{M}_{\mathcal{U}}^{\text{H}} + \sigma^2\mathbf{I} \tag{1.57}$$

$$= \mathbf{R} - \mathbf{M}_{\mathcal{D}}\mathbf{M}_{\mathcal{D}}^{\dagger} \tag{1.58}$$

is the covariance matrix for the undetected users.

The feedforward filter for user $k$ is therefore the linear MMSE filter assuming that only users in $\mathcal{U}$ are present. The resulting MMSE is:

$$\mathcal{E}_{\text{dfd},kk} = 1 - \mathbf{m}_k^{\dagger}\mathbf{R}_{\mathcal{U}}^{-1}\mathbf{m}_k, \tag{1.59}$$

where $\mathcal{E}_{\text{dfd}} = E[\mathbf{e}_{\text{dfd}}\mathbf{e}_{\text{dfd}}^{\dagger}]$ is the error covariance matrix, which has the same form as the MMSE for a linear receiver, except that **R** is replaced by $\mathbf{R}_{\mathcal{U}}$. In the absence of error propagation, interference from all users in set $\mathcal{D}$ is therefore eliminated, and user $k$ is affected only by the users in set $\mathcal{U}$. That is, the MMSE DFD cancels interference from the users in set $\mathcal{D}$, while suppressing interference from users in $\mathcal{U}$ in an MMSE sense.[17]

An alternative interpretation for the DFD filters can be obtained by minimizing the sum MSE:

$$\xi_{\text{dfd}} = \sum_{k=1}^{K} E\left[|\mathbf{e}_{\text{dfd},k}|^2\right] = tr[\mathcal{E}_{\text{dfd}}], \tag{1.60}$$

with respect to **F** and **B**. Minimizing with respect to **F** with fixed **B** gives:

$$\mathbf{F} = \mathbf{F}_{\text{lin}}(\mathbf{I} + \mathbf{B}). \tag{1.61}$$

where $\mathbf{F}_{\text{lin}} = \mathbf{R}^{-1}\mathbf{M}$ is the MMSE linear filter. The feedforward filter is therefore a concatenation of the linear MMSE filter with the filter $\mathbf{I} + \mathbf{B}$, as shown in Figure 1.6.

[17]This result is analogous to earlier results for decision-feedback, or "data-aided," equalization [22,56]. It is shown there that conditioned on perfect feedback, the feedback filter cancels the associated intersymbol interference (ISI), and the feedforward filter suppresses the remaining ISI.

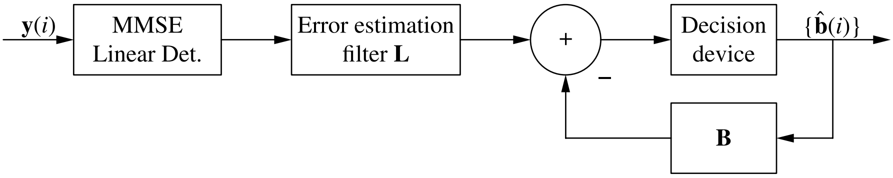



**Figure 1.6.** DFD with an error estimation filter.

From (1.52), (1.53), and (1.61) the error covariance matrix can be expressed as:

$$\mathcal{E}_{\text{dfd}} = (\mathbf{I} + \mathbf{B})^{\dagger}\mathcal{E}_{\text{lin}}(\mathbf{I} + \mathbf{B}), \tag{1.62}$$

where $\mathcal{E}_{\text{lin}}$ is the error covariance matrix for the *linear* MMSE filter. That is:

$$\mathcal{E}_{\text{lin}} = E[\mathbf{e}_{\text{lin}}\mathbf{e}_{\text{lin}}^{\dagger}] = \mathbf{I} - \mathbf{M}^{\dagger}\mathbf{R}^{-1}\mathbf{M} = (\sigma^{-2}\mathbf{I} + \mathbf{M}^{\dagger}\mathbf{M})^{-1} \tag{1.63}$$

where $\boldsymbol{e}_{\text{lin}} = \mathbf{b}(i) - \mathbf{F}_{\text{lin}}^{\dagger}\mathbf{y}(i)$ is the error for the linear MMSE filter, and the last equality follows from the matrix inversion lemma.

We therefore wish to determine the $\mathbf{B}_k$, which minimizes $\text{tr}[\mathcal{E}_{\text{dfd}}]$ (equivalently, $[\mathcal{E}_{\text{dfd}}]_{kk}$), for a given set $\mathcal{D}$. Let $\mathbf{B}_{k,\mathcal{D}}$ be the $D \times 1$ vector containing the elements of $\mathbf{B}_k$ with indices in $\mathcal{D}$. By definition, all other elements of $\mathbf{B}_k$ are zero. Minimizing (1.62) over $\mathbf{B}_k$ gives:

$$\mathbf{B}_{k,\mathcal{D}} = (\mathbf{I} - \mathbf{M}_{\mathcal{D}}^{\dagger}\mathbf{R}^{-1}\mathbf{M}_{\mathcal{D}})^{-1}\mathbf{M}_{\mathcal{D}}^{\dagger}\mathbf{R}^{-1}\mathbf{m}_k \tag{1.64}$$

$$= -[\mathcal{E}_{\text{lin}}]_{\mathcal{D},\mathcal{D}}^{-1}[\mathcal{E}_{\text{lin}}]_{k,\mathcal{D}} \tag{1.65}$$

where $[\mathcal{E}_{\text{lin}}]_{k,\mathcal{D}}$ is the $k$th column of $\mathcal{E}_{\text{lin}}$ taking only rows in set $\mathcal{D}$ and $[\mathcal{E}_{\text{lin}}]_{\mathcal{D},\mathcal{D}}$ is the matrix formed from only those rows and columns of $\mathcal{E}_{\text{lin}}$ in $\mathcal{D}$.

The expression for $\mathbf{B}_k$, given by (1.65), is the MMSE estimation filter for the error $\mathbf{e}_{\text{lin},k}$ ($k$th component of $\mathbf{e}_{\text{lin}}$) given $\mathbf{e}_{\text{lin},\mathcal{D}}$ (i.e., $\mathbf{e}_{\text{lin},m}$ for $m \in \mathcal{D}$). That is, combining (1.52) and (1.61) with (1.53) gives:

$$\mathbf{e}_{\text{dfd}} = (\mathbf{I} + \mathbf{B})^{\dagger}\mathbf{e}_{\text{lin}} \tag{1.66}$$

so that:

$$\mathbf{e}_{\text{dfd},k} = \mathbf{e}_{\text{lin},k} + \mathbf{B}_{k,\mathcal{D}}^{\dagger}\mathbf{e}_{\text{lin},\mathcal{D}}. \tag{1.67}$$

Selecting $\mathbf{B}_{k,\mathcal{D}}$ to minimize $E[|\mathbf{e}_{\text{dfd},k}|^2]$ gives (1.65). Note that the orthogonality principle implies:

$$E[\mathbf{e}_{\text{dfd},k}\mathbf{e}_{\text{lin},m}^{*}] = 0 \tag{1.68}$$

where $m \in \mathcal{D}$. We also remark that the expression for $\mathbf{B}_k$ in (1.56) follows directly from (1.65) by applying the Matrix Inversion Lemma [32, Sec. 13.2].

### 1.6.1 Successive Decision Feedback

For the S-DFD, we have for user $k$:

$$\mathcal{D} = \{1, \ldots, k-1\}, \qquad \mathcal{U} = \{k, \ldots, K\}. \tag{1.69}$$

From (1.68) and (1.67), it is easily shown that:

$$E[\mathbf{e}_{\mathrm{dfd},k}\mathbf{e}^{*}_{\mathrm{dfd},m}] = 0, \quad k \neq m, \tag{1.70}$$

which from (1.66) implies that:

$$\mathcal{E}_{\mathrm{dfd}} = (\mathbf{I} + \mathbf{B})^{\mathrm{H}}\mathcal{E}_{\mathrm{lin}}(\mathbf{I} + \mathbf{B}) \tag{1.71}$$

is *diagonal*. That is, $\mathbf{I} + \mathbf{B}$ can be interpreted as an *error whitening filter* [15]. In this case, $\mathbf{B}$ can be computed via a Cholesky factorization of the error covariance matrix $\mathcal{E}_{\mathrm{lin}}$ in (1.63).

### 1.6.2 Parallel Decision Feedback

For the P-DFD, assuming that all $K$ streams are detected, we have for user $k$:

$$\mathcal{U} = \{k\}, \qquad \mathcal{D} = \{1, \ldots, k-1, k+1, \ldots, K\} \tag{1.72}$$

The initial symbol estimates for feedback might be obtained from the linear MMSE filter, or from an S-DFD. From (1.56) we have:

$$\mathbf{F}_k = \mathbf{R}_{\mathcal{U}}^{-1}\mathbf{m}_k = a\mathbf{m}_k \tag{1.73}$$

where $a = 1/(\|\mathbf{m}_k\|^2 + \sigma^2)$, and combining (1.61) with (1.73) gives:

$$\mathbf{I} + \mathbf{B} = a(\mathbf{M}^{\dagger}\mathbf{M} + \sigma^2\mathbf{I}) \tag{1.74}$$

That is, the feedforward filter is a bank of scaled matched filters, and the off-diagonal components of the feedback matrix are cross-correlations between columns of $\mathbf{M}$. Ideal interference cancellation is therefore achieved with correct feedback estimates. Substituting (1.58) into (1.59) gives the corresponding MMSE:

$$[\mathcal{E}_{\mathrm{p-dfd}}]_{kk} = 1 - \mathbf{m}_k^{\dagger}(\mathbf{m}_k\mathbf{m}_k^{\dagger} + \sigma^2\mathbf{I})^{-1}\mathbf{m}_k = a\sigma^2, \tag{1.75}$$

which applies to a single user in the absence of interference.

For the general case in which $\mathcal{D}$ is an arbitrary subset of transmitted streams, from (1.56) it is apparent that the feedforward filter uses the available degrees of freedom to suppress *only* interference from symbols in $\mathcal{U}$. This applies, for example, to a CDMA system in which $\mathcal{D}$ contains users within the cell and $\mathcal{U}$ contains users in other cells.

From (1.68) and (1.67), it can be shown that the error covariance matrix for the P-DFD is given by:

$$\mathcal{E}_{\text{p-dfd}} = (\mathbf{I} - \mathbf{B})\mathbf{D}_p \tag{1.76}$$

where $\mathbf{D}_p$ is the diagonal matrix with $[\mathbf{D}_p]_{k,k} = E[|\mathbf{e}_{\text{p-dfd},k}|^2]$. In contrast with the S-DFD, the error covariance matrix for the P-DFD is generally a full matrix, and the matrix $\mathbf{I} + \mathbf{B}$ no longer has the interpretation of an error whitening filter.

### 1.6.3 Filter Adaptation

According to the previous discussion, the optimal feedforward and feedback filters depend on knowledge of user signatures, channel characteristics, and the noise variance. However, as with a linear multiuser detector, the DFD filters can be estimated adaptively with a training sequence. For example, if the transmitter transmits a training sequence $\{\mathbf{b}(i)\}$ known to the receiver, then the DFD filters $\mathbf{F}$ and $\mathbf{B}$ can be selected to minimize the Least Squares cost function:

$$\mathcal{C}(i) = \sum_{m=1}^{i} \| \mathbf{e}_{\text{dfd}}(m) \|^2 \tag{1.77}$$

where $\mathbf{e}_{\text{dfd}(m)}$ is given by (1.66). Note that in the multiuser scenario this requires all users to train at the same time (although they may be symbol-asynchronous).

The filters, which minimize $\mathcal{C}(i)$, have the same form as the MMSE filters given in (1.56), and (1.61) and (1.65), where associated expectations are replaced by sums over the time index. Hence, the filters can be computed from either the least squares version of (1.56), or (1.61) and (1.65), where $\mathbf{F}_{\text{lin}}$ is replaced by the linear least squares filter. We also remark that as for the linear receiver, reduced-rank versions of the DFD can be derived to reduce training. Namely, referring to the decomposition of the feedforward filter $\mathbf{F} = \mathbf{F}_{\text{lin}}(\mathbf{I} + \mathbf{B})$, we can replace $\mathbf{F}_{\text{lin}}$ by a reduced-rank filter. The error estimation filter (1.65) can also be approximated as a reduced-rank filter, although the order of this filter may not be large enough to benefit from rank reduction (e.g., see [74]).

### 1.6.4 Error Propagation and Iterative Decision Feedback

The performance of the DFD can be substantially degraded by decision errors in the feedback estimates. This is especially problematic at low SNRs when the detection (DFD) and decoding for error control are separated. (That is, the decoder generates hard decisions based on DFD outputs alone.) In that case, the accumulation of feedback errors can degrade the performance of the DFD relative to a linear detector.[18]

The effects of error propagation can be mitigated by using successive cancellation and demodulation rather than parallel cancellation. Namely, provided that the rate for a particular stream (e.g., user) is less than the capacity, treating the undecoded streams as

[18] Computing an exact expression for the error probability associated with a particular transmitted stream, accounting for error propagation, is generally difficult. However, it is possible to determine the probability that *at least one element* of $\hat{\mathbf{b}}$ is incorrect [59,83].

noise, that stream can be reliably decoded and then cancelled for purposes of detecting the remaining streams. This type of successive cancellation achieves the sum capacity associated with the CDMA or MIMO channel [17,28], but requires that the transmitted streams be coded at the appropriate rates. (Those would have to be determined at the receiver, given the mixing matrix $\mathbf{M}$, and fed back to the transmitters.) Also, the decoding order can affect performance (e.g., see [83]). An associated disadvantage of the S-DFD, relative to the P-DFD for some applications, is that it generally does not provide uniform performance over the multiplexed data streams. Also, since each data stream must be decoded before the next stream can be detected, streams towards the bottom of the decoding order may experience substantial latency.

DFD performance can be improved by combining *soft* estimates for feedback with *iterative* refinements of those estimates. Soft estimation typically refers to weighting the feedback symbols by their "reliabilities." For example, the hard estimate of the symbol can be replaced with its expected value, given some assumptions about the error statistics (e.g., Gaussian). This mitigates the effect of hard decision errors, since symbols in error are likely to have lower reliabilities than correct decisions. (A review of related non-iterative interference cancellation techniques is given in [2].)

Iterative versions of the DFD are obtained by updating and feeding back the estimates $\hat{\mathbf{b}}$. That is, $\hat{\mathbf{b}}$ in iteration $i$ is used to generate the input to the decision device $\mathbf{x}$ in (1.52), which is used to generate the estimate $\hat{\mathbf{b}}$ at iteration $i+1$. This type of iterative cancellation approach to multiuser detection for CDMA was initially proposed in [84], assuming hard decisions for $\hat{\mathbf{b}}$.

When the decision device in Figure 1.5 generates hard estimates, it has been observed that additional DFD iterations typically provide a modest reduction in the error probability [37,94]. Feeding back soft decisions for cancellation further improves performance somewhat; however, much more substantial performance gains are obtained by iterating the soft decision DFD with a soft decision decoder (assuming the presence of error control coding), analogous to iterative decoding of a concatenated (turbo) code. This type of joint iterative multiuser detection and decoding is discussed in detail in the next chapter.

### 1.6.5 Application to MIMO Channels

Both the S- and P-DFD can be directly applied to single-user MIMO channels, where the matrix $\mathbf{M}$ is the channel matrix, and $\mathbf{b}$ is the vector of symbols across transmit antennas. From the preceding discussion, either DFD can be applied independently to each received vector $\mathbf{y}(i)$ to obtain the estimate $\hat{\mathbf{b}}(i)$. The application of the S-DFD to estimate $K$ transmitted streams (one per transmit antenna), where $K \leq N$, was proposed in [18], and further evaluated in [17].[19] An underlying assumption is that each stream is individually coded, hence each transmitted stream corresponds to a user in a CDMA system, where the signature for user (stream) $k$ is the $k$th column of the channel

[19]This combined transmission/reception scheme without coding is given the acronymn "Vertical (V)-BLAST (Bell Laboratories Layered Space-Time)" in [18], since each received vector $\mathbf{y}$ is processed individually as a column (vertically). With coding the acronymn is changed in [17] to "Horizontal (H)-BLAST," since each stream is coded "horizontally" across time.

matrix $\mathbf{M}$ (array of channel gains from transmit antenna $k$ to the receive antennas). The DFD output can then be passed to a bank of $K$ decoders, one for each stream.[20]

Processing each received vector individually by the DFD, as just described, and subsequently decoding each stream is suboptimal in the sense that it cannot achieve the sum capacity of the MIMO channel. This is due to error propagation associated with the DFD. As discussed in the preceding section, two ways to improve performance are (1) successively decode each entire *stream* (i.e., corresponding to a frame or packet) before using the estimates to cancel interference for the next stream; and (2) use iterative methods to refine the DFD estimates. An important feature of the MIMO channel, as opposed to CDMA, is that the transmitter can coordinate the transmission of the $K$ codewords across space and time (e.g., see [79, Ch. 8], [17]). The pattern can therefore be chosen to facilitate successive interference cancellation, and also to equalize the code rates across streams, so that feeding back a set of $K$ rates is no longer required. Alternatively, the transmitter can interleave a single codeword across space and time, and the receiver can apply iterative detection methods with either an S- or P-DFD.

## 1.7 INTERFERENCE MITIGATION AT THE TRANSMITTER

In addition to interference suppression and cancellation at the receiver, in some situations it may be possible to mitigate, or *avoid* interference at the transmitter. In general, this is often accomplished by transmitting on different frequencies or time slots. For the model (1.1), this can also be accomplished by optimizing a precoding matrix for a MIMO channel, or by optimizing signatures assigned to different users.[21]

To illustrate, here we assume that the mixing matrix $\mathbf{M} = \mathbf{HS}$, where $\mathbf{H}$ is a given $N \times K$ channel matrix, and $\mathbf{S}$ is a $K \times K'$ precoding or signature matrix, where $K'$ is the number of multiplexed data streams. That is, the columns of $\mathbf{S}$ can be interpreted as signatures across antennas (for a MIMO channel), or across time (for CDMA). In the case of a MIMO channel, we can choose the matrix $\mathbf{S}$ to optimize a performance metric over all data streams (e.g., the sum rate). In that case we refer to the data streams as being *coordinated*. In contrast, if each column of $\mathbf{S}$ is optimized independently to maximize a performance metric for the associated data stream, then the data streams are *uncoordinated*.

### 1.7.1 Precoding for Coordinated Data Streams

Precoding is illustrated in Figure 1.7, where the channel is decomposed according to the eigen-decomposition:

$$\mathbf{H}^\dagger\mathbf{H} = \mathbf{V}\Lambda\mathbf{V}^\dagger \tag{1.78}$$

[20]The successive canceller for V-BLAST, presented in [18], corresponds to the form of the S-DFD given in (1.56). The equivalence between this form and the S-DFD described in Section 1.6.1 and shown in Figure 1.6, where the "error estimation filter" is a whitening filter, was pointed out in [23].

[21]Note that assigning channels and time slots can be viewed as a special case of "signature" assignment, since the signatures can be defined across frequency and/or time.

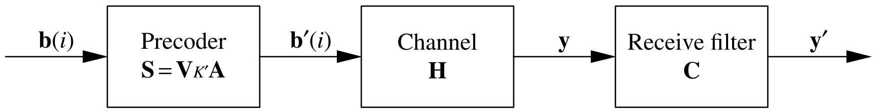



**Figure 1.7.** Linear precoding at the transmitter for a MIMO channel with coordinated data streams.

where $\mathbf{V}$ is a $K \times K$ unitary matrix, the columns of which are the eigen-vectors of $\mathbf{H}^\dagger\mathbf{H}$, and $\Lambda$ is the $K \times K$ diagonal matrix of nonnegative eigen-values. In Figure 1.7, the precoding matrix $\mathbf{S} = \mathbf{V}_{K'}\mathbf{A}$, where $\mathbf{A}$ is a real $K' \times K'$ diagonal matrix, which allocates power across the $K'$ streams, and the columns of $\mathbf{V}_{K'}$ are the eigen-vectors (columns of $\mathbf{V}$) corresponding to the $K'$ largest eigen-values (i.e., $\mathbf{V}_{K'}$ is $K \times K'$). The transmitted symbol vector is then $\mathbf{b}' = \mathbf{V}_{K'}\mathbf{A}\mathbf{b}$, and the receiver multiplies the channel output $\mathbf{y}$ with the receive filter $\mathbf{C}^\dagger = \mathbf{A}\mathbf{V}_{K'}^\dagger\mathbf{H}^\dagger$ to obtain:

$$\mathbf{y}' = \mathbf{A}\mathbf{V}_{K'}^\dagger\mathbf{H}^\dagger\mathbf{H}\mathbf{V}_{K'}\mathbf{A}\mathbf{b} + \mathbf{A}\mathbf{V}_{K'}^\dagger\mathbf{H}^\dagger\mathbf{n} \tag{1.79}$$

$$= \Lambda\mathbf{A}^2\mathbf{b} + \mathbf{n}' \tag{1.80}$$

where $\mathbf{n}' = \mathbf{A}\mathbf{V}_{K'}^\dagger\mathbf{H}^\dagger\mathbf{n}$ and $E[\mathbf{n}'\mathbf{n}'^\dagger] = \sigma^2\mathbf{A}^2\Lambda$. Hence choosing "signatures" (columns of $\mathbf{V}_{K'}$) aligned with the eigen-vectors of $\mathbf{M}$ diagonalizes the channel matrix, thereby eliminating interference.

With additive white Gaussian noise, the corresponding maximum achievable rate, or sum capacity, with precoding is given by [79, Ch. 7]:

$$C = \log\det(\mathbf{I}_{K'} + \sigma^{-2}\mathbf{A}\mathbf{V}^\dagger\mathbf{H}^\dagger\mathbf{H}\mathbf{V}\mathbf{A}) \tag{1.81}$$

$$= \sum_{i=1}^{K'} \log\left(1 + A_i^2\lambda_i/\sigma^2\right) \tag{1.82}$$

where $A_i$ and $\lambda_i$ are the corresponding diagonal elements of $\mathbf{A}$ and $\Lambda$. To maximize $C$, the amplitudes are chosen according to the water-pouring distribution $A_i^2 = \max\{\eta - \sigma^2/\lambda_i, 0\}$, where $\eta$ is the water-pouring level selected to enforce the power constraint $\sum_{i=1}^{K'} A_i^2 \le P_{av}$.

Jointly optimizing the precoding matrix with the receiver clearly must perform better than interference mitigation at the receiver alone. The performance improvement can be substantial at high loads when the receiver is constrained to be suboptimal (e.g., a linear receiver).[22]

With an optimal receiver, the additional gain due to optimal precoding may be modest (e.g., see [79, Ch. 8]); however, precoding generally simplifies the coding, decoding, and detection. Namely, the preceding discussion for coordinated data

[22]The load refers to the number of independent data streams relative to the number of degrees of freedom (e.g., number of antennas or processing gain).

streams implies that with optimal precoding a linear matched filter achieves the sum capacity. Furthermore, codes designed for the single-input/single-output AWGN channel can be applied directly by coding over the channel modes (eigen-vectors). This approach to combined error control coding and linear precoding avoids the use of more complicated space-time codes, and achieves the associated channel capacity.

The main drawback of precoding is that the channel or precoding matrix must be estimated at the receiver, and then relayed back to the transmitter through a feedback channel. This must be done frequently in situations where the data streams and channels are time-varying (e.g., due to mobility).

***1.7.1.1 Precoding for Equalizing SNR Performance*** The capacity given by (1.82) can be achieved by transmitting random Gaussian codewords, where the elements of the codewords are spread across the different channel eigen-vectors (columns of $\mathbf{V}_{K'}$). The variance of the elements are chosen according to the water pouring distribution. Practical codes, however, choose the codeword elements from discrete constellations. Hence in practice, coded bits can be mapped to a set of constellation points, one for each eigen-mode, where the size of a constellation is matched to the associated channel gain (eigen-value). This leads to "bit loading" schemes in which the rate for the $i$th eigen-mode (i.e., information bits per symbol) is selected to approximate $\log(1 + A_i^2\lambda_i^2/\sigma^2)$, where $A_i^2$ is the optimized power [8].[23]

Constraining the set of constellations, e.g., restricting the codeword elements to be selected from the *same* constellation over all modes, limits the ability to do bit loading. In that situation it is desirable to decompose the channel into effective modes or "sub-channels" having the *same* SNR [i.e., as opposed to the eigen-decomposition (1.78), which gives the set of SNRs $\{A_i^2\lambda_i^2/\sigma^2\}$]. This type of *uniform* channel decomposition is presented in [42,95], and relies on the following observations:

1. Replacing $\mathbf{S}$ in Figure 1.7 with $\mathbf{S}\Phi$, where $\Phi$ is any unitary matrix, does not change the sum capacity.
2. Any $N \times K$ matrix $\mathbf{M}$ with rank $r$ can be written as:

$$\mathbf{M} = \mathbf{QUP}, \tag{1.83}$$

   where $\mathbf{Q}$ is $N \times r$, $\mathbf{P}$ is $K \times r$, both $\mathbf{P}$ and $\mathbf{Q}$ have orthonormal columns, and $\mathbf{U}$ is an $r \times r$ upper triangular matrix with equal diagonal elements.

The first observation gives the additional degrees of freedom needed to vary the set of received SNRs per transmitted stream without sacrificing total rate. The second observation is used to compute the precoder and the receiver. (The proof of (1.83) is given in [41,96].)

[23] In practice, the same power can be used on each active channel without sacrificing much rate.

Replacing the optimal precoding matrix in Figure 1.7, $\mathbf{S} = \mathbf{V}_{K'}\mathbf{A}$, by $\mathbf{V}_{K'}\mathbf{A}\Phi$, where $\Phi$ is orthonormal, introduces interference among the transmitted substreams, hence the linear receiver $\mathbf{C}$ shown in Figure 1.7, is no longer optimal. For the uniform channel decomposition, we therefore replace $\mathbf{C}$ by an S-DFD, which has front-end filter $\mathbf{F}$ given by (1.61). Let $\tilde{\mathbf{M}} = \mathbf{HS} = \mathbf{HV}_{K'}\mathbf{A}\Phi$ denote the combined channel and precoder matrix. Then from (1.71) and (1.63) we have:

$$\tilde{\mathbf{R}} = \tilde{\mathbf{M}}^\dagger \tilde{\mathbf{M}} + \sigma^{-2}\mathbf{I}_{K'} \tag{1.84}$$

$$= (\mathbf{I}_{K'} + \mathbf{B})\mathbf{D}^{-1}(\mathbf{I}_{K'} + \mathbf{B})^\dagger \tag{1.85}$$

where $\mathbf{D}$ is diagonal.

As discussed in Section 1.6.1, assuming perfect cancellation, the MSE for the $k$th substream is $\mathbf{D}_{kk}$. Although the sum capacity $C = \log\det[\tilde{\mathbf{R}}] = \sum_{k=1}^{K'} \log \mathbf{D}_{kk}^{-1}$ does not depend on $\Phi$, it can be used to change the MSEs (or SNRs) across substreams. This can be achieved by factoring out $\Phi$ in (1.84), and writing $\tilde{\mathbf{R}} = \Phi^\dagger \mathbf{U}^\dagger \mathbf{U}\Phi$, where $\mathbf{U}$ is upper triangular. The preceding decomposition (1.83) can then be applied to $\tilde{\mathbf{R}}^{1/2} = \mathbf{QU}\Phi$, where $\mathbf{Q}$ is selected along with $\Phi = \mathbf{P}$. Algorithms for determining $\Phi$ and $\mathbf{Q}$ are presented in [42,95]. The resulting precoding matrix $\mathbf{S}\Phi$ combined with an S-DFD at the receiver then gives the same SNR across all $K'$ data streams.

### 1.7.2 Signature Optimization with Uncoordinated Data Streams

In some scenarios, such as a peer-to-peer network, the data streams correspond to different users, who may not be able to coordinate transmissions. In that case, a particular user may adjust his own signature (for spread spectrum transmission), or precoding matrix (with multiple antennas), to optimize his own performance, ignoring the effect on other users.

This is illustrated with the system model:

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{s}_k A_k b_k + \sum_{i \neq k} \mathbf{H}_i \mathbf{s}_i A_i b_i + \mathbf{n} \tag{1.86}$$

where $\mathbf{y}_k$ is the received signal for user $k$, and $\mathbf{H}_i$ is the channel from transmitter $i$ to receiver $k$. Suppose that user $k$ has a linear MMSE receiver, and can select the signature $\mathbf{s}_k$ to minimize the MSE $E[|\mathbf{c}^\dagger \mathbf{y}_k - b_k|^2]$. It can be shown that, assuming all other signatures and channels are fixed, the optimized signature $\mathbf{s}_k$ is the eigenvector of the matrix $\mathbf{H}_k^\dagger \mathbf{R}_k^{-1} \mathbf{H}_k$ corresponding to the maximum eigen-value, where $\mathbf{R}_k = \sum_{i \neq k} A_i^2 \mathbf{H}_i \mathbf{s}_i \mathbf{s}_i^\dagger \mathbf{H}_i^\dagger + \sigma^2 \mathbf{I}$ is the interference-plus-noise covariance matrix at receiver $k$.

The optimal signature depends on the channels, so that significant feedback may be required with mobile users. The model (1.86) also applies to the multiple access channel (i.e., with co-located receivers), and in that case it is desirable to select the set of signatures $\{\mathbf{s}_1, \ldots, \mathbf{s}_K\}$ to maximize a global performance objective, such as sum rate over the users. A basic question is whether or not the individual (or "greedy") optimization described here, which does not account for interference to other users, can maximize such a global performance objective. This is discussed in detail in Chapter 7.[24]

### 1.7.3 Network Configurations

The ability to coordinate different data streams depends on the network configuration. Coordination is relatively easy when the transmitted data streams are co-located, as in a single-user MIMO channel. Another example in which coordinated precoding can provide substantial benefits is the downlink of a single cell in a cellular system. There the optimal precoding, which maximizes the total sum rate over the users, is *nonlinear*. In addition, the transmission to a particular user depends on the messages being transmitted to other users [9, Ch. 14]. This optimal precoder and associated performance are discussed in Chapter 8.

For the uplink the base station can, in principle, compute a set of optimized signatures for the users given knowledge of the uplink channels [e.g., $\mathbf{H}_i$, $i = 1, \ldots, K$ in (1.86)]. Coordination in that scenario mainly consists of synchronizing the users. Unlike the downlink, users are not likely to be able to coordinate codewords. The uplink sum capacity is achieved by the dual operation of successive decoding and cancellation of interfering users at the base station. (See [79, Ch. 10] and the discussion in Chapter 8.)

As previously mentioned, coordination of data streams may be difficult in an ad hoc, or peer-to-peer network in which transmitter-receiver pairs are not co-colocated. In those scenarios each transmitter-receiver pair may optimize a local performance objective, ignoring the effect on other transmitter-receiver pairs. The performance of such a network can be studied within the framework of game theory [71]. Namely, the transmitter-receiver pairs can be viewed as *non-cooperative agents*, each acting according to their own self-interest. The users can then be expected to adjust their parameters (e.g., powers, signatures) until they reach an equilibrium in which no further improvements to local performance are possible.[25] This game-theoretic framework is currently being used to study distributed resource allocation in a variety of communications scenarios (e.g., see [4]).

In general, a wireless network may consist of a combination of coordinated and uncoordinated nodes. For example, it is generally easier to coordinate users within a cell than it is to coordinate users across cells. Optimization of signaling methods along with signatures, or precoding matrices, in such a general environment is a challenging direction for future work.

[24]For the model considered here (uncoordinated users each with a single signature) greedy optimization generally does not maximize the sum rate, although it typically comes close [61].

[25]More precisely, a *Nash equilibrium* is achieved when each user has no incentive to change its current power and/or signature given a fixed set of choices for powers and signatures across the other users.

## 1.8 OVERVIEW OF REMAINING CHAPTERS

We conclude this chapter by relating the previous discussion and overview of multiuser detection techniques to the advances discussed in subsequent chapters. Generally speaking, these advances fall into two main categories: advances in detection techniques, and advances in performance analysis. With the exception of iterative techniques, discussed in Chapter 2, the advances in detection techniques are motivated by more elaborate, or different modeling assumptions (i.e., that account for multipath, fading, and downlink transmissions). The advances in performance analysis are motivated by parallel advances in relevant mathematical methods.

Referring to the fundamental performance limits shown in Figure 1.1, optimal multiuser-detection offers a significant increase in spectral efficiency relative to the linear detector at high SNRs and high loads. As discussed in Section 1.6, that has motivated the development of relatively simple nonlinear decision feedback and iterative techniques that can close this gap. When combined with error control coding, the multiuser detector can exploit reliability information generated by the decoder. Iterating between the (soft output) multiuser detector and the soft decoder is analogous to iterative (Turbo) decoding, and the performance approaches that corresponding to a single isolated user (i.e., without interference) over a wide range of loads and SNRs. An overview of these types of iterative techniques is presented in Chapter 2.

The results in Figure 1.1 are based on the model (1.1), where $\mathbf{M}$ has *i.i.d.* elements. Although this model has been quite useful for developing and analyzing multiuser detection techniques, it does not account for channel impairments encountered in practice, e.g., associated with wireless channels. In particular, multipath introduces fading and intersymbol interference, which pose additional challenges for multiuser receiver design. Linear receivers for multiuser channels with multipath are developed in Chapter 3, including variants of reduced-rank receivers, which are based on eigen-decompositions of appropriate signal subspaces.

For the basic multiuser model (1.1), where $\mathbf{M}$ is random, it is often desirable to obtain performance measures that are averaged over realizations of $\mathbf{M}$. This can be difficult in general, but sometimes becomes tractable in the large system limit in which the dimensions of $\mathbf{M}$ are scaled with a fixed ratio. This has already been suggested by the performance results shown in Figure 1.1. This type of large system analysis is discussed Chapter 4, where derivations of some key multiuser performance measures are presented for a few variations on the channel model (1.1). For the random matrix model (1.1) the techniques in Chapter 4 cannot be applied in a straightforward way to determine the performance (error probability) of the *optimal* detector. To evaluate that performance measure in the large system limit, it has been necessary to apply techniques from the statistical physics literature. That analysis is introduced in Chapter 5, along with a general estimation framework for interpreting the performance of different (suboptimal) multiuser detectors.

As previously mentioned, the model (1.1) applies to both channels with multiple antennas and multi-user systems. This connection is developed in more detail in Chapter 6, where fundamental (capacity) limits associated with multi-antenna systems are presented along with performance results for linear receivers. In particular, results for the random matrix model (1.1) are presented for finite dimensions, which complement the large system results in Chapter 4.

Chapters 7 and 8 discuss methods for *jointly optimizing* transmitters and receivers, as previously discussed in Section 1.7. The discussion in Chapter 7 is focused on *avoiding* interference through selection of signatures at a transmitter. As previously pointed out in Section 1.7, that can be viewed as a linear precoding technique. In contrast, Chapter 8 discusses optimal multiuser precoding for the downlink, which processes the input symbols in a nonlinear fashion. There the optimization criterion is the rate summed over users. The discussion is generalized to the situation in which there are multiple antennas at the base station and at each mobile (i.e., the MIMO downlink). In those scenarios substantial performance gains are achievable through optimal precoding.

Finally, while there have been many advances in multiuser detection over the past few years, there are a number of important related directions for future work. These include extensions and generalizations of the techniques and analysis presented here to other channel models (e.g., combinations of wideband, MIMO, slow or fast fading) and network configurations (e.g., multi-cell, peer-to-peer, multi-hop). Additional modifications may be motivated by the presence (or lack) of limited information about channels and interferers, which may (or may not) be available in practice.

From a practical point of view, numerous challenges arise when trying to incorporate many of the techniques discussed here in actual systems. These include adaptive channel estimation in the presence of time- and frequency-selective fading (for interferers as well as for the desired data stream), allocation of system (computing) resources within a receiver (e.g., between decoding and detection), power management (consumed by the device as well as transmitted power), limited feedback schemes for relaying channel (or decoder state) information from the receiver to a transmitter, in addition to managing system cost. Some of these issues, such as channel estimation, become more complicated as the number of degrees of freedom increases (e.g., by increasing bandwidth or number of antennas). It is our hope that the advances described in this collection will serve as a springboard for future investigations into all of these areas.

## REFERENCES

1. N. Amitay and J. Salz. Linear equalization theory in digital data transmission over dually polarized fading radio channels. *Bell System Technical Journal*, pages 2215–2259, December 1984.
2. J. G. Andrews. Interference cancellation for cellular systems: a contemporary overview. *IEEE Wireless Communications*, 12(2):19–29, April 2005.
3. M. Andrews, K. Kumaran, K. Ramanan, A. L. Stolyar, R. Vijayakumar, and P. Whiting. Providing quality of service over a shared wireless link. *IEEE Communications Magazine*, 39(2):150–154, 2001.
4. T. Basar, J. Huang, N. B. Mandayam, V. P. Palomar, J. Walrand, and S. B. Wicker, editors. *IEEE Journal on Selected Areas in Communications*, 2008. Special Issue on Game Theory in Communication Systems.

5. P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi. CDMA/HDR: a bandwidth-efficient high-speed wireless data service for nomadic users. *IEEE Communications Magazine*, pages 70–77, July 2000.
6. E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor. *MIMO Wireless Communications*. Cambridge University Press, 2007.
7. W. Chen, U. Mitra, and P. Schniter. On the equivalence of three reduced rank linear estimators with applications to DS-CDMA. *IEEE Transactions on Information Theory*, 48(9):2609–2614, 2002.
8. P. S. Chow, J. M. Cioffi, and J. A. C. Bingham. A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels. *IEEE Transactions on Communications*, 43(234):773–775, April 1995.
9. T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons, 2nd edition, 2006.
10. M. O. Damen, H. El Gamal, and G. Caire. On maximum-likelihood detection and the search for the closest lattice point. *IEEE Transactions on Information Theory*, 49(10):2389–2402, 2003.
11. O. Damen, A. Chkeif, and J. C. Belfiore. Lattice code decoder for space-time codes. *IEEE Communications Letters*, 4(5):161–163, 2000.
12. G. K. E. Dietl. Linear estimation and detection in Krylov subspaces. *Foundations in Signal Processing, Communications, and Networking*, 1, 2007.
13. A. Duel-Hallen. Equalizers for multiple input/multiple output channels and PAM systems with cyclostationary input sequences. *IEEE Journal on Selected Areas in Communications*, 10(3):630–639, April 1992.
14. A. Duel-Hallen. Decorrelating decision-feedback multiuser detector for synchronous code-division multiple-access channel. *IEEE Transactions on Communications*, 41(2):285–290, February 1993.
15. A. Duel-Hallen. A family of multiuser decision-feedback detectors for asynchronous code-division multiple-access channels. *IEEE Transactions on Communications*, 43(2/3/4):421–434, February/Mar/Apr 1995.
16. U. Fincke and M. Pohst. Improved methods for calculating vectors of short length in a lattice, including a complexity analysis. *Mathematics of Computation*, 44(170):463–471, 1985.
17. G. J. Foschini, D. Chizhik, M. J. Gans, C. Papadias, and R. A. Valenzuela. Analysis and performance of some basic space-time architectures. *IEEE Journal on Selected Areas in Communications*, 21(3):303–320, April 2003.
18. G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky. Simplified processing for high spectral efficiency wireless communication employing multi-element arrays. *IEEE Journal on Selected Areas in Communications*, 17(11):1841–1852, November 1999.
19. G. J. Foschini and M. J. Gans. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications (Kluwer)*, 6:311–335, 1998.
20. C. D. Frank, E. Visotsky, and U. Madhow. Adaptive interference suppression for the downlink of a direct sequence CDMA system with long spreading sequences. *Journal of VLSI Signal Processing*, 30:273–291, 2002.
21. R. Gallager. A perspective on multiaccess channels. *IEEE Transactions on Information Theory*, 31(2):124–142, March 1985.

22. A. Gersho and T. L. Lim. Adaptive cancellation of intersymbol interference for data transmission. *Bell System Technical Journal*, 60:1997–2021, November 1981.
23. G. Ginis and J. M. Cioffi. On the relation between V-BLAST and the GDFE. *Communication Letters*, 5(9):364–366, September 2001.
24. A. Goldsmith. *Wireless Communication*. Cambridge University Press, 2005.
25. J. S. Goldstein and I. S. Reed. Reduced rank adaptive filtering. *IEEE Trans. Signal Processing*, 45(2):492–496, February 1997.
26. J. S. Goldstein, I. S. Reed, and L. L. Scharf. A multistage representation of the Wiener filter based on orthogonal projections. *IEEE Trans. Inform. Theory*, 44(7), November 1998.
27. A. Grant and C. Schlegel. Convergence of linear interference cancellation multiuser receivers. *IEEE Transactions on Communications*, 10(49):1824–1834, October 2001.
28. T. Guess and M. K. Varanasi. An information-theoretic framework for deriving canonical decision-feedback receivers in Gaussian channels. *IEEE Transactions on Information Theory*, 51(1):173–187, January 2005.
29. D. Guo, L. K. Rasmussen, S. Sun, and T. J. Lim. A matrix-algebraic approach to linear parallel interference cancellation in CDMA. *IEEE Transactions on Communications*, 48:152–161, January 2000.
30. M. Haardt, A. Klein, R. Koehn, S. Oestreich, M. Purat, V. Sommer, and T. Ulrich. The TD-CDMA based UTRA TDD mode. *Selected Areas in Communications, IEEE Journal on*, 18(8):1375–1385, 2000.
31. B. Hassibi and H. Vikalo. On the sphere-decoding algorithm I. Expected complexity. *IEEE Transactions on Signal Processing*, 53(8 Part 1):2806–2818, 2005.
32. S. Haykin. *Adaptive Filter Theory*. Prentice Hall, 4th edition, 2002.
33. B. M. Hochwald and S. Ten Brink. Achieving near-capacity on a multiple-antenna channel. *IEEE Transactions on Communications*, 51(3):389–399, 2003.
34. M. L. Honig. Adaptive linear interference suppression for packet DS-CDMA. *European Transactions on Telecommunications and Related Technologies*, 9(2):173–181, March/apr 1998.
35. M. L. Honig and J. S. Goldstein. Adaptive reduced-rank interference suppression based on the multistage Wiener filter. *IEEE Transactions on Communications*, 50(6):986–994, June 2002.
36. M. L. Honig, S. L. Miller, M. J. Shensa, and L. B. Milstein. Performance of adaptive linear interference suppression in thepresence of dynamic fading. *IEEE Transactions on Communications*, 49(4): 635–645, 2001.
37. M. L. Honig and R. Ratasuk. Large-system performance of iterative multiuser decision-feedback detection. *IEEE Transactions on Communications*, 51(8):1368–1377, August 2003.
38. M. L. Honig and W. Xiao. Performance of reduced-rank linear interference suppression. *IEEE Transactions on Information Theory*, 47(5):1928–1946, July 2001.
39. H. Huang and S. Verdú. Linear differentially coherent multiuser detection for multipath channels. *Wireless Personal Communications*, 6(1/2):113–136, 1998.
40. J. Jalden and B. Ottersten. On the Complexity of Sphere Decoding in Digital Communications. *IEEE Transactions on Signal Processing*, 53(4):1474–1484, 2005.

41. Y. Jiang, J. Li, and W. W. Hager. Joint transceiver design for MIMO communications using geometric mean decomposition. *IEEE Transactions on Signal Processing*, 53(10):3791–3803, October 2005.

42. Y. Jiang, J. Li, and W. W. Hager. Uniform Channel Decomposition for MIMO Communications. *IEEE Transactions on Signal Processing*, 53(11):4283–4294, November 2005.

43. M. Joham, Y. Sun, M. D. Zoltowski, M. Honig, and S. Goldstein. A new backward recursion for the multi-stage nested Wiener filter employing Krylov subspace methods. In *Proc. MILCOM 2001*, pages 1210–1213, McLean, Va., October 2001.

44. D. H. Johnson and D. E. Dudgeon. *Array Signal Processing*. Prentice Hall, Signal Processing Series, 1992.

45. M. Kavehrad and J. Salz. Cross-polarization cancellation and equalization in digital transmission over dually polarized multipath fading channels. *Bell System. Technical Journal*, pages 2211–2245, December 1985.

46. A. R. Kaye and D. A. George. Transmission of multiplexed PAM signals over multiple channel and diversity systems. *IEEE Trans. Commun. Techn.*, pages 520–526, October 1970.

47. Kiran and D. N. C. Tse. Effective interference and effective bandwidth of linear multiuser receivers in asynchronous CDMA systems. *IEEE Transactions on Information Theory*, 46(4):1426–1447, July 2000.

48. L. Li, A. M. Tulino, and S. Verdú. Design of reduced-rank MMSE multiuser detectors using random matrix methods. *IEEE Transactions on Information Theory*, 50(6):986–l008, 2004.

49. X. Liu, E. K. P̃. Chong, and N. Shroff. Opportunistic transmission scheduling with resource sharing constraints in wireless networks. *IEEE Journal on Selected Areas in Communications*, 19(10), October 2001.

50. P. Loubaton and W. Hachem. Asymptotic analysis of reduced rank Wiener filters. *Proc. 2003 Information Theory Workshop*, pages 328–331, March 2003.

51. W.-K. Ma, T. N. Davidson, K. M. Wong, Z.-Q. Luo, and P.-C. Ching. Quasi-maximum-likelihood multiuser detection using semi-definite relaxation with application to synchronous CDMA. *IEEE Transactions on Signal Processing*, 50(4):912–922, April 2002.

52. U. Madhow, K. Bruvold, and L. J. Zhu. Differential MMSE: A Framework for Robust Adaptive Interference Suppression for DS-CDMA Over Fading Channels. *IEEE Transactions on Communications*, 53(8):1377–1390, 2005.

53. U. Madhow and M. L. Honig. MMSE interference suppression for direct-sequence spread spectrum CDMA. *IEEE Transactions on Communications*, 42(12):3178–3188, December 1994.

54. N. B. Mandayam and S. Verdú. Analysis of an approximate decorrelating detector. *Wireless Personal Communications: An International Journal*, 6(l–2):97–111, January 1998.

55. S. Moshavi, E. G. Kanterakis, and D. L. Schilling. Multistage linear receivers for DS-CDMA systems. *International Journal of Wireless Information*, 3:1–17, January 1996.

56. M. S. Mueller and J. Salz. A unified theory of data-aided equalization. *Bell System Technical Journal*, 60(9):2023–2038, November 1981.

57. T. Ojanpera and R. Prasad. An overview of air interface multiple access for UMTS/IMT-2000. *IEEE Commun. Mag*, 36(9):82–95, 1998.

58. D. A. Pados and S. N. Batalama. Low-complexity blind detection of DS/CDMA signals: Auxiliary-vector receivers. *IEEE Transactions on Communications*, 45(12):1586–1594, December 1997.

59. N. Prasad and M. K. Varanasi. Analysis of decision feedback detection for MIMO Rayleigh-fading channels and the optimization of power and rate allocations. *IEEE Transactions on Information Theory*, 50(6):1009–1025, June 2004.

60. J. G. Proakis and M. Salehi. *Digital Communications*. McGraw-Hill, 5th edition, 2008.

61. G. S. Rajappan and M. L. Honig. Signature sequence adaptation for DS-CDMA with multipath. *IEEE Journal on Selected Areas in Communications*, 20(2):384–395, February 2002.

62. D. J. Ryan, I. B. Collings, and I. V. L. Clarkson. GLRT-Optimal noncoherent lattice decoding. *IEEE Transactions on Signal Processing*, 55(7 Part 2):3773–3786, 2007.

63. J. Salz. Digital transmission over cross-coupled linear channels. *AT&T Technical Journal*, 64(6):1147–1159, July 1985.

64. C. Sankaran and A. Ephremides. Solving a class of optimum multiuser detection problems with polynomial complexity. *IEEE Transactions on Information Theory*, 44(5): 1958–1961, 1998.

65. L. L. Scharf. The SVD and reduced rank signal processing. *Signal Processing*, 25(2):113–133, 1991.

66. C. Schlegel and A. Grant. Polynomial complexity optimal detection of certain multiple-access systems. *IEEE Transactions on Information Theory*, 46(6):2246–2248, 2000.

67. K. S. Schneider. Optimum detection of code division multiplexed signals. *IEEE Transactions on Aerospace and Electronic Systems*, 15(1):181–185, January 1979.

68. R. Singh and L. B. Milstein. Interference suppression for DS/CDMA. *IEEE Transactions on Communications*, 47(3):446–453, March 1999.

69. N. Slonimsky. *Lexicon of Musical Invective: Critical Assaults on Composers since Beethoven's Time*. University of Washington Press, 1987.

70. P. Spasojevic and C. N. Georghiades. The slowest descent method and its application to sequence estimation. *IEEE Transactions on Communications*, 49(9):1592–1604, 2001.

71. V. Srivastava, J. Neel, A. B. Mackenzie, R. Menon, L. A. Dasilva, J. E. Hicks, J. H. Reed, and R. P. Gilles. Using game theory to analyze wireless ad hoc networks. *IEEE Communications Surveys & Tutorials*, 7(4):46–56, 2005.

72. T. Starr, J. M. Cioffi, and P. J. Silverman. Understanding digital subscriber line technology. *Prentice Hall Communications Engineering and Emerging Technologies Series*, 1999.

73. Y. Sun. *Transmitter and Receiver Techniques for Wireless Fading Channels*. PhD thesis, Northwestern University, Evanston, IL, December 2004.

74. Y. Sun and M. L. Honig. Performance of reduced-rank equalization. *IEEE Transactions on Information Theory*, 52(10):4548–4562, October 2005.

75. P. H. Tan and L. K. Rasmussen. Multiuser detection in CDMA—a comparison of relaxations, exact, and heuristic search methods. *IEEE Transactions on Wireless Communications*, 3(5):1802–1809, 2004.

76. P. H. Tan, L. K. Rasmussen, and T. J. Lim. Constrained maximum-likelihood detection in CDMA. *IEEE Transactions on Communications*, 49(1):142–153, 2001.

77. E. Telatar. Capacity of multi-antenna gaussian channels. *Euro. Trans. Telecommun.*, 10:585–595, Nov.–Dec. 1999.
78. L. F. Trichard, J. S. Evans, and I. B. Collings. Large system analysis of linear multistage parallel interference cancellation. *IEEE Transactions on Communications*, 50(11):1778–1786, 2002.
79. D. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
80. W. Van Etten. Maximum likelihood receiver for multiple channel transmission systems. *IEEE Transactions on Communications*, pages 276–283, February 1976.
81. B. D. Van Veen and K. M. Buckley. Beamforming: a versatile approach to spatial filtering. *IEEE Signal Processing Magazine*, 5(2):4–24, 1988.
82. M. K. Varanasi. Noncoherent detection in asynchronous multiuser channels. *IEEE Transactions on Information Theory*, 39(1):157–176, 1993.
83. M. K. Varanasi. Decision feedback multiuser detection: a systematic approach. *IEEE Transactions on Information Theory*, 45(l):2l9–240, January 1999.
84. M. K. Varanasi and B. Aazhang. Multistage detection in asynchronous code-division multiple-access communications. *IEEE Transactions on Communications*, 38(4): 509–519, April 1990.
85. M. K. Varanasi and A. Russ. Noncoherent decorrelative detection for nonorthogonal multi-pulsemodulation over the multiuser Gaussian channel. *IEEE Transactions on Communications*, 46(12):1675–1684, 1998.
86. S. Verdri. Minimum probability of error for asynchronous Gaussian multiple-access channels. *IEEE Transactions on Information Theory*, 32(l):85–96, January 1986.
87. S. Verdú. *Multiuser Detection*. Cambridge University Press, 1998.
88. S. Verdú and S. Shamai. Spectral efficiency of CDMA with random spreading. *IEEE Transactions on Information Theory*, 45(2):622–640, March 1999.
89. H. Vikalo and B. Hassibi. On the sphere-decoding algorithm II. Generalizations, second-order statistics, and applications to communications. *IEEE Transactions on Signal Processing, [see also IEEE Transactions on Acoustics, Speech, and Signal Processing]*, 53(8 Part l):2819–2834, 2005.
90. H. Vikalo, B. Hassibi, and T. Kailath. Iterative decoding for MIMO channels via modified sphere decoding. *IEEE Transactions on Wireless Communications*, 3(6):2299–2311, 2004.
91. E. Visotsky and U. Madhow. Noncoherent multiuser detection for CDMA systems with nonlinearmodulation: a non-Bayesian approach. *IEEE Transactions on Information Theory*, 47(4):1352–1367, 2001.
92. D. Warrier and U. Madhow. Noncoherent communication in space and time. *IEEE Transactions on Information Theory*, 48:651–668, 2002.
93. M. E. Weippert, J. D. Hiemstra, J. S. Goldstein, and M. D. Zoltowski. Insights from the relationship between the multistage Wiener filter and the method of conjugate gradients. *Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2002*, pages 388–392, August 2002.
94. G. Woodward, R. Ratasuk, M. L. Honig, and P. B. Rapajic. Minimum mean-squared error multiuser decision-feedback detectors for DS-CDMA. *IEEE Transactions on Communications*, 50(12):2104–2112, December 2002.

95. J. K. Zhang, T. N. Davidson, and K. M. Wong. Uniform decomposition of mutual information using MMSE decision feedback detection. *IEEE International Symposium on Information Theory*, pages 714–718, September 2005.

96. J. K. Zhang, A. Kavtic, and K. M. Wong. Equal-diagonal QR decomposition and its application to precoder design for successive-cancellation detection. *IEEE Transactions on Information Theory*, 51(1):154–172, January 2005.