Wireless Multi-user Communications

Mérouane Debbah
Alcatel-Lucent Chair on Flexible Radio
SUPELEC
Plateau de Moulon
3 rue Joliot-Curie
B.P. 91192
GIF SUR YVETTE CEDEX, France
merouane.debbah@supelec.fr.

ABSTRACT. Until recently, one-way single link transmission was considered i.e where a single transmitter and a single receiver are involved. However, practical applications are almost always based on multilink multiway systems, where a (possibly very large) number of transmitters and receivers exchange information through some shared transmission media. We refer to these systems as *multiuser systems*, where a "user" denotes an abstract entity responsible of communicating over the system (both transmitting to and receiving from one or more terminals). In these notes, we will introduce some wireless multi-user communication features and discuss their performance. CDMA is used as a benchmark case in the sense that all other vector channel cases (Virtual MIMO, cooperative communications, ad-hoc networks,..) can be treated in this setting with some slight modifications.

1 Introduction

We distinguish between point-to-point networks and systems in which users share a common transmission resource (see Fig. 1). Point-to-point networks are defined by a set of one-way or two-way links between *nodes*. Problems related to these type of networks are topology, routing, switching, congestion control, etc... [1]. These types of networks are based on cooperative communications: since a direct link between a given transmitter-receiver pair may not exist, intermediate nodes act as relays in order to allow the desired communication between any sender-receiver pair. Shared-resource systems are simply defined by sets of transmitters and receivers all connected to the same physical channel. Therefore, a direct link between any transmitter and any receiver is always possible. However, users generate mutual interference that may impair reliable communication. Then, some *multiple access* technique must be used to cope with the destructive effect of interference.

Most wireless communication systems (in particular, *cellular* systems) belong to the class of shared-resource systems, where the shared resource is the system radio channel. A key concept of wireless networks is the exploitation of the spatial component. Because of attenuation (a rapidly increasing function of the distance between transmitter and receiver) two users sufficiently separated in space can share the same bandwidth without causing mutual interference. This consideration leads to the so called "frequency reuse" in cellular systems: the total system

Debbah, M. (2007) Wireless Multi-User Communications. In *Emerging Wireless Technologies* (pp. 5-1 – 5-22). Educational Notes RTO-EN-IST-070, Paper 5. Neuilly-sur-Seine, France: RTO. Available from: http://www.rto.nato.int.

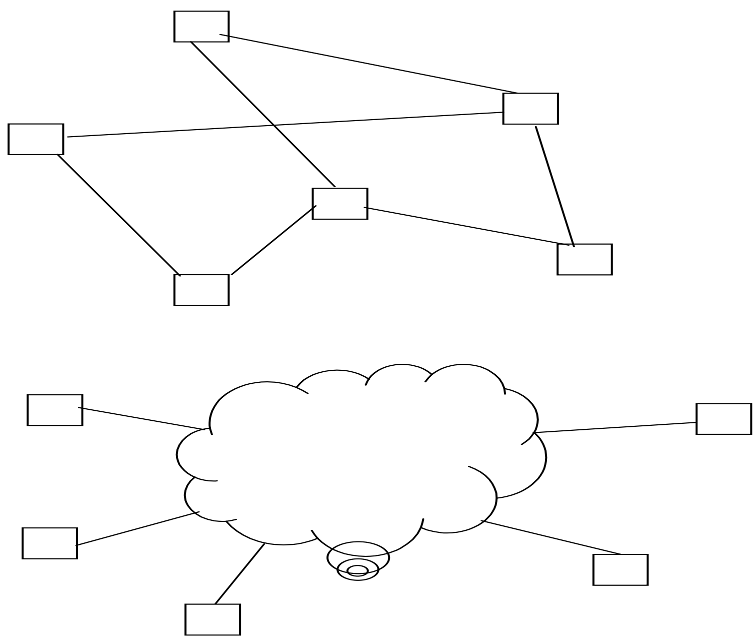

Figure 1: Point-to-point versus shared-resource network topologies.

bandwidth is partitioned into m disjoint subbands. These are allocated to the cells in a way such that two cells reusing the same subband have a given minimum space separation that guarantees sufficiently low mutual interference. The choice of the optimal reuse factor m depends on the type of modulation and coding used for transmission and on the environment propagation characteristics.

2 Duplexing and multiple access

In cellular systems we identify two sets of terminals: base stations (BS) and mobile terminals (MT). The link from BS to MT is called *downlink* and the link from MT to BS is called *uplink*, and the way of sharing the transmission resource between uplink and downlink is called *duplexing*. There are two main duplexing schemes: frequency-division duplexing (FDD) and time-division duplexing (TDD).

In FDD, uplink and downlink are assigned to two separate frequency bands. In this way, no particular synchronization between terminals is needed in order to avoid interference between uplink and downlink. However, operating ad different frequency bands requires normally a more complicated hardware. In TDD, uplink and downlink are assigned to the same frequency band, but to different time slots. In order to avoid interference between uplink and downlink, all terminals must be approximately synchronized to a common time reference (need for BSs synchronization). Also, MTs located in different positions in their cell have different propagation delays. In order to be able to align their uplink slots, their relative delay must be very small with respect to the slot duration. Therefore, TDD can be effectively implemented only for cells

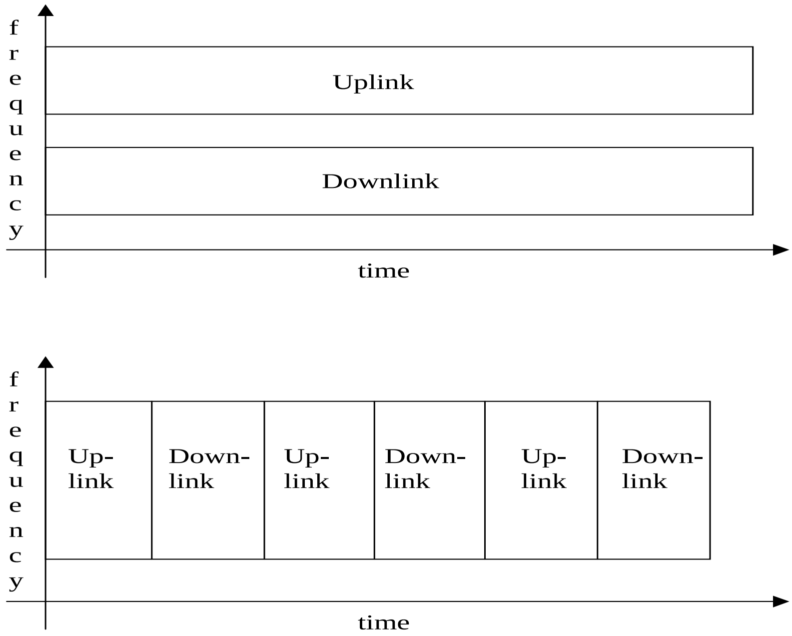

Figure 2: FDD and TDD schemes on the time-frequency plane.

of small size (indoor, picocells, microcells). Fig. 2 gives a graphical representation of FDD and TDD on the time-frequency plane.

Let K be the number of users (i.e., active communication links) in either directions. Multiple access techniques can be divided into *orthogonal* and *non-orthogonal*.

2.1 Orthogonal multiple access

Multiple Access Interference (MAI) is avoided by making the users orthogonal in frequency (frequency division multiple access, or FDMA) in time (time-division multiple access, or TDMA) in time-frequency (TDMA with frequency hopping, or FH-TDMA) or, more in general, in the signal space (orthogonal code-division multiple access, or orthogonal CDMA).

In FDMA, users are assigned to K different subbands. In TDMA, users are assigned to K different time slots. In orthogonal CDMA, a set of K orthogonal signals $\{s_1(t), \ldots, s_K(t)\}$ is chosen and signal $s_k(t)$ is assigned to user k ($s_k(t)$ is referred to as the *signature* of user k). Each user k transmit a linearly modulated signal in the form:

$$x_k(t) = \sum_n a_{k,n} s_k(t - nT) \tag{1}$$

where $a_{k,n}$ are complex modulation symbols. If the users are synchronous and their propagation channel is just a constant multiplicative gain c_k, the received signal is given by

$$y(t) = \sum_{k=1}^{K} c_k x_k(t) + \nu(t)$$

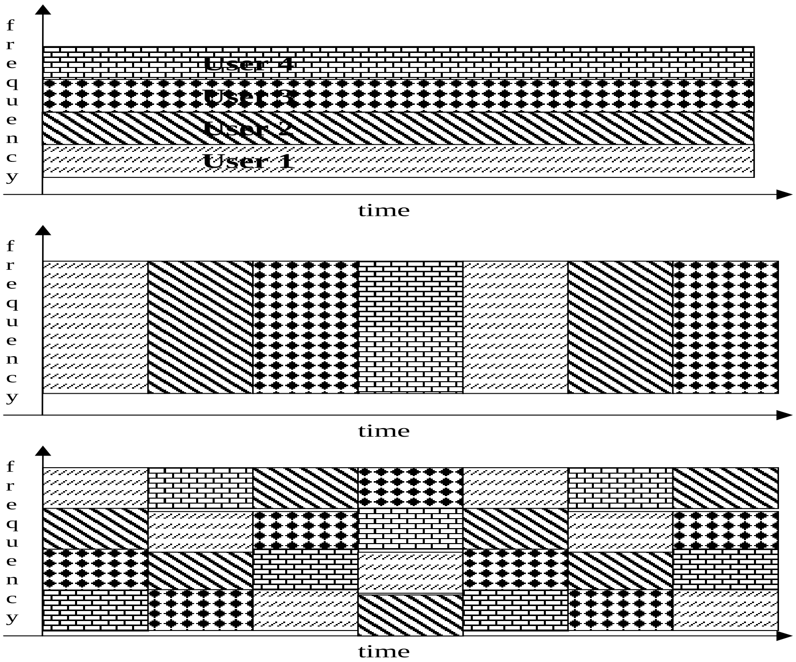

Figure 3: FDMA, TDMA and FH-TDMA in the time-frequency plane, with $K = 4$ users.

where $\nu(t)$ is the background AWGN, with power spectral density N_0. The receiver for user k is based on a filter matched to user k only, with impulse response $\frac{1}{\sqrt{\mathcal{E}_k}} s_k(-t)^*$, where $\mathcal{E}_k = \|s_k(t)\|^2$. Assuming the following conditions:

1. Zero ISI: $\int s_k(t)s_k(t-nT)^*dt = \mathcal{E}_k\delta_{n,0}$.

2. Zero MAI: $\int s_k(t)s_j(t-nT)^*dt = \mathcal{E}_k\delta_{k,j}$.

the output of a filter matched to the k-th user signature waveform, sampled at the symbol rate, yields

$$y_{k,n} = \sqrt{\mathcal{E}_k} c_k a_{k,n} + \nu_{k,n}$$

where $\nu_{k,n} \sim \mathcal{N}_{\mathbb{C}}(0, N_0)$.

2.2 Non-orthogonal multiple access

Non-orthogonal multiple access schemes do not try to avoid MAI but cope with it and "separate" the user information messages by using signal processing at the receiver, channel coding and random access protocols, or any combination of these. In this chapter we consider only signal processing and channel coding techniques (see [1] for a presentation of random access protocols).

Non-orthogonal multiple access can be seen as a generalization of orthogonal CDMA. User signals have still the form (1), but the user signature waveforms are no longer mutually orthogonal. The waveforms $s_k(t)$ are chosen to have a large time-bandwidth product, so that these systems are normally referred to as *spread-spectrum* and the access scheme is called Spread-Spectrum Multiple Access (SSMA).

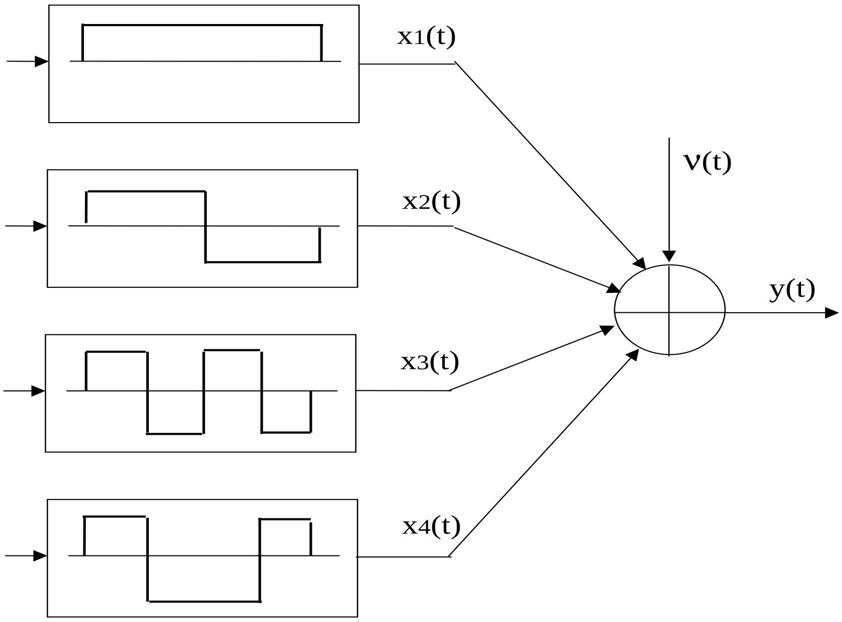

Figure 4: An example of orthogonal CDMA system with $K = 4$ users.

The continuous-time received signal is given by

$$y(t) = \sum_{k=1}^{K} \int c_k(t, \tau) x_k(t - \tau) d\tau + \nu(t) \tag{2}$$

Under the usual assumption of discrete multipath and slowly-varying fading we have

$$\int c_k(t, \tau) x_k(t - \tau) d\tau = \sum_{n} \sum_{p=0}^{P-1} c_{k,p,n} a_{k,n} s_k(t - nT - \tau_{k,p}) \tag{3}$$

where $c_{k,p,n}$ is the p-th path gain of user k channel around the n-th symbol interval and $\tau_{k,p}$ is the p-th path delay of user k channel.

A technique for SSMA particularly important for wireless cellular applications is Direct-Sequence CDMA (DS-CDMA) [2] (for example, this is the technique used in IS-95 and in the UMTS standards). Fig. 5 represents a typical block diagram of a DS-CDMA transmission scheme for a generic user k. A channel encoder produces code words $\mathbf{x}_k$. Code words are interleaved, as in single-user transmission. The sequence of interleaved symbols $a_{k,n}$ is fed into a repetition encoder of length L, so that

$$\ldots, a_{k,n-1}, a_{k,n}, a_{k,n+1}, \ldots$$

is turned into

$$\ldots, \underbrace{a_{k,n-1}, \ldots, a_{k,n-1}}_{L \text{ times}}, \underbrace{a_{k,n}, \ldots, a_{k,n}}_{L \text{ times}}, \underbrace{a_{k,n+1}, \ldots, a_{k,n+1}}_{L \text{ times}}, \ldots$$

Repeated symbols are called "chips". The chips produced by repetition encoding are multiplied chipwise by a pseudo-random sequence of chips $s_{k,\ell,n}$, for $n \in \mathbb{Z}$ and $\ell = 0, \dots, L-1$, called *spreading sequence* or *signature sequence* of user k. The ℓ-th chip of the n-th symbol interval of the resulting sequence is given by

$$b_{k,\ell,n} = a_{k,n} s_{k,\ell,n} \tag{4}$$

Finally, the sequence $b_{k,\ell,n}$ is linearly modulated to generate the transmit signal complex complex envelope

$$x_k(t) = \sum_n \sum_{\ell=0}^{L-1} b_{k,\ell,n} \psi(t - nT - \ell T_c) \tag{5}$$

where $\psi(t)$ is the *chip-shaping pulse*, T is the symbol interval and $T_c = T/L$ is the chip interval. Normally, $\psi(t)$ satisfies the Nyquist criterion with respect to the chip interval, i.e., $\int \psi(t)\psi(t - nT_c)dt = \delta_{0,n}$, where we assume $\|\psi(t)\|^2 = 1$.

Spreading sequences might be periodic or aperiodic. In the case of *periodic spreading*, $s_{k,\ell,n} = s_{k,\ell}$ for all $n \in \mathbb{Z}$ and user k is identified by the sequence of length L $\mathbf{s}_k = (s_{k,0}, \dots, s_{k,L-1})^T$. In this case, $x_k(t)$ can be written as

$$x_k(t) = \sum_n a_{k,n} s_k(t - nT) \tag{6}$$

where

$$s_k(t) = \sum_{\ell=0}^{L-1} s_{k,\ell} \psi(t - \ell T_c) \tag{7}$$

Since the two-sided bandwidth of $\psi(t)$ is about $1/T_c$ and the duration of $s_k(t)$ is of the order of $LT_c = T$, the time-bandwidth product of signature signals is $\approx T/T_c = L$. In usual DS-CDMA, $L \gg 1$ (e.g., from 16 to 256 and more), and the system is Spread Spectrum.

3 DS-CDMA with conventional detection

For simplicity, we treat the case of periodic spreading. Generalization to aperiodic spreading is trivial. The received signal given by the superposition of K DS-CDMA users plus AWGN is given by (see (2) and (3))

$$y(t) = \sum_{k=1}^{K} \sum_n a_{k,n} \sum_{p=0}^{P-1} c_{k,p,n} s_k(t - nT - \tau_{k,p}) + \nu(t) \tag{8}$$

Let

$$\psi'_k(n,t) = \sum_{p=0}^{P-1} c_{k,p}[n] s_k(t - \tau_{k,p})$$

and focus on the detection of user 1 (our reference user). A suboptimal but simple way to detect user 1 symbols is to treat all other users as Gaussian white background noise and use a matched filter for user 1 only. This approach is referred to as *single-user matched filter* (SUMF) detector. The filter matched to user 1 waveform is

$$f_1(n,t) = \psi'_1(n,-t)^* = \sum_{p=0}^{P-1} c^*_{1,p,n} s_1(-t - \tau_{1,p})^*$$

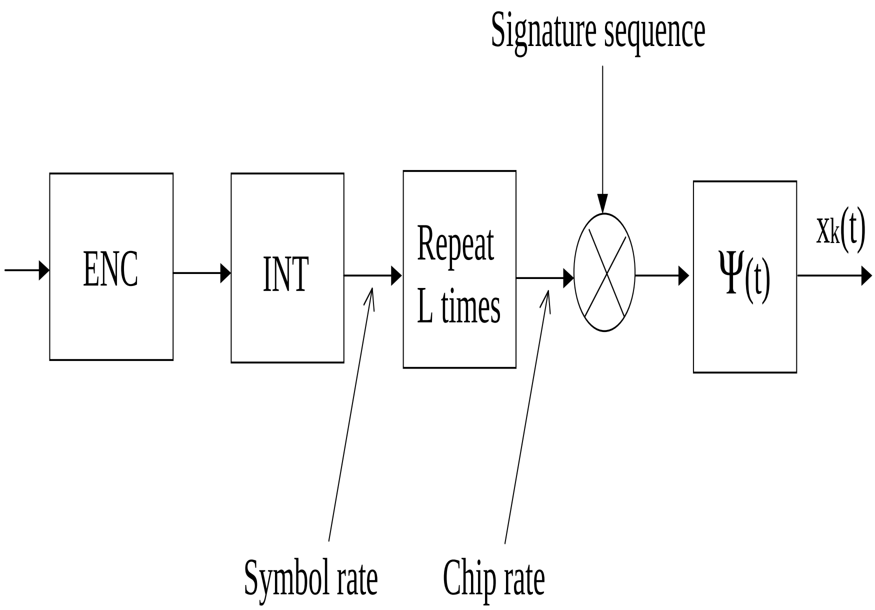

Figure 5: Block diagram of a general DS-CDMA modulator.

The SUMF output, sampled at the symbol rate, is given by

$$\begin{aligned} y_{1,n} &= \int y(t)\psi_1'(t-nT)^* dt \\ &= \sum_{p=0}^{P-1} c_{1,p,n}^* \int y(t)s_1(t-nT-\tau_{1,p})^* dt \\ &= \sum_{p=0}^{P-1} c_{1,p,n}^* \sum_{\ell=0}^{L-1} s_{1,\ell}^* \int y(t)\psi(t-nT-\ell T_c-\tau_{1,p})^* dt \end{aligned} \quad (9)$$

From the last line of the above equation, we see that $y_{1,n}$ can be computed from the samples output by a filter matched to the chip-shaping pulse $\psi(t)$ taken at $nT + \ell T + \tau_{1,p}$, for all $n \in \mathbb{Z}$, $\ell = 0, \ldots, L-1$ and $p = 0, \ldots, P-1$. Then, the DS-CDMA signal format yields the implementation of the rake receiver represented in Fig. 6.

First, the chip-matched filter is sampled at rate N_s/T_c (N_s samples per chip). For large N_s (typically, $N_s = 4$), the arbitrary path delays $\tau_{1,p}$ can be well approximated as

$$\tau_{1,p} \approx M_p T_c + m_p T_c/N_s$$

where $M_p = \lfloor \tau_{1,p}/T_c \rfloor$ and where $m \in \{0, \ldots, N_s - 1\}$. For each rake finger p, the "integer" and "fractional" parts of the delays ($M_p T_c$ and mT_c/N_s, respectively) are handled differently. The integer part is obtained by delaying the generation of the spreading sequence by M_p chips. For periodic spreading, this means a cyclic shift of the local spreading sequence by M_p positions.

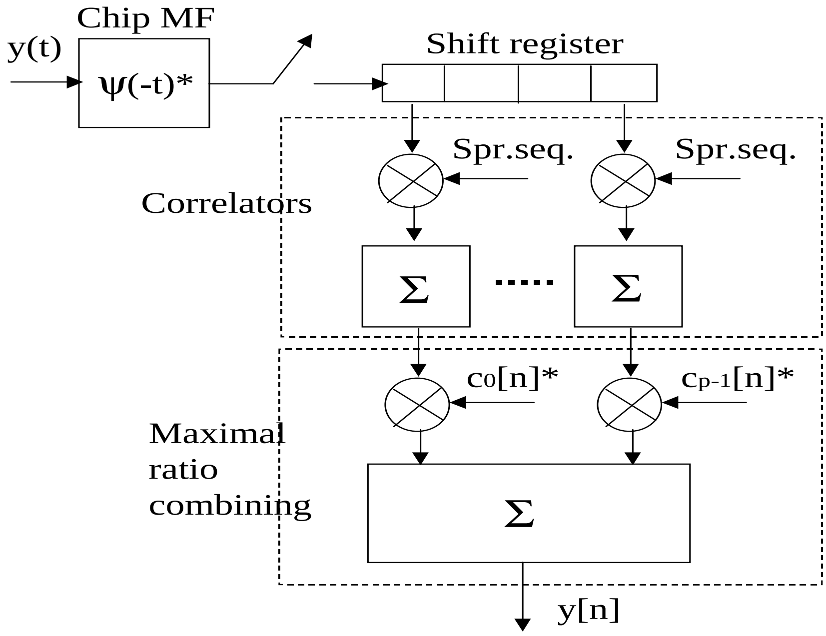

Figure 6: Rake receiver for DS-CDMA.

The fractional part is achieved by a shift register of $N_s - 1$ memory elements. The p-th finger reads the correlator input data from the m_p-th memory, at the chip rate.

Then, the p-th finger correlates the chip-rate samples with the spreading sequence appropriately shifted by M_p positions. The correlator output is

$$y_{1,p,n} \approx \sum_{\ell=0}^{L-1} s_{1,\ell}^* \int y(t)\psi(t - nT - \ell T_c - \tau_{1,p})^* dt$$

Finally, the P finger outputs are combined via MRC, in order to obtain

$$y_{1,n} = \sum_{p=0}^{P-1} c_{1,p,n}^* y_{1,p,n}$$

DS-CDMA with conventional SUMF detection is usually analyzed by making a Gaussian approximation of the interference term at the rake output and by neglecting the self-interference of the useful wignal with the delayed versions of itself. In brief, we assume that user k signal received through path p is seen by the receiver of user 1 as an independent white additive Gaussian noise with power spectral density $|c_{k,p,n}|^2 \mathcal{E}_k/L$. Notice the key role of the processing gain L: without fading, each user is received with power $\mathcal{P}_k = \mathcal{E}_k/T$ over a bandwidth of approximately $W = L/T$ Hz, because of spreading. Then, the power spectral density of each user is approximately flat and equal to $\mathcal{E}_k/L$. This explains why, for large L, DS-CDMA modulation makes users look like extra additive white noise to each other. Subject to the above assumption, the interference term at each rake finger output is Gaussian with variance

$$N_I = \frac{1}{L}\sum_{k=2}^{K}\mathcal{E}_k\sum_{p=0}^{P-1} g_{k,p,n}$$

where $g_{k,p,n} = |c_{k,p,n}|^2$ is the power gain of the p-th path of user k channel. The SINR at the SUMF output (after rake combining) is given by

$$\text{SINR}_1 = \frac{\mathcal{E}_1 \sum_{p=0}^{P-1} g_{1,p,n}}{N_0 + \frac{1}{L}\sum_{k=2}^{K}\mathcal{E}_k\sum_{p=0}^{P-1} g_{k,p,n}} \tag{10}$$

Normally, DS-CDMA is considered for large systems, i.e., for systems with large L and K. We define the system loading factor $\beta = K/L$ to be the number of users per chip. From the law of large numbers, under mild convergence conditions we have the limit in probability

$$\lim_{K\to\infty}\frac{1}{K}\sum_{k=2}^{K}\mathcal{E}_k\sum_{p=0}^{P-1} g_{k,p,n} = \sum_{p=0}^{P-1} E[\mathcal{E}_k g_{k,p,n}]$$

where expectation is with respect to the ensemble of the users in the system and with respect to the (common) fading statistics. Then, for large systems we have

$$\text{SINR}_1 = \frac{\mathcal{E}_1}{N_0 + \beta\mathcal{E}_I}\sum_{p=0}^{P-1} g_{1,p,n} \tag{11}$$

where $\mathcal{E}_I = \sum_{p=0}^{P-1} E[\mathcal{E}_k g_{k,p,n}]$ is the average received energy per symbol from an interfering user. We conclude that with the Gaussian approximation and the assumption of large system, the performance analysis of any (coded) modulation scheme follows exactly the same lines developed in the previous chapters for single-user systems by replacing the noise power spectral density N_0 with an equivalent noise plus interference power spectral density $I_0 = N_0 + \beta\mathcal{E}_I$.

It is apparent that non-orthogonal CDMA with conventional SUMF detection is interference-limited, in the sense that even if $N_0 \to 0$, the SINR at the SUMF output for each user remains finite. Therefore, error probability cannot be decreased beyond a certain limit by simply increasing the transmitted power of all users.

4 Multiuser detection

In order to make a DS-CDMA not interference-limited, we have to exploit the structure of interference at the receiver. The ensemble of signal-processing and decoding techniques which exploit the structure of the interfering signals in order to combat the MAI is usually known as *Multiuser Detection* (MUD) [3]. Here, we present the main algorithms in the simple case of a synchronous DS-CDMA system with periodic spreading. This system model has mainly a theoretical interest but serves as the basis for developing any sort of MUD algorithms. Of course, these algorithms should then be extended to handle the time-varing multipath asynchronous environment given in (8), typical of wireless systems.

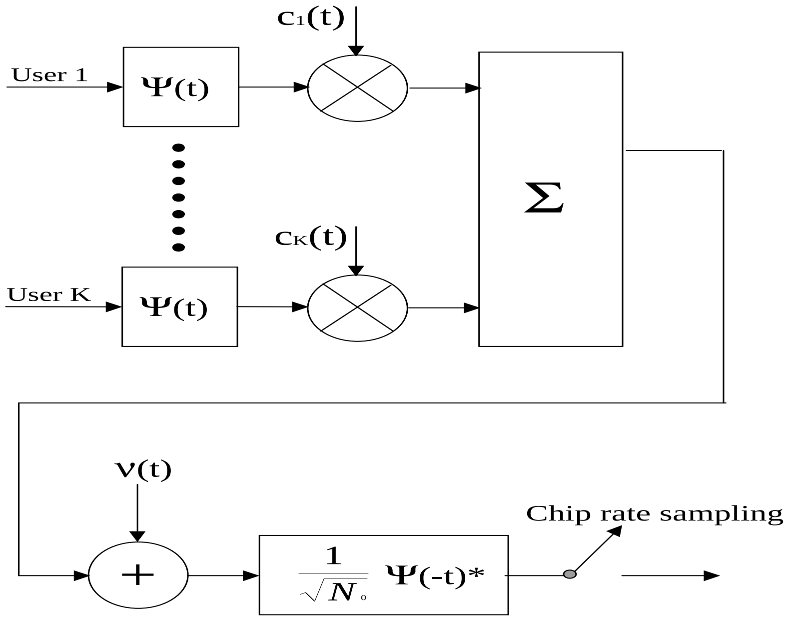

Figure 7: Chip-synchronous flat-fading DS-CDMA channel model.

4.1 Canonical DS-CDMA model

Consider a DS-CDMA system with K users, synchronous transmission and flat fading. The received signal can be written as

$$y(t) = \sum_{k=1}^{K} \sqrt{\mathcal{E}_k} \sum_{n} c_{k,n} a_{k,n} s_k(t - nT) + \nu(t) \tag{12}$$

where the signature waveforms are given by (7). The average transmitted energy per symbols of user k is $\mathcal{E}_k$.

In this case, an optimal receiver front-end is provided by a chip-matched filter $\frac{1}{\sqrt{N_0}}\psi(-t)^*$ whose output is sampled at the chip-rate $1/T_c$ (see Fig. 7). The resulting discrete-time channel can be written in vector form by collecting the L samples corresponding to each symbol interval. We define the matrix $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_K] \in \mathbb{C}^{L \times K}$ containing the spreading sequences and the diagonal matrix of complex channel amplitudes $\mathbf{W}_n = \text{diag}(w_{1,n}, \dots, w_{K,n})$, where $w_{k,n} = c_{k,n}\sqrt{\mathcal{E}_k/N_0}$. Then, the received vector during the n-th symbol interval is

$$\mathbf{y}_n = \mathbf{S}\mathbf{W}_n\mathbf{a}_n + \boldsymbol{\nu}_n \tag{13}$$

where $\boldsymbol{\nu}_n \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I})$.

4.2 ML multiuser joint decoder

Each user k transmits independently selected code words $\mathbf{x}_k$ of a given channel code $\mathcal{C}_k$ of length N. We let $\mathbf{X}$ be the $K \times N$ matrix obtained by arranging the K user code words by row (we assume block-synchronous transmission). Then, the vector $\mathbf{a}_n$ of modulation symbols in (13) is the n-th column of $\mathbf{X}$.

Assuming perfect knowledge of all user spreading sequences and instantaneous SNRs, the ML joint sequence detector is given by

$$\widehat{\mathbf{X}} = \arg \min_{\mathbf{X} \in \mathcal{C}} \sum_{n=1}^{N} |\mathbf{y}_n - \mathbf{S}\mathbf{W}_n \mathbf{a}_n|^2 \tag{14}$$

where $\mathcal{C} = \mathcal{C}_1 \times \ldots \times \mathcal{C}_K$ is the Cartesian-product of all user codes. In principle, the above decoder can be implemented by a Viterbi Algorithm acting on the Cartesian-product trellis of the user codes. The complexity of the ML joint decoder is $O(\prod_{k=1}^{K} M_k)$ where M_k is the number of trellis states of code $\mathcal{C}_k$. Since complexity is essentially exponential in the number of users, the joint ML decoder is not suited to practical implementations in typical CDMA systems, where K is generally large.

4.3 Symbol-by-symbol ML and MAP detectors

Most literature on MUD considered the suboptimal receiver scheme of Fig. 8, where a symbol-by-symbol (sbs) detector produces "estimates" $\widetilde{a}_{k,n}$ of the user modulation symbols, that are fed to a bank of independent channel decoders [3]. The simplest way to make use of symbol estimates $\widetilde{a}_{k,n}$ is to produce sbs hard decisions $\widehat{a}_{k,n}$. Next, we focus on the n-th symbol interval and we drop index n for the sake of notation simplicity.

The main performance measures proposed for sbs-MUD are the asymptotic multiuser efficiency (AME), the near-far resistance (NFR) and the output SINR. AME and NFR are asymptotic (for high SNR) performance measures that apply to hard-decisions at the detector output. We let $\gamma_k = |w_k|^2$ be the instantaneous SNR of user k. For a given k, let $P_k^{\rm su}(\gamma_k)$ be the probability of error of the optimal single-user detector for user k in the absence of MAI, for SNR equal to γ_k, and let $P_k(\gamma_1, \ldots, \gamma_K)$ be the probability of error of a given sbs-MUD detector for user k. The AME is defined as the SNR asymptotic penalty factor payed by the sbs-MUD with respect to the single-user optimal performance in the limit for $N_0 \to 0$. This is given by

$$\eta_k = \sup \left\{ r \in [0,1] \ : \ \lim_{N_0 \to 0} \frac{P_k(\gamma_1, \ldots, \gamma_K)}{P_k^{\rm su}(r\gamma_k)} = 0 \right\}$$

The NFR is given by the worst-case AME over all possible γ_j for $j \neq k$, i.e.,

$$\bar{\eta}_k = \inf_{\gamma_j \in \mathbb{R}: j \neq k} \eta_k$$

In order words, the NFR measures the asymptotic SNR penalty for the worst-case interferers SNRs. A detector for which $\bar{\eta}_k > 0$ is said to be *near-far resistant*. In this case, the error probability for user k can be made arbitrarily small for any choice of the other users power, simply by increasing user k power. If $\bar{\eta}_k = 0$, the receiver is said to be near-far non-resistant, and user k performance is *interference-limited*.

If the multiuser detector provides soft estimates $\widetilde{a}_k$ of a_k, error probability at the detector output is rather meaningless since normally these soft-estimates are passed to a soft-decoding algorithm without making hard decisions. In this case, a more meaningful performance measure is the output SINR, defined by

$$\text{SINR}_k = \frac{|E[\widetilde{a}_k | a_k]|^2}{\text{var}(\widetilde{a}_k | a_k)}$$

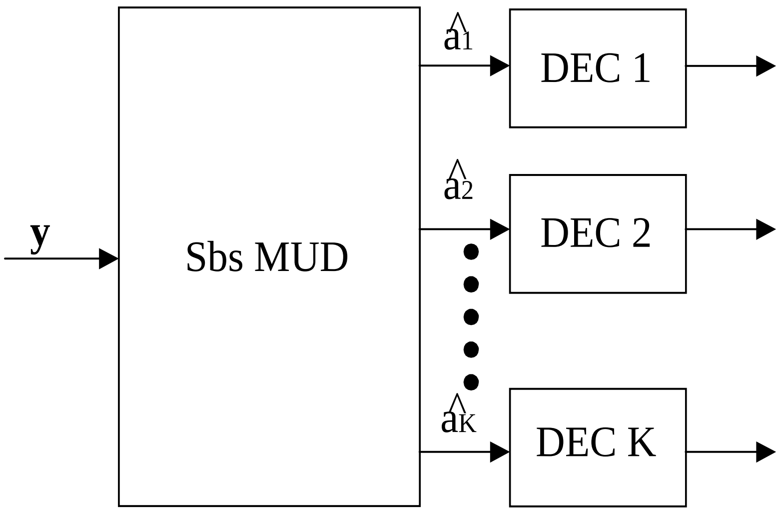

Figure 8: Symbol-by-symbol MUD followed by single-user decoders

where $\mathrm{var}(\widetilde{a}_k|a_k) = E[|\widetilde{a}_k|^2|a_k] - |E[\widetilde{a}_k|a_k]|^2$ is the conditional variance of $\widetilde{a}_k$ given a_k.

The ML sbs detector is based on the decision rule

$$\widehat{\mathbf{a}} = \arg \min_{\mathbf{a}\in\mathcal{A}^K} |\mathbf{y} - \mathbf{SWa}|^2 \tag{15}$$

where $\mathcal{A}$ is the modulation symbol alphabet (assumed common to all users). This detector provides intrinsically hard decisions and minimizes the probability that $\widehat{\mathbf{a}} \neq \mathbf{a}$, assuming that all users symbols are i.i.d. over $\mathcal{A}$.

A different approach is to minimize the probability of error for every user k by the sbs MAP rule

$$\widehat{a}_k = \arg \max_{a\in\mathcal{A}} \mathrm{APP}_k(a) \tag{16}$$

where the *a posteriori probability* $\mathrm{APP}_k(a)$ is given by

$$\begin{aligned}
\mathrm{APP}_k(a) &= \Pr(a_k = a|\mathbf{y}) \\
&= \frac{p(\mathbf{y}|a_k = a)\Pr(a_k = a)}{\sum_{b\in\mathcal{A}} p(\mathbf{y}|a_k = b)\Pr(a_k = b)} \\
&\propto p_k(a) \sum_{\substack{\mathbf{a}\in\mathcal{A}^K \\ a_k = a}} p(\mathbf{y}|\mathbf{a}) \prod_{j\neq k} p_j(a_j) \\
&\propto p_k(a) \sum_{\substack{\mathbf{a}\in\mathcal{A}^K \\ a_k = a}} \exp\left(-|\mathbf{y} - \mathbf{SWa}|^2\right) \prod_{j\neq k} p_j(a_j)
\end{aligned} \tag{17}$$

where $p_j(a_j)$ is the a priori probability of the j-th user symbol (the a priori joint distribution of the user symbols is assumed to be in product form, i.e., the symbols are a priori independent). It is well-known that, if the a priori symbol distribution is uniform for all symbols, the following approximation holds for high SNR

$$\log \text{APP}_k(a) \approx -\min_{\substack{\mathbf{a}\in\mathcal{A}^K \\ a_k=a}} |\mathbf{y} - \mathbf{SWa}|^2 + \text{const.}$$

Therefore, maximizing $\log \text{APP}_k(a)$ over $a \in \mathcal{A}$ for all k yields approximately the same result as minimizing $|\mathbf{y} - \mathbf{SWa}|^2$ over $\mathbf{a} \in \mathcal{A}^K$. However, the MAP sbs detector is able to provide also *soft reliability information* about the symbol decisions in the form of APPs (equivalently, logarithms of APP, or log-likelihood ratios if data are binary). In particular, the a posteriori minimum-mean square error (MMSE) estimate $\widetilde{a}_k$ of a_k for the APP distribution $\text{APP}_k(a)$ is given by the conditional mean

$$\widetilde{a}_k = E[a_k|\mathbf{y}] = \sum_{a\in\mathcal{A}} a\text{APP}_k(a) \tag{18}$$

The complexity of both the ML and the MAP detectors is still exponential in the number of users K, in general.

Consider the pairwise error probability $P(\mathbf{a} \to \mathbf{a}')$, where $\mathbf{a}'$ differs from $\mathbf{a}$ in the k-th position. We have

$$\begin{aligned} P(\mathbf{a} \to \mathbf{a}') &= \Pr\left(|\mathbf{y} - \mathbf{SWa}'|^2 - |\mathbf{y} - \mathbf{SWa}|^2 < 0 \middle| \mathbf{a}\right) \\ &= \Pr\left(2\text{Re}\{\boldsymbol{\nu}^H \mathbf{SWd}\} > \mathbf{d}^H \mathbf{W}^H \mathbf{S}^H \mathbf{SWd}\right) \\ &= Q\left(\sqrt{\frac{1}{2}\mathbf{d}^H \mathbf{W}^H \mathbf{RWd}}\right) \end{aligned} \tag{19}$$

where we let $\mathbf{R} = \mathbf{S}^H\mathbf{S}$ and $\mathbf{d} = \mathbf{a}' - \mathbf{a}$. By using a modification of the union bounding technique it can be shown [3] that there exists finite constants C_L and C_U such that

$$C_L Q(\sqrt{\eta_k \gamma_k |d_k|^2/2}) \le P_k(\gamma_1, \ldots, \gamma_K) \le C_U Q(\sqrt{\eta_k \gamma_k |d_k|^2/2})$$

where

$$\eta_k = \min_{\mathbf{d}\in\mathcal{D}_k} \mathbf{d}^H \mathbf{A}^H \mathbf{RAd} \tag{20}$$

In the above expression, $\mathcal{D}_k$ denotes the set of all normalized difference vectors $\mathbf{d} = (\mathbf{a}' - \mathbf{a})/|d_k|$ such that $d_k \neq 0$ and $\mathbf{A} = \mathbf{W}/\sqrt{\gamma_k}$ (notice that, by definition, the vector $\mathbf{d}$ has k-th element with magnitude 1 and the matrix $\mathbf{A}$ is diagonal with the k-th diagonal element of magnitude 1).

The AME of the ML sbs detector is difficult to calculate since it involves the minimization of a (non-negative definite) quadratic form over a discrete and finite set of vectors. However, the corresponding NFR is surprisingly easy to find. We have

$$\begin{aligned} \bar{\eta}_k &= \inf_{\mathbf{A}} \min_{\mathbf{d}\in\mathcal{D}_k} \mathbf{d}^H \mathbf{A}^H \mathbf{RAd} \\ &= \inf_{\substack{\mathbf{v}\in\mathbb{C}^K \\ v_k=1}} \mathbf{v}^H \mathbf{Rv} \\ &= \min_{\mathbf{u}\in\mathbb{C}^{K-1}} \left\{1 + 2\text{Re}\{\mathbf{r}_k^H \mathbf{u}\} + \mathbf{u}^H \mathbf{R}_k \mathbf{u}\right\} \end{aligned} \tag{21}$$

where $\mathbf{R}_k$ is the $(K-1)\times(K-1)$ matrix obtained by eliminating the k-th row and column from $\mathbf{R}$ and $\mathbf{r}_k$ is the vector of length $K-1$ obtained from the k-th column of $\mathbf{R}$ by eliminating the k-th element (equal to 1). The quadratic form in the last line of (21) is minimized by $\mathbf{u} = -\mathbf{R}_k^{-1}\mathbf{r}_k$. By substituting this into (21) we obtain the resulting NFR

$$\bar{\eta}_k = 1 - \mathbf{r}_k^H \mathbf{R}_k^{-1} \mathbf{r}_k \tag{22}$$

By using the 2×2 block matrix inversion lemma we can show that

$$\bar{\eta}_k = \frac{1}{[\mathbf{R}^{-1}]_{k,k}} \tag{23}$$

Also, the NFR can be expressed in terms of the normalized signature sequences as

$$\bar{\eta}_k = \mathbf{s}_k^H[\mathbf{I} - \mathbf{S}_k(\mathbf{S}_k^H\mathbf{S}_k)^{-1}\mathbf{S}_k^H]\mathbf{s}_k \tag{24}$$

where $\mathbf{S}_k$ is the $L\times(K-1)$ matrix obtained from $\mathbf{S}$ by eliminating its k-th column. In order to verify the above formula, notice that $\mathbf{s}_k^H\mathbf{s}_k = 1$, $\mathbf{r}_k = \mathbf{S}_k^H\mathbf{s}_k$ and $\mathbf{R}_k = \mathbf{S}_k^H\mathbf{S}_k$, and use (22).

It is easy to verify that $\bar{\eta}_k = 0$ if and only if $\mathbf{s}_k$ is contained in the linear span of the other users, i.e., in the column space of $\mathbf{S}_k$. In all other cases, the ML sbs detector is near-far resistant.

4.4 Linear multiuser detectors

Because of the exponential complexity of the ML and MAP sbs detectors it is meaningful to investigate suboptimal low-complexity alternatives which still provide optimum NFR. In particular, *linear* MUD for user k is defined by a filter with coefficients vector $\mathbf{f}_k$ whose output

$$z_k = \mathbf{f}_k^H \mathbf{y} \tag{25}$$

is used to detect user k symbol. The SUMF corresponds to the particular choice $\mathbf{f}_k = \mathbf{s}_k$. The output z_k of the filter is a soft estimate of the user symbol a_k. If the filter output is used directly as input of a soft decoder for user k channel, we simply let $\widetilde{a}_k = z_k$. As an alternative, the filter output can be used as the input of a sbs hard detector, defined by

$$\widehat{a}_k = \mathrm{dec}(z_k)$$

where $\mathrm{dec}(\cdot)$ is any detection rule suited for user k modulation symbols. For example, in the case of binary antipidal modulation we have $\widehat{a}_k = \mathrm{sign}(\mathrm{Re}\{z_k\})$.

The SINR at the output of the linear filter is given by

$$\begin{aligned} \mathrm{SINR}_k &= \frac{\gamma_k \mathbf{f}_k^H \mathbf{s}_k \mathbf{s}_k^H \mathbf{f}_k}{\mathbf{f}_k^H(\mathbf{I} + \mathbf{S}_k \mathbf{W}_k \mathbf{W}_k^H \mathbf{S}_k^H)\mathbf{f}_k} \\ &= \frac{\gamma_k |\mathbf{f}_k^H \mathbf{s}_k|^2}{|\mathbf{f}_k|^2 + \sum_{j\neq k} \gamma_j |\mathbf{f}_k^H \mathbf{s}_j|^2} \end{aligned} \tag{26}$$

where $\mathbf{W}_k$ is the $(K-1)\times(K-1)$ diagonal matrix obtained from $\mathbf{W}$ by deleting the k-th row and column. The symbol error probability with hard sbs detection depends on the modulation format and on the choice of $\mathbf{f}_k$. As far as performance analysis is concerned, the Gaussian approximation can be used to model the filter output as a virtual single-user additive noise channel with given SNR equal to SINR_k. By dividing the filter output by $w_k \mathbf{f}_k^H \mathbf{s}_k$ in order to remove the bias, the virtual channel is given by

$$z_k = a_k + \nu_k$$

where $\nu_k \sim \mathcal{N}_{\mathbb{C}}(0, I_0)$ and $I_0 = 1/\text{SINR}_k$.

The two most common and useful design criteria for the filter $\mathbf{f}_k$ are the zero-forcing (ZF) and the minimum MSE (MMSE) criteria. These can be regarded as the analogous of ZF and MMSE linear equalization for ISI channels.

ZF linear multiuser detection. The ZF linear detector (also known as *decorrelator* [3]) projects the received vector $\mathbf{y}$ onto the orthogonal complement of the column span of $\mathbf{S}_k$, and then performs a matched filtering operation in the projected space. The resulting filter is given by

$$\mathbf{f}_k = \alpha(\mathbf{I} - \mathbf{S}_k(\mathbf{S}_k^H \mathbf{S}_k)^{-1} \mathbf{S}_k^H)\mathbf{s}_k \tag{27}$$

where α is any proportionality non-zero constant. It is easy to check that the AME of the ZF linear detector is zero if and only if $\mathbf{s}_k$ is contained in the linear span of the columns of $\mathbf{S}_k$. Otherwise, the AME is independent of γ_j for all $j \neq k$ and it is equal to the optimal NFR $\bar{\eta}_k$ given by (22), (23) and (24).

The output SINR is given by

$$\text{SINR}_k = \bar{\eta}_k \gamma_k \tag{28}$$

Notice that since typically $0 < \bar{\eta}_k < 1$, the ZF linear detector provides SINR degradation with respect to a SUMF operating in the absence of MAI. Since $\bar{\eta}_k$ is independent of the interferers powers γ_j, $j \neq k$, this degradation is present even in the case of very weak MAI. In fact, in the case where the CDMA system is noise-limited (i.e., the MAI power is much less than the background noise power), the ZF linear detector may perform significantly worse than the SUMF. However, in interference-limited conditions (large SNR for all users), the ZF linear detector is near-optimal.

MMSE linear detector. The MMSE linear detector (see [3] and references therein) is the filter $\mathbf{f}_k$ minimizing the MSE

$$\epsilon_k^2 = E[|a_k - z_k|^2]$$

By applying the orthogonality principle, we obtain the system of linear equations

$$E[(a_k - z_k)^* \mathbf{y}] = \mathbf{0}$$

whose solution is given by the well-known linear Wiener filter

$$\mathbf{f}_k = w_k \mathbf{\Sigma}^{-1} \mathbf{s}_k$$

where w_k is the k-th diagonal element of $\mathbf{W}$, and $\mathbf{\Sigma} = E[\mathbf{y}\mathbf{y}^H]$ is the covariance matrix of $\mathbf{y}$. We can obtain an alternative form for the filter $\mathbf{f}_k$ by using the fact that

$$\begin{aligned} \mathbf{\Sigma} &= \mathbf{I} + \mathbf{S}\mathbf{W}\mathbf{W}^H \mathbf{S}^H \\ &= \mathbf{I} + \mathbf{S}_k \mathbf{W}_k \mathbf{W}_k^H \mathbf{S}_k^H + \gamma_k \mathbf{s}_k \mathbf{s}_k^H \\ &= \mathbf{\Sigma}_k + \gamma_k \mathbf{s}_k \mathbf{s}_k^H \end{aligned} \tag{29}$$

where $\mathbf{\Sigma}_k$ is the covariance of the noise plus interference contained in the received signal. By applying the matrix inversion lemma to the last line of (29), we obtain

$$\mathbf{f}_k = \frac{w_k}{1+\mu_k}\mathbf{\Sigma}_k^{-1}\mathbf{s}_k \tag{30}$$

where $\mu_k = \gamma_k \mathbf{s}_k^H \mathbf{\Sigma}_k^{-1}\mathbf{s}_k$ is real and positive. The resulting SINR is immediately obtained as

$$\begin{aligned} \text{SINR}_k &= \frac{\gamma_k \mathbf{f}_k^H \mathbf{s}_k \mathbf{s}_k^H \mathbf{f}_k}{\mathbf{f}_k^H \mathbf{\Sigma}_k \mathbf{f}_k} \\ &= \frac{\mu_k^2/(1+\mu_k)^2}{\mu_k/(1+\mu_k)^2} \\ &= \mu_k \end{aligned} \tag{31}$$

It can be shown that the MMSE linear detector achieves the maximum output SINR among all linear detectors. Moreover, any other filter $\mathbf{f}'_k \propto \mathbf{f}_k$ achieves the same optimal SINR. It can also be shown that the MMSE linear filter converges to the ZF linear filter as $N_0 \to 0$ (therefore, it has the same AME and NFR of ZF). However, in very noisy conditions (i.e., for large N_0) the MMSE linear filter converges to the SUMF. This illustrates intuitively the fact that the MMSE linear detector does not suffer from the noise enhancement effect of the ZF detector.

4.5 Formulation in the matched filter output domain.

MMSE and ZF (decorrelator) detectors can be formulated in a completely equivalent way with respect to the output of a bank of SUMFs, given by

$$\begin{aligned} \mathbf{r} &= \mathbf{S}^H \mathbf{y} \\ &= \mathbf{S}^H \mathbf{S} \mathbf{W} \mathbf{a} + \mathbf{S}^H \boldsymbol{\nu} \\ &= \mathbf{R} \mathbf{W} \mathbf{a} + \mathbf{n} \end{aligned} \tag{32}$$

where, as already defined previously, we let $\mathbf{R} = \mathbf{S}^H\mathbf{S}$ and $\mathbf{n} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, N_0\mathbf{R})$.

Hence, it is easy to show that the output of the ZF multiuser detector $\mathbf{z} = (z_1, \ldots, z_K)^T$ is given, up to a scalar factor, by

$$\mathbf{z} = \mathbf{W}^{-1}\mathbf{R}^{-1}\mathbf{r} \tag{33}$$

and the output of the MMSE linear multiuser detector is given, up to a multiplicative factor, by

$$\mathbf{z} = \left(\mathbf{W}^H \mathbf{R} \mathbf{W} + N_0 \mathbf{I}\right)^{-1} \mathbf{W}^H \mathbf{r} \tag{34}$$

4.6 Multistage approximation of linear MUD

The MMSE and the ZF (decorrelator) linear detectors require the inversion of a $L \times L$ matrix or of a $K \times K$ matrix. For large L and K this might be still too complex for efficient practical implementation.

For the sake of reducing further the computational complexity of the receiver, multistage approximations of linear multiuser detectors have been proposed. As an example, consider the decorrelator formulated in the matched filter output domain (i.e., given by (33)). We let $\widetilde{\mathbf{R}} = \mathbf{R} - \mathbf{I}$, and use the first order Taylor expansion

$$\frac{1}{1+x} = 1 - x + o(x)$$

so that

$$\mathbf{R}^{-1} = (\mathbf{I} + \widetilde{\mathbf{R}})^{-1} \approx \mathbf{I} - \widetilde{\mathbf{R}} = 2\mathbf{I} - \mathbf{R}$$

Therefore, the first-order approximated ZF detector is given by

$$\begin{aligned} \mathbf{z} &= \mathbf{W}^{-1}(2\mathbf{I} - \mathbf{R})\mathbf{r} \\ &= \mathbf{W}^{-1}\left(\mathbf{r} - (\mathbf{S}^H\mathbf{S} - \mathbf{I})\mathbf{r}\right) \end{aligned} \tag{35}$$

The last line has a simple and nice interpretation in terms of *parallel interference cancellation* with *linear* feedback. Indeed, the output of the matched filter back $\mathbf{r}$ can be seen as a soft estimate of the modulation symbols $\mathbf{Wa}$. Indeed, for quasi-orthogonal spreading waveforms and for very high SNR we have

$$\mathbf{r} = \mathbf{S}^H\mathbf{S}\mathbf{W}\mathbf{a} + \mathbf{S}^H\boldsymbol{\nu} \approx \mathbf{W}\mathbf{a}$$

Then, the last line of (35) can be seen as a refinement of this approximation where k-th component of $\mathbf{z}$ is given by

$$z_k = \frac{1}{w_k}\left(r_k - \sum_{j\neq k}[R]_{k,j}r_j\right) \tag{36}$$

The above idea can be extended to any m-th order Taylor expansion. However, truncating an infinite series does not provide the best approximation of a function for a given finite order m. Then, in general the problem of finding the optimal m-th order linear detector can be formulated as follows: given $m > 0$, find the weights $\{\delta_\ell : \ell = 0, \dots, m\}$ such that the linear detector defined by

$$\mathbf{z} = \sum_{\ell=0}^{m} \delta_\ell \left(\mathbf{W}^H\mathbf{R}\mathbf{W}\right)^\ell \mathbf{W}^H\mathbf{r} \tag{37}$$

maximizes the output SINR for all users. Methods for finding the polynomial coefficients δ_ℓ are proposed, for example, in [4, 5, 6].

Once the coefficients are found, the fast implementation of the multistage linear detector can be obtained via banks of *correlators* and re-spreading operations. The key is Horner's rule to evaluate a polynomial:

$$p(x) = \sum_{\ell=0}^{m} \delta_\ell x^\ell = (\delta_0 + x(\delta_1 + x(\delta_2 + \cdots x(\delta_{m-1} + \delta_m x)))\cdots) \tag{38}$$

Then, the output $\mathbf{z}$ of (37) can be written as

$$\mathbf{z} = (((\cdots(\delta_m\mathbf{W}^H\mathbf{S}^H\mathbf{S}\mathbf{W} + \delta_{m-1}\mathbf{I})\mathbf{W}^H\mathbf{S}^H\mathbf{S}\mathbf{W} + \delta_{m-2}\mathbf{I})\cdots + \delta_1\mathbf{I})\mathbf{W}^H\mathbf{S}^H\mathbf{S}\mathbf{W} + \delta_0\mathbf{I})\mathbf{W}^H\mathbf{r} \tag{39}$$

Notice that the multiplication $\mathbf{W}^H\mathbf{S}^H\mathbf{S}\mathbf{W}\mathbf{v}$, where $\mathbf{v}$ is any K-dimensional vector, corresponds to the direct spreading followed by summing and matched filtering block. Finally, the overall m-th order linear detector is implemented by concatenating m such stages.

5 Performance in the random large-system regime

The performance of linear multiuser detectors (in particular, of the SUMF, decorrelator and MMSE filters) is characterized by the SINR at the filter output for any given user k. In general, for a filter $\mathbf{f}_k$ this is given by

$$\text{SINR}_k = \frac{|w_k|^2 |\mathbf{f}_k^H \mathbf{s}_k|^2}{N_0 |\mathbf{f}_k|^2 + \sum_{j \neq k} |w_j|^2 |\mathbf{f}_k^H \mathbf{s}_j|^2} \tag{40}$$

It is clear that, generally speaking, the SINR depends on the received instantaneous user SNRs $\gamma_j = |w_j|^2/N_0$, on the user spreading sequences and on the receiving filter $\mathbf{f}_k$. If the $\mathbf{s}_k$'s are randomly assigned to the users, the SINR is a random variable.

Recently it has been recognized [7, 8, 9] that in the limit of large system, i.e., for large K, L with fixed load $K/L = \beta$, under certain convergence conditions, the SINR at the output of SUMF, decorrelator and MMSE filters converges almost surely to the value

$$\text{SINR}_k = \gamma_k \eta \tag{41}$$

where η is a deterministic constant that depends on the system parameters and on the receiver. This result is of fundamental importance since it allows the characterization of the performance of any given user k in terms of its own individual SNR γ_k via a proportionality factor that depends on the system.

Assumptions.

1. The spreading sequences $\mathbf{s}_k$ have i.i.d. elements $s_{\ell,k} = \frac{1}{\sqrt{L}} v_{\ell,k}$, where $v_{\ell,k}$ are i.i.d. random variables with mean zero, variance 1, and finite fourth order moment.

2. Let $F^{(K)}(z)$ denote the empirical cdf of the user received SNRs, given by

$$F^{(K)}(z) = \frac{1}{K} \sum_{k=1}^{K} 1\{\gamma_k \leq z\} \tag{42}$$

   We assume that as $K \to \infty$, $F^{(K)}(z) \to F(z)$, where $F(z)$ is a given distribution function.

Main results. As $K \to \infty$ with $K/L = \beta$, we have

1. The SINR at the output of the k-th user SUMF converges almost surely to

$$\text{SINR}_k = \frac{\gamma_k}{1 + \beta \int_0^\infty z dF(z)} \tag{43}$$

2. The SINR at the output of the k-th user decorrelator, for $0 \leq \beta < 1$, converges almost surely to

$$\text{SINR}_k = (1 - \beta)\gamma_k \tag{44}$$

3. The SINR at the output of the k-th user MMSE filter converges almost surely to the quantity $\text{SINR}_k = \gamma_k \eta^{\text{mmse}}$, where η^{mmse} is the unique non-negative solution of the equation

$$\eta = \frac{1}{1 + \beta \int_0^\infty \frac{z}{1+\eta z} dF(z)} \tag{45}$$

Sketch of proof. The first fact from the theory of large random matrices that we shall use in order to prove the above results is the following lemma and its corollary:

Lemma. Let $\mathbf{v}$ and $\mathbf{A}$ be a L-dimensional random vector and a $L \times L$ dimensional random matrix, mutually independent, where the elements of $\mathbf{v}$ are i.i.d., with mean zero, variance 1 and finite fourth order moment and $\mathbf{A}$ has bounded spectral radius, for all L. Then, the limit

$$\lim_{L\to\infty} \left| \frac{1}{L}\mathbf{v}^H\mathbf{A}\mathbf{v} - \frac{1}{L}\mathrm{tr}(\mathbf{A}) \right| = 0$$

with probability 1.

Corollary. Suppose that the sequence of Hermitian symmetric matrices $\mathbf{A}$, for increasing L, has a limiting eigenvalue distribution, i.e., the empirical cdf of the eigenvalues $\{\lambda_1, \ldots, \lambda_L\}$, defined by

$$G^{(L)}(z) = \frac{1}{L}\sum_{\ell=1}^{L} 1\{\lambda_\ell \leq z\}$$

converges with probability 1 to a fixed distribution $G(z)$, then with the same assumptions as above,

$$\lim_{L\to\infty} \frac{1}{L}\mathbf{v}^H\mathbf{A}\mathbf{v} = E[\lambda] = \int_{-\infty}^{\infty} z dG(z)$$

□

Given a cdf $F(z)$, the *Stieltjes transform* of $F(z)$ is defined by

$$m_F(s) = \int \frac{1}{s+z} dF(z)$$

in the complex half-plane $\mathrm{Im}\{s\} < 0$. The Stieltjes transform specifies uniquely the associated distribution $F(z)$, and viceversa. The main fact from the theory of large random matrices that we use in the proof is the following:

Theorem [10]. Let $\mathbf{A} \in \mathbb{C}^{L\times K}$ be a matrix of i.i.d. random variables $a_{i,j} = v_{i,j}/\sqrt{L}$, such that $v_{i,j}$ has mean zero, variance 1 and finite fourth order moment, and let $\mathbf{\Gamma}$ be a $K \times K$ diagonal matrix with real elements $\gamma_1, \ldots, \gamma_K$. Assume that, as $K \to \infty$ with $K/L = \beta$, the empirical distribution of the γ_k's, defined in (42) converges to a non-random fixed distribution $F(z)$. Then, the empirical distribution of the eigenvalues of the matrix $\mathbf{M} = \mathbf{A}\mathbf{\Gamma}\mathbf{A}^H$ converges with probability 1 to the non-random limit $G(z)$, where the Stieltjes transform $m_G(s)$ of $G(z)$ is given implicitly by the solution of the equation

$$m_G(s) = \frac{1}{s + \beta \int \frac{z}{1+zm_G(s)} dF(z)} \tag{46}$$

□

Now, we start proving the large-system result for the SUMF. The SUMF is defined by the filter $\mathbf{f}_k = \mathbf{s}_k$. The noise plus MAI covariance matrix is given by

$$N_0 \left(\mathbf{I} + \mathbf{S}_k\mathbf{\Gamma}_k\mathbf{S}_k^H\right) = N_0\mathbf{\Sigma}_k$$

where $\mathbf{S}_k$ contains as columns all $\mathbf{s}_j$ but $\mathbf{s}_k$, and $\mathbf{\Gamma}_k$ contains on the diagonal all γ_j's but γ_k. Then, the SINR can be written as

$$\text{SINR}_k = \frac{\gamma_k |\mathbf{s}_k|^2}{\mathbf{s}_k^H \mathbf{\Sigma}_k \mathbf{s}_k} \tag{47}$$

Notice that $\mathbf{\Sigma}_k$ does not contain $\mathbf{s}_k$, therefore, $\mathbf{s}_k$ and $\mathbf{\Sigma}_k$ are independent. Hence, by the trace lemma, we have

$$\lim_{L\to\infty} \mathbf{s}_k^H \mathbf{\Sigma}_k \mathbf{s}_k = \lim_{L\to\infty} \frac{1}{L} \text{tr}(\mathbf{\Sigma}_k)$$

where $G(z)$ is the limiting distribution of the eigenvalues of $\mathbf{\Sigma}_k$.

By inspection, we see that all eigenvalues of $\mathbf{\Sigma}_k$ are given by $1 + \lambda_\ell$, where $\{\lambda_\ell\}$ are the eigenvalues of $\mathbf{S}_k \mathbf{\Gamma}_k \mathbf{S}_k^H$. Hence, we have

$$\begin{aligned}
\lim_{L\to\infty} \frac{1}{L} \text{tr}(\mathbf{\Sigma}_k) &= 1 + \lim_{L\to\infty} \frac{1}{L} \text{tr}(\mathbf{S}_k \mathbf{\Gamma}_k \mathbf{S}_k^H) \\
&= 1 + \beta \lim_{K\to\infty} \frac{1}{K} \text{tr}(\mathbf{S}_k^H \mathbf{S}_k \mathbf{\Gamma}_k) \\
&= 1 + \beta \lim_{K\to\infty} \frac{1}{K} \sum_{j \neq k} |\mathbf{s}_j|^2 \gamma_j \\
&= 1 + \beta \int z dF(z)
\end{aligned}$$

where the last line follows from the strong law of large numbers, for which

$$\lim_{L\to\infty} |\mathbf{s}_j|^2 = \lim_{L\to\infty} \frac{1}{L} \sum_{\ell=1}^{L} |v_{\ell,j}|^2 = E[|v_{\ell,j}|^2] = 1$$

For the same reason, the numerator in (47) converges with probability 1 to γ_k. Therefore, (43) is proved.

Now, consider the SINR at the output of the decorrelator, given by (see (28)), $\text{SINR}_k = \bar{\eta}_k \gamma_k$, where

$$\bar{\eta}_k = \mathbf{s}_k^H [\mathbf{I} - \mathbf{S}_k (\mathbf{S}_k^H \mathbf{S}_k)^{-1} \mathbf{S}_k] \mathbf{s}_k \tag{48}$$

The matrix $\mathbf{P}_k^\perp = [\mathbf{I} - \mathbf{S}_k (\mathbf{S}_k^H \mathbf{S}_k)^{-1} \mathbf{S}_k]$ is the orthogonal projector on the subspace orthogonal complement of the subspace spanned by interference. It is well-known that all orthogonal projectors have eigenvalues either equal to one or equal to zero, and the number of non-zero eigenvalues is equal to the dimension of the subspace. Hence, by using the trace lemma we have

$$\begin{aligned}
\lim_{L\to\infty} \bar{\eta}_k &= \lim_{L\to\infty} \frac{1}{L} \text{tr}(\mathbf{P}_k^\perp) \\
&= \lim_{L\to\infty} \frac{1}{L} \sum_{\ell=1}^{L-K+1} 1 \\
&= 1 - \beta
\end{aligned}$$

where we assumed that $\beta < 1$, otherwise the projector is not defined. Then, (44) is proved.

Finally, for the linear MMSE we use the SINR formula (see (31)), $\text{SINR}_k = \gamma_k \mathbf{s}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{s}_k$. From the trace lemma we have

$$\begin{aligned} \lim_{L\to\infty} \mathbf{s}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{s}_k &= \lim_{L\to\infty} \frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{1+\lambda_\ell} \\ &= \int \frac{1}{1+z} dG(z) \\ &= m_G(1) \end{aligned}$$

where again $\{\lambda_\ell\}$ are the eigenvalues of $\mathbf{S}_k \mathbf{\Gamma}_k \mathbf{S}_k^H$, we used the definition of Stieltjes transform and where $G(z)$ is the limiting eigenvalue distribution of $\mathbf{S}_k \mathbf{\Gamma}_k \mathbf{S}_k^H$. From the Theorem, $m_G(1)$ is given by the solution of

$$m_G(1) = \frac{1}{1 + \beta \int \frac{z}{1+z m_G(1)} dF(z)}$$

Hence, we have shown that in the limit for large L, K the SINR is given by $\gamma_k \eta$, where $\eta = m_G(1)$ is the solution of the above equation, as desired.

References

[1] D. Bertsekas and R. Gallager, *Data Networks, 2nd ed.*, Prentice-Hall, Upper Saddle River, NJ, 1987.

[2] A. J. Viterbi, *CDMA - Principles of spread-spectrum communications*, Addison-Wesley, Reading, MA, 1995.

[3] S. Verdu, *Multiuser detection*, Cambridge University Press, Cambridge, UK, 1998.

[4] R. Mueller and S. Verdu, "Design and analysis of low-complexity interference mitigation on vector channels," *IEEE J. Select. Areas Commun.*, vol. 19, no. 8, pp. 1429–1441, August 2001.

[5] S. Moshavi, E. Kanterakis, and D. Schilling, "Multistage linear receivers for DS-CDMA systems," *Int. J. wireless Inform. Networks*, vol. 3, no. 1, pp. 1–17, January 1996.

[6] D. Guo, L. Rasmussen, and T. Lim, "Linear parallel interference cancellation in long-code CDMA multiuser detection," *IEEE J. Select. Areas Commun.*, vol. 17, no. 6, pp. 2074–2081, December 1999.

[7] D. Tse and S. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and capacity," *IEEE Trans. on Inform. Theory*, vol. 45, no. 2, pp. 641–675, March 1999.

[8] S. Verdu and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. on Inform. Theory*, vol. 45, no. 2, pp. 622–640, March 1999.

[9] S. Verdu and S. Shamai, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. on Inform. Theory*, vol. 47, no. 4, pp. 1302–1327, May 2001.

[10] J. Silverstein, *Eigenvalues and eigenvectors of large dimensional sample covariance matrices, in : Random Matrices and Theory Applications. Ed: J. Cohen, H. Kesten and C. Newman*, American Math. Soc., Providence (RI), 1986.