

1 Joint Antenna Combining and Multiuser Detection

Ralf Müller[†] and Laura Cottatellucci[†]

Smart antenna technology combined with code-division multiple-access (CDMA) is common in 3rd generation cellular communication systems. The uplink of a cellular CDMA system where each mobile and each base station is equipped with multiple element antennas, is exemplarily explored. Multiple reception due to multiple receive antennas and multiple reception due to symbol repetition over time—this is commonly referred to as spreading—are identified as two peculiarities of the same concept. It is shown that joint processing of antenna signals and multi-user interference is superior to a separated approach in terms of performance. Utilizing multistage detection and certain properties of random matrices, antenna combining and multi-user detection can be implemented jointly without the need for matrix multiplications or matrix inversions. Hereby, the complexity per bit of the presented algorithms scales linearly with the number of users. —

1.1 INTRODUCTION

In recent years, intense efforts have been devoted to two important lines of works: multiuser techniques for direct sequence CDMA systems and signal processing techniques in systems with antenna arrays. Multiuser receivers mitigate the cross-talk between users taking into account the structure of interference from other users. Typically, these techniques exploit the degrees of freedom given by the frequency diversity intrinsically present in CDMA systems. Antenna arrays provide spatial diversity and smart antennas make use of these degrees of freedom to enhance the system capacity. The resource pooling result, due to Hanly and Tse [1], enlightens the fact that degrees of freedom in space and frequency are interchangeable. Moreover, the total number of degrees of freedom is the product of the degrees of freedom in space and frequency. A system with spreading factor N and N_R receive antennas is in many respects equivalent to a system with a single antenna and spreading factor NN_R . This suggests the idea to treat the two effects in

[†]Forschungszentrum Telekommunikation Wien, Vienna, Austria

the same way performing antenna array processing and multiuser detection jointly. Joint processing outperforms techniques that try to exploit separately the degree of freedom in space and in frequency significantly. However, the optimal algorithms for this task are known for prohibitive complexity. A joint array processing and multiuser detection approach implementable in real-time systems is offered by linear multistage detectors with asymptotic weighting: they reach an excellent trade-off between performance and complexity. In fact, they have a complexity per transmitted bit which is linear in the number of transmitting antennas and in the number of users—the same is true for a single user matched filter—and attain almost the same performance as the joint linear minimum mean-squared error receiver.

In Section 1.2, we introduce the system model. Sections 1.3 and 1.4 stress the duality between degrees of freedom in space and frequency and explain the resource pooling result. Section 1.5 shows why the separation of antenna array processing and multiuser detection is a severely suboptimum approach by means of an intuitive example. Linear multistage detectors are defined in Section 1.6. Their structure and their properties are described in Section 1.7. In Section 1.8, we introduce asymptotic weighting—this approach works excellently for systems with many users—to overcome the weight design problem which multistage detectors can suffer from. The state of art about multistage receivers with large systems weighting is summarized in Section 1.10.

1.2 SYSTEM MODEL

Consider the uplink (reverse link) of an asynchronous CDMA system where each of the K users is equipped with N_T transmitting antenna elements, the base station receiver is equipped with N_R receiving antenna elements, the channel is frequency-selective with impulse responses $h_{k,p,q}(t)$, $1 \leq k \leq K$, $1 \leq p \leq N_T$, $1 \leq q \leq N_R$ and impaired by additive white Gaussian noise $v_q(t)$. Transmitting a block of M symbols and denoting the spreading sequence of user k at antenna element p and symbol time m by $s_{k,p,m}[n]$, $0 \leq n \leq N - 1$ and the respective data symbol by $a_{k,p}[m]$, $0 \leq m \leq M - 1$, the signal transmitted by user k from antenna element q is given (in complex baseband notation) by

$$b_{k,p}(t) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \psi(t - nT_c - mNT_c) a_{k,p}[m] s_{k,p,m}[n] \quad (1.1)$$

with T_c denoting the time duration of one chip and $\psi(t)$ denoting the chip waveform. The latter combines the actual chip waveform, often a rectangular waveform, and the band-limiting influence of the pulse-shaping filter.

The received signal at antenna element q can be written in continuous time as

$$r_q(t) = \sum_{k=1}^K \sum_{p=1}^{N_T} \int_{-\infty}^{+\infty} b_{k,p}(t-\tau) h_{k,p,q}(\tau) d\tau + v_q(t). \quad (1.2)$$

At the receiver front end, it is passed through a chip matched filter¹ $\psi^*(-t)$ giving

$$\tilde{r}_q(t) = \int_{-\infty}^{+\infty} r_q(\tau) \psi^*(\tau-t) d\tau. \quad (1.3)$$

In order to get sufficient statistics after sampling at the chip rate, the signal $\tilde{r}_q(t)$ should be passed through a bank of KN_T filters matched to the channel impulse responses $h_{k,p,q}(t)$, $1 \leq k \leq K$, $1 \leq p \leq N_T$, before being sampled. Alternatively, sufficient statistics could be obtained by oversampling $\tilde{r}_q(t)$ with a rate sufficiently large to exploit the band-limitation of the chip waveform.

The first approach, matched filtering with the channel impulse response, is difficult on wireless channels due to the time-variant nature of the channel impulse response and the difficulties in channel estimation in the continuous time domain. The second approach, oversampling, is the method of choice for implementations of UMTS. For chip waveforms with roll-off factor (bandwidth extension factor) $\alpha < \frac{1}{2}$, threefold oversampling is sufficient.

In the oversampled discrete-time domain, the matched filter bank can be implemented by KN_T digital filters per receive antenna, one for each user and each transmit antenna. Their output signals in chip time n are given by

$$y_{k,p,q}[n] = \int_{-\infty}^{+\infty} \tilde{r}_q(\tau) h_{k,p,q}^*(\tau - nT_c) d\tau \quad (1.4)$$

for user k , transmit antenna p , receive antenna q , and chip time $1 - L \leq n \leq NM + L - 2$ with L denoting the length of the longest channel impulse response in chips.

The discrete-time signal representation in (1.4) is very redundant. It contains replicas of the data symbols in time due to spreading—spreading is only a repetition code with time-variant mapping of the code symbols to the signal constellation—and multiple receive antennas. The MKN_T data symbols $a_{k,p}[m]$, $1 \leq k \leq K$, $1 \leq p \leq N_T$, $0 \leq m \leq M - 1$ are represented by the $(MN + 2L - 1)KN_T N_R$ received signal samples $y_{k,p,q}[n]$, $1 \leq k \leq K$, $1 \leq p \leq N_T$, $1 \leq q \leq N_R$, $1 - L \leq n \leq NM + L - 2$. Without loss of information, the redundant signal dimensions can be eliminated by spatial and

¹In practice, the chip matched filter is often implemented in discrete-time via over-sampling.

4 JOINT ANTENNA COMBINING AND MULTIUSER DETECTION

temporal matched filtering

$$x_{k,p}[m] = \sum_{n=1}^N s_{k,p,m}^*[n + mN] \sum_{q=1}^{N_R} y_{k,p,q}[n + mN]. \quad (1.5)$$

Note that there is no weighting of the signals of different receive antennas in (1.5), as the correct weighting with the channel impulse responses has already taken place in (1.4).

The overall communication channel from the data symbols $a_{k,p}[m]$ to the symbol-time sufficient statistics $x_{k,p}[m]$ is a discrete-time KN_T -dimensional additive noise channel. It is canonically described in vector notation by

$$\mathbf{x}[m] = \sum_{\ell=-l}^l \Phi_{\ell}[m] \mathbf{a}[\ell] + \mathbf{v}[m] \quad (1.6)$$

with $l = \lceil (L - 1)/N \rceil$,

$$\mathbf{x}[m] = \begin{bmatrix} x_{1,1}[m] \\ \vdots \\ x_{K,1}[m] \\ x_{1,2}[m] \\ \vdots \\ \vdots \\ x_{K,N_T}[m] \end{bmatrix} \quad \text{and} \quad \mathbf{a}[m] = \begin{bmatrix} a_{1,1}[m] \\ \vdots \\ a_{K,1}[m] \\ a_{1,2}[m] \\ \vdots \\ \vdots \\ a_{K,N_T}[m] \end{bmatrix} \quad (1.7)$$

where $\Phi_m[\ell]$ is a matrix-valued time-variant weighting function and $\mathbf{v}[m]$ is spatially and temporally correlated Gaussian noise.

1.3 SPREADING VS. RECEIVE ANTENNAS

The descriptions of the multiuser MIMO channels (1.5) and (1.4) in both symbol-time and chip-time are very involved and little illustrative. More light onto the effect of the characterizing parameters K , N_T , N_R , N is set by the concept of resource-pooling discovered by Hanly and Tse [1]. For the purpose of explaining the concept, consider the following two simplifications: Let the auto-correlation function of the chip waveforms fulfill the Nyquist criterion and the impulse responses satisfy

$$h_{k,p,q}(t) = h_{k,p,q} \delta(t) \quad \forall k, p, q. \quad (1.8)$$

The latter condition means that the channels of all users are frequency-flat and the users are synchronized.

In discrete time notation the simplifications introduced above result in

$$\mathbf{x}[m] = \mathbf{Q}[m]^H \mathbf{Q}[m] \mathbf{a}[m] + \mathbf{v}[m] \quad (1.9)$$

with the $NN_R \times KN_T$ virtual spreading matrix

$$\mathbf{Q}[m] = \begin{bmatrix} \mathbf{S}_1[m] \mathbf{H}_{1,1} & \mathbf{S}_2[m] \mathbf{H}_{2,1} & \cdots & \mathbf{S}_{N_T}[m] \mathbf{H}_{N_T,1} \\ \mathbf{S}_1[m] \mathbf{H}_{1,2} & \mathbf{S}_2[m] \mathbf{H}_{2,2} & \cdots & \mathbf{S}_{N_T}[m] \mathbf{H}_{N_T,2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_1[m] \mathbf{H}_{1,N_R} & \mathbf{S}_2[m] \mathbf{H}_{2,N_R} & \cdots & \mathbf{S}_{N_T}[m] \mathbf{H}_{N_T,N_R} \end{bmatrix} \quad (1.10)$$

composed of the $N \times K$ true spreading matrices

$$\mathbf{S}_p[m] = \begin{bmatrix} s_{1,p,m}[0] & s_{2,p,m}[0] & \cdots & s_{K,p,m}[0] \\ s_{1,p,m}[1] & s_{2,p,m}[1] & \cdots & s_{K,p,m}[1] \\ \vdots & \vdots & \ddots & \vdots \\ s_{1,p,m}[N-1] & s_{2,p,m}[N-1] & \cdots & s_{K,p,m}[N-1] \end{bmatrix} \quad (1.11)$$

depending on symbol time m and transmit antenna p , the diagonal channel weight matrices $\mathbf{H}_{p,q} = \text{diag}(h_{1,p,q}, h_{2,p,q}, \dots, h_{K,p,q})$ depending on transmit and receive antenna indices p and q , and the vector-valued additive temporally white Gaussian noise $\mathbf{v}[m] \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}[m]^H \mathbf{Q}[m] \sigma^2)$.

Compared to a communication channel where only a single element antennas is used, the structure of the mathematical formulation of the channel has not changed. The effect of multiple transmit antennas is to blow up the user dimension of the spreading matrix by a factor equal to the number of transmit antenna. The effect of multiple receive antennas is to blow up the spreading factor dimension of the spreading matrix by a factor equal to the number of receive antenna. None of these two effects is surprising: additional transmit antennas create additional data signals in the same way as additional users do, provided that the data signals at the multiple transmit antennas are statistically independent (an assumption which is violated if space-time codes are used). Additional receive antennas create replicas of the same data disturbed by independent noise samples (if the noise is spatially white) in the same way as spreading creates replicas of the same data over time which are also disturbed by independent noise samples if the noise is temporally white. Unless the expansion of the spreading matrix into a virtual spreading matrix by multiple transmit and receive antennas has a significant effect on the crosscorrelation properties of the (virtual) spreading sequences, spreading and multiple receive antennas must be seen as different peculiarities of the same concept. Transmit antennas and multiple user are then related to each other in precisely the same way.

It is hard to investigate how much crosscorrelation properties change when a spreading matrix is blown up into virtual spreading matrix for general spreading sequences. Obviously, if the spreading sequences are orthogonal,

the virtual ones will not be. In UMTS, however, spreading sequences of different users are not orthogonal, but pseudo-random sequences. Therefore, it is more natural and of higher practical importance to consider random spreading sequences as done in the following.

1.4 RESOURCE POOLING FOR RANDOM SPREADING

Let the chips of the true spreading sequences be jointly independent and identically distributed random variables with zero mean and variance $1/N$. Let the channel weights of all users corresponding to all antenna pairs be independent random variables. Let the channel weights belonging to the same user be identically distributed with zero mean and variance P_k . These assumptions translate into the following properties of the system:

- Pseudo-random spreading or scrambling sequences are used. The sequences for different transmit antennas of the same user are different—if they were not, a severe drawback in performance would occur.
- The dual antenna array channels are spatially uncorrelated and without line-of-sight component.
- The receiving antennas are located at the same base station, i.e. there is no soft hand-off.

In addition, we will assume that the system is large, i.e. there are many users $K \gg 1$ and the spreading factor $N \gg 1$ is large, but the load

$$\alpha = \frac{K}{N} \quad (1.12)$$

is fixed. The large system assumption is accurate up to fractions of decibels in predicted performance for spreading factors of about $N = 32$ or greater.

For a channel like (1.9), many performance measures, such as channel capacity with side information and signal-to-interference and noise ratios of the most popular linear multiuser detectors, depend mainly or even solely on the singular values of the channel matrix and the signal-to-noise ratio. Showing that two channels have channel matrices with the same singular value distribution is therefore a strong indicator, in some cases even a proof, that the two channels are equivalent in terms of their capabilities to transport information.

In order to state the resource pooling result more precisely, let us denote the singular values of any $m \times n$ matrix \mathbf{A} by $\sigma_1(\mathbf{A}) \leq \sigma_2(\mathbf{A}) \leq \dots \leq \sigma_{\min\{m,n\}}(\mathbf{A})$ and the singular value distribution of the matrix \mathbf{A} by

$$F_{\mathbf{A}}(x) = \frac{1}{\min\{m,n\}} \sum_{i: \sigma_i(\mathbf{A}) \leq x} 1. \quad (1.13)$$

Resource Pooling: Let the number of users K grow large for fixed load α and fixed number of transmit and receive antennas N_T and N_R , respectively, then the following singular value distributions converge with probability 1 to the same common singular value distribution $F_\infty(x)$

$$F_{\mathbf{Q}[m]}(x) \longrightarrow F_\infty(x) \quad \forall m \quad (1.14)$$

$$F_{\tilde{\mathbf{S}}\tilde{\mathbf{H}}_p}(x) \longrightarrow F_\infty(x) \quad \forall p \quad (1.15)$$

where the $NN_R \times KN_T$ matrix $\tilde{\mathbf{S}}$ is composed of independent identically distributed entries with zero mean and variance $1/(NN_R)$ and the singular value distribution of the diagonal matrix $\tilde{\mathbf{H}}_p$ is identical to the singular value distribution of

$$\sqrt{\sum_{q=1}^{N_R} \mathbf{H}_{p,q}^H \mathbf{H}_{p,q}} \quad (1.16)$$

The resource pooling result² was proven for $N_T = 1$ under a few additional technical assumptions (the spreading sequences shall be Gaussian distributed and the fading weights are uniformly bounded from above) by Hanly and Tse [1]. A generalization to arbitrary number of transmit antennas N_T and correlated fading is given in [14] and Section 1.10.

The resource pooling result states that, when going from a single antenna CDMA system with KN_T users, spreading factor NN_R and a fading distribution with N_R^{th} order diversity to a multiple antenna CDMA system with K users, spreading factor N , N_T transmit antennas per user, N_R receive antennas at the base station and a fading distribution without diversity, most performance measures stay the same. This means that *receive antennas are interchangeable with the spreading factor*—if we double the number of receive antennas, we can halve the spreading factor and, therefore, save half the spectral bandwidth. It also means that *transmit antennas are interchangeable with users*. Note that unlike to single user antenna array systems, there is no need to have the number of transmit antennas scale with the number of receive antennas to have the throughput scale with the number of receive antennas. Adding users—which means adding antennas, since every user has at least one antennas—does equally well in terms of total, but not per-user throughput.

1.5 TWO MISCONCEPTIONS

In order to exploit the potential of the resource pooling concept, the two conceptually identical tasks antenna combining and multi-user detection must be performed jointly. Sometimes a joint approach is questioned for sake of

²Actually, Hanly and Tse even showed a stronger result, i.e. the convergence of the signal-to-interference and noise ratio of the linear minimum mean-squared error multiuser detector.

saving complexity. It is the central aim of this contribution to challenge the idea that a separate approach can be given justification. In Sections 1.6 to 1.8, we will show how joint antenna combining and multi-user detection can be implemented with linear complexity per bit. In this section, we aim to give an intuitive understanding of why a separate approach performs badly. For the latter purpose, we restrict ourselves to two receive antennas and translate the abstract problem of joint antenna combining and multi-user detection into the more familiar problem of equalization for analog double-sideband amplitude modulation.

Consider an analog double-sideband amplitude modulation in the equivalent complex baseband representation. Let the frequency-selective channel be as shown in Fig. 1.1. The two sidebands correspond to the two signals received

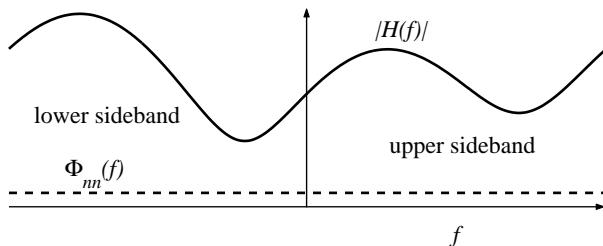


Fig. 1.1 Example of transfer function $H(f)$ and noise power density $\Phi_{nn}(f)$ in double sideband amplitude modulation.

at the two elements of the receiving antenna array. The magnitudes of the transfer functions in each sideband correspond to the singular value spectra of the first and the second block row of the virtual spreading matrix $\mathbf{Q}[m]$, these are $[\mathbf{S}_1[m]\mathbf{H}_{1,1} \cdots \mathbf{S}_{N_T}[m]\mathbf{H}_{N_T,1}]$ and $[\mathbf{S}_1[m]\mathbf{H}_{1,2} \cdots \mathbf{S}_{N_T}[m]\mathbf{H}_{N_T,2}]$. The phases of the transfer functions in each sideband correspond to the singular vectors of these two block rows.

Separating antenna combining and multiuser detection is equivalent to either ignore the phase relations when combining the two sidebands or to ignore the existence of the other sideband when equalizing the channel. Fig. 1.2 compares the resulting signals after combining the sidebands when the phase relations are taken into account vs. them being not taken into account. It is not surprisingly that after equalization, a price in terms of a low signal-to-noise ratio within a certain frequency range has to be paid for the ignorance of the phase relations, cf. Fig. 1.3. This figure also shows what happens when first an equalizer, which ignores the co-existence of the sidebands, is applied and then the two equalized (constant phase) signals are combined. This setting would correspond in multi-antenna CDMA to the case when both receiver chains employ separate multi-user detectors and then combine the signals of the two separate antenna chains at a later stage—this approach is common in

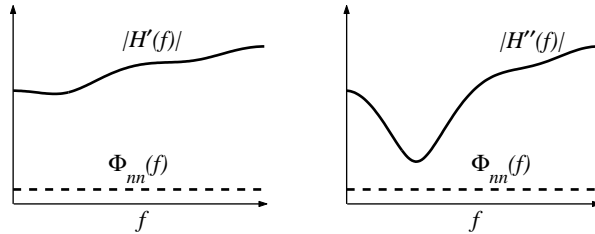


Fig. 1.2 Transfer functions after combining the two sidebands with an optimized frequency-dependent weighting (left hand side) vs. an optimized frequency-flat weighting (right hand side).

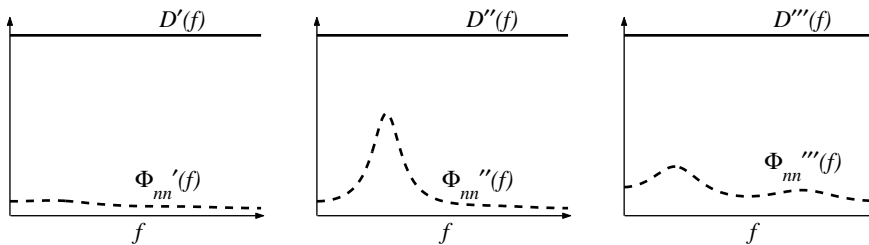


Fig. 1.3 Noise spectral density for the two cases of Fig. 1.2 after zero-forcing equalization (left and middle) vs. an equalization first, combining later approach (right).

macrodiversity scenarios where the antenna elements are not co-located. The result is also not encouraging.

It seems questionable whether there remains a noticeable gain at all due to the deployment of multiple-antennas, i.e. two sidebands, if signals are not processed jointly, since, in some cases, the mismatch of the detection procedure could outweigh the benefits of multiple reception. Fig. 1.4 addresses this point. There the previous approaches to re-construct the double sideband

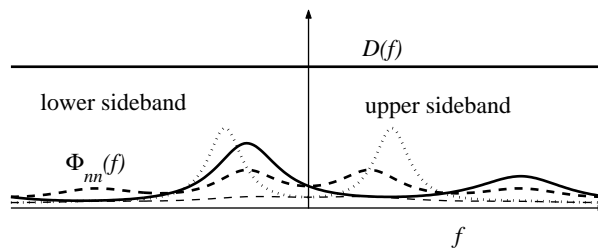


Fig. 1.4 Noise spectral densities after re-construction of the double-sideband signals from its single-sidebands in Fig. 1.3 (dashed and dotted lines) vs. direct equalization in the double-sideband (solid line).

from the overlay of single sidebands are compared to straightforward equalization without any combining. In this case, it indeed appears that, when first combining the two sidebands separately ignoring the phase relation, the overall worst-case performance becomes inferior to no combining at all.

1.6 LINEAR MULTI-STAGE RECEIVERS

The sufficient discrete-time statistics (1.6) is fed into a receiver algorithm which aims to recover the transmitted data symbols in the presence of noise and interference. This algorithm is a detector in case of uncoded transmission providing decisions on each individual symbol. If channel coding is used, the receiver algorithm has to perform two tasks: exploit the structure of the interference (detection) and exploit the code laws (decoding). Optimally, the two tasks are performed by a jointly exhaustive search. For complexity reasons, they are often performed separately. Multi-user joint decoders are feasible only if they are implemented by successive iterations over detection and decoding.

For sake of simplicity, consider here a separated approach where the detector output is fed into a decoding unit. In order to not suffer from an unnecessary hit in performance, the decoder should be fed with soft decisions—these are (in general non-linear) estimates calculated from the sufficient statistics in such way that they minimize the mean-squared error to the true symbols.

The optimum non-linear function is, in general and also in practice, a sum of at least 2^{KN_T} terms. In practice, the exponential complexity is overcome when the estimate is constrained to be a linear functional of the sufficient statistics. The best linear functional, in terms of achievable mean-squared error, for the channel in (1.9) is given by [2]

$$\mathbf{d}[m] = (\mathbf{Q}[m]^H \mathbf{Q}[m] + \sigma^2 \mathbf{I})^{-1} \mathbf{x}[m]. \quad (1.17)$$

The need for one matrix inversion per discrete time instant makes sure that this detector is not the method of choice in systems with many users.

In systems with many users, we can make use of the resource pooling results and the convergence of the singular values of the virtual spreading matrix $\mathbf{Q}[m]$ to some deterministic limit which is, as (1.14) shows, independent from the time index m . In order to illustrate this approach, consider the general problem of inverting a non-singular matrix \mathbf{R} with eigenvalues $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{K'}$. Note that any matrix annihilates its own characteristic polynomial (Cayleigh-Hamilton Theorem)

$$\prod_{k=1}^{K'} (\mathbf{R} - \lambda_k \mathbf{I}) = \mathbf{0}. \quad (1.18)$$

Expanding the product and solving for the identity matrix gives

$$\mathbf{I} = \sum_{k=1}^{K'} \alpha_k \mathbf{R}^k \quad (1.19)$$

with some coefficients α_k which depend only on the eigenvalues λ_1 to $\lambda_{K'}$. Multiplying both sides of (1.19) by the inverse of \mathbf{R}

$$\mathbf{R}^{-1} = \sum_{k=0}^{K'-1} \alpha_{k+1} \mathbf{R}^k \quad (1.20)$$

shows how the inverse of a $K' \times K'$ matrix can be written as a $(K' - 1)^{\text{st}}$ order matrix polynomial if its eigenvalues are known.

The eigenvalues of the matrix to be inverted for the linear minimum mean-squared error (LMMSE) detector (1.17) are given by the squared singular values of the virtual spreading matrix shifted by the noise variance. Due to the resource pooling result, they are time-invariant, if the number of users is large, and so are the coefficients of the matrix polynomial in (1.20).

In practice, one would neither like to calculate the matrix polynomial up to the full order nor need to do so. In fact, the number of terms to achieve satisfactory performance for applications in LMMSE detectors is quite small (less than ten) and does not scale with the number of users [3]. Therefore, we will choose the order of the polynomial to be $D - 1$ with $D \ll K'$

$$\mathbf{d}[m] \approx \sum_{k=0}^{D-1} \tilde{w}_k (\mathbf{Q}[m]^H \mathbf{Q}[m] + \sigma^2 \mathbf{I})^k \mathbf{x}[m] \quad (1.21)$$

$$= \sum_{k=0}^{D-1} w_k (\mathbf{Q}[m]^H \mathbf{Q}[m])^k \mathbf{x}[m]. \quad (1.22)$$

This procedure to approximate the LMMSE detector was first proposed by Moshavi et al. [4].

It is not a trivial problem to choose the right coefficients w_k for an approximating polynomial with reduced order. In particular, it is not a wise choice to just use the first D coefficients of the full order polynomial. Moshavi et al. [4] proposed to choose the coefficients in order to minimize the mean-squared error between the output of the exact LMMSE detector and its approximation. This leads to the following system of Yule-Walker equations for the choice of the coefficients

$$\begin{bmatrix} m_1 \\ \vdots \\ m_D \end{bmatrix} = \begin{bmatrix} m_2 + \sigma^2 m_1 & \cdots & m_{D+1} + \sigma^2 m_D \\ \vdots & \ddots & \vdots \\ m_{D+1} + \sigma^2 m_D & \cdots & m_{2D} + \sigma^2 m_{D-1} \end{bmatrix} \begin{bmatrix} w_0 \\ \vdots \\ w_{D-1} \end{bmatrix} \quad (1.23)$$

with the empirical eigenvalue moments

$$m_n = \frac{1}{K'} \sum_{k=1}^{K'} \lambda_k^n. \quad (1.24)$$

In complete analogy to the full order polynomial, the coefficients depend only on the singular values of the virtual spreading matrix denoted by $\sqrt{\lambda_k}$, $1 \leq k \leq KN_T$.

The resource pooling result states that for many users the singular value distribution (and thus all its moments) of the virtual spreading matrix for multiple antenna systems is identical to the corresponding singular value distribution in a single antenna system with an appropriately modified fading distribution. Thus, we can apply the same approach to find the empirical eigenvalue moments in CDMA systems with multiple antennas as in CDMA systems with single element antennas. This approach is presented in Section 1.8. For considerations concerning asynchronous systems, the reader is referred to [5].

1.7 BLOCK STRUCTURE OF RECEIVER

Approximation (1.22) leads to the very simple implementation structure of the linear multi-stage receiver shown in Figure 1.5. The receiver consists of $D - 1$

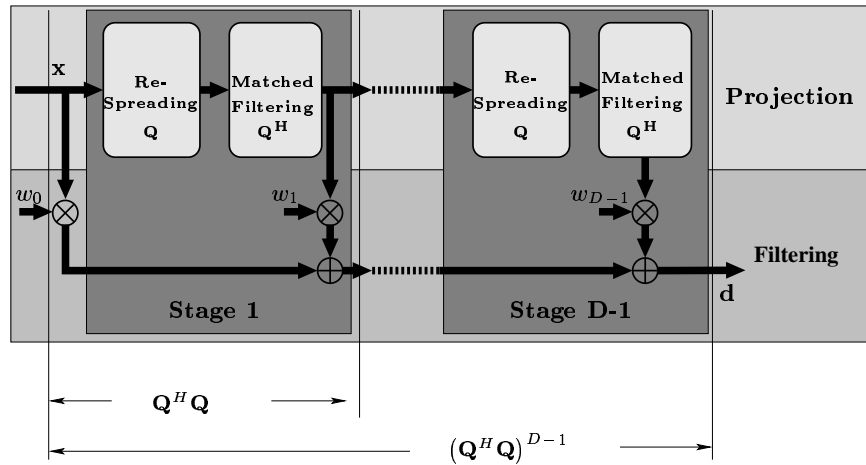


Fig. 1.5 Linear multistage receiver structure.

identical stages each of whose performs a re-spreading of the input signals (filtering by the "virtual" spreading matrix $Q[m]$) and subsequent matched filtering (filtering by $Q[m]^H$).

Figure 1.5 suggests an alternative interpretation of the linear multi-stage receiver as subspace technique, an approach followed in [3]. Each stage provides the projection of the received signal onto a one-dimensional subspace. The subsequent weighting corresponds to a bank of filters working in a D -dimensional space. In contrast to other subspace methods, this technique has the following properties:

- It does not require tracking of the signal subspace.
- It shows the surprising property that the rank D required to achieve a target SINR (e.g. within an small ϵ of the full rank LMMSE receiver) does not scale with the system size, i.e. with KN_T and NN_R . The signal space becomes larger but the dimension of the subspace needed to achieve certain performance level saturates.
- A few stages are sufficient for near full-rank performance. The output SINR of the reduced rank receiver converges *exponentially fast* in the rank D towards output SINR of the full rank receiver.
- The complexity per transmitted bit of the projection module in Figure 1.5 scales in the same way with KN_T as the complexity of the single-user matched filter. However, the computation of the empirical eigenvalue moments has cubic complexity in KN_T . This drawback can be overcome using the asymptotic weighting discussed in Section 1.8.

1.8 LARGE-SYSTEM WEIGHTING

The empirical moments m_d needed in (1.23) to calculate the weights depend on the spreading matrices $\mathbf{S}_p[m]$, $p = 1, \dots, N_T$ and the channel realizations. Thus, the weights have to be re-computed if one of these parameters change. In mobile communications the channel may change rapidly and the spreading matrices may vary from symbol to symbol (e.g. in the frequency-division duplex mode in UMTS). The weights have to be updated at each symbol interval or, at most, at each frame interval. When the size of the virtual spreading matrix $\mathbf{Q}[m]$ is large, it is computationally too demanding to determine the detector weights according to (1.23). However, it is possible to exploit a property of random matrices which allows for a significant simplification of the weight-computation. With some conditions, usually fulfilled in practice, for the virtual spreading matrix $\mathbf{Q}[m]$, all empirical moments m_d converge to deterministic limits as the system size grows large

$$\lim_{K \rightarrow \infty} m_d = \mu_d. \quad (1.25)$$

The limit μ_d depends on a small set of system parameters and can be computed without much effort. Large-system weights approximate the exact

weights in (1.23) by substituting the empirical moments m_d with its asymptotic limits μ_d . When the system size is sufficiently large, $m_d \approx \mu_d$ and the performance degradation due to this approximation is negligible.

Let us consider a simple example. Let all the entries of $\mathbf{Q}[m]$ be statistically independent and identically distributed (i.i.d.) random variables with zero mean and unitary variance. Consider matrices $\mathbf{Q}[m]$ with different sizes, but identical aspect ratios³. For each of them generate 100 independent realizations. In Figure 1.6, the moments m_1 and m_2 are shown for all realizations.

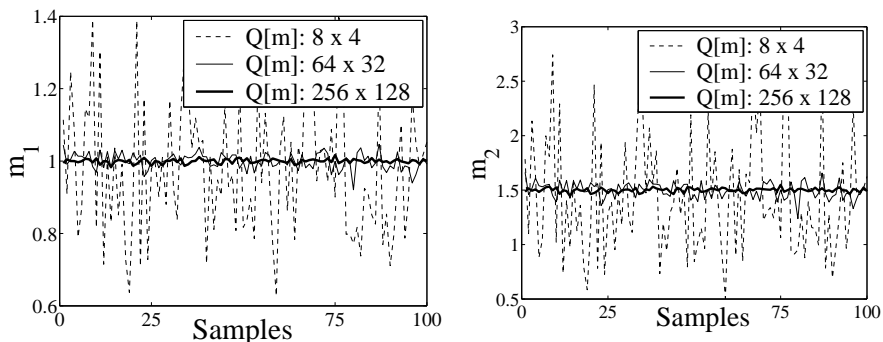


Fig. 1.6 First two moments for 100 realizations of $\mathbf{Q}[m]$ with i.i.d. Gaussian entries.

It is apparent that m_1 and m_2 converge to 1 and $\frac{3}{2}$, respectively, as the size of $\mathbf{Q}[m]$ increases. The asymptotic moments μ_d depend only on the aspect ratio β and are given by

$$\mu_d = \sum_{i=0}^{d-1} \binom{d}{i} \binom{d}{i+1} \frac{\beta^i}{d}. \quad (1.26)$$

Let us consider now the matrix $\mathbf{Q}[m]$ with the structure defined in (1.10). Moreover, assume the following:

- The elements $s_{k,p,m}$ are i.i.d. with zero mean and variance $\frac{1}{N}$.
- The channel fading coefficients $h_{k,p,q}$ are independent across k , p and q . Their distribution is uniform in phase over $[0, 2\pi)$ and arbitrary, but upper bounded⁴ in amplitude. For a given transmitting antenna p and given user k , the fading coefficients $h_{k,p,1}, h_{k,p,2}, \dots, h_{k,p,N_R}$ are identically distributed. The assumption of independence is verified in scenarios with rich scattering and sufficient separation between antennas.

³This is the ratio of the number of columns to the number of rows of a matrix

⁴The widely used Rayleigh distribution is not upper bounded, but by physical reasons any fading coefficient is upper bounded.

The identical distribution of the fading coefficients for a given transmitting antenna element models micro-diversity scenarios, in which all the receive antenna elements are placed at the same base station.

Now, we define the powers

$$P_{k,p} = \sum_{q=1}^{N_R} |h_{k,p,q}|^2 \quad (1.27)$$

and denote with η_d the d^{th} moment of their cumulative distribution $F_P(\pi)$, i.e. $\eta_d = \int \pi^d d F_P(\pi)$. The positive asymptotic moments μ_d can be determined from the moments of the powers η_d by the following recursive algorithm:

Initialization: Let $\beta = \alpha \frac{N_T}{N_R}$, $\rho_0(x) = 1$ and $\mu_0 = \beta^{-1}$.

Recursion: 1. Define

$$\rho_{d+1}(x) = \beta x \sum_{s=0}^d \rho_s \mu_{d-s}$$

and write it as polynomial in x .

2. Replace all the monomials x, x^2, \dots, x^{d+1} in the polynomial $\rho_{d+1}(x)$ by $\eta_1, \eta_2, \dots, \eta_{d+1}$, respectively, and assign the result to μ_{d+1} .

The asymptotic eigenvalue moments for $\mathbf{Q}[m]$ depend only on the aspect ratio and on the eigenvalue moments of the distribution $F_P(\pi)$. They are easily computed. The first six asymptotic eigenvalue moments are

$$\begin{aligned} \mu_1 &= \eta_1 \\ \mu_2 &= \beta \eta_1^2 + \eta_2 \\ \mu_3 &= \beta^2 \eta_1^3 + 3\beta \eta_1 \eta_2 + \eta_3 \\ \mu_4 &= \beta^3 \eta_1^4 + 6\beta^2 \eta_1^2 \eta_2 + 4\beta \eta_1 \eta_3 + 2\beta \eta_2^2 + \eta_4 \\ \mu_5 &= \beta^4 \eta_1^5 + 10\beta^3 \eta_1^3 \eta_2 + 10\beta^2 \eta_1^2 \eta_3 + 10\beta^2 \eta_1 \eta_2^2 + 5\beta \eta_1 \eta_4 + 5\beta \eta_2 \eta_3 + \eta_5 \\ \mu_6 &= 20\beta^3 \eta_3 \eta_1^3 + 3\beta \eta_3^2 + 6\beta \eta_4 \eta_2 + 5\beta^2 \eta_3^2 + 15\beta^4 \eta_1^4 \eta_2 + 6\beta \eta_5 \eta_1 \\ &\quad + 15\beta^2 \eta_4 \eta_1^2 + 30\beta^3 \eta_1^2 \eta_2^2 + 30\beta^2 \eta_1 \eta_2 \eta_3 + \beta^5 \eta_1^6 + \eta_6 \end{aligned} \quad (1.28)$$

1.9 PERFORMANCE

In this section, we study the performance of linear multistage receivers with asymptotic weighting proposed in Section 1.8 and compare it to the exact

weighting in (1.23), the large system approximation of the multistage Wiener receiver in \mathbf{Q} , described in Section 1.10 and the full rank LMMSE receiver.

The simulation results presented in this section were obtained for uncoded transmission in flat Rayleigh fading, using $\frac{\pi}{4}$ -offset QPSK modulation, and assuming perfect knowledge of the channel. The receivers are compared in terms of their bit error rate (BER) evaluated as a function of the normalized signal-to-noise ratio E_b/N_0 where E_b is mean energy per bit and N_0 is the one sided noise spectral density. Figure 1.7 shows the BER versus E_b/N_0 for a 5-stage detector and $\beta = 0.5$ ($K = 64, N = 64, N_T = 62, N_R = 4$). The performance degradation due to the large system approximation of weights is completely negligible. Figure 1.8 shows the performance improvements of the detector with large-system weighting for increasing number of stages.

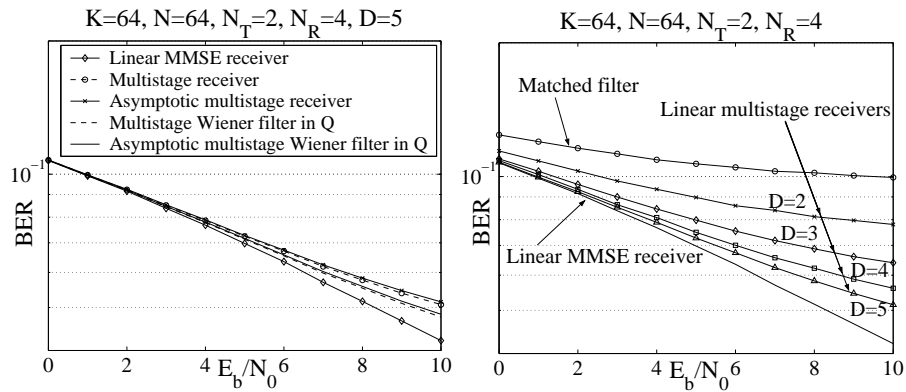


Fig. 1.7 BER versus $\frac{E_b}{N_0}$ for $\beta = \frac{1}{2}$.

Fig. 1.8 BER versus $\frac{E_b}{N_0}$ for $\beta = \frac{1}{2}$ for $D = 2, \dots, 5$.

1.10 RELATED WORKS

The linear multistage receivers in (1.22) with weighting (1.23) belong to a wider class of receivers described by

$$\tilde{\mathbf{d}}[m] = \sum_{k=0}^{D-1} \mathbf{W}_k (\mathbf{Q}[m]^H \mathbf{Q}[m])^k \mathbf{x}[m]. \quad (1.29)$$

with $\mathbf{W}_k = \text{diag}(w_{k,1}, w_{k,2}, \dots, w_{k,K'})$.

The weighted linear parallel interfering cancelling (W-PIC) receivers, generalized to multiple antenna systems provide a simple example of receivers in the class described by (1.29).

A non-adaptive implementation of the multistage Wiener filters [3] is obtained from (1.29) enforcing \mathbf{W}_d , $d = 0, \dots, D-1$, to minimize the mean-squared error $E\{\|\tilde{\mathbf{d}}[m] - \mathbf{d}[m]\|^2\}$. Throughout, we refer to it as the *multistage Wiener filter in \mathbf{Q}* . This receiver was proposed and its asymptotic performance for CDMA systems in flat fading channels analyzed in [6]. Let $\mathbf{Q}_{k'}[m]$ be the matrix obtained from $\mathbf{Q}[m]$ by suppressing the virtual spreading corresponding to the signal k' . Under the conjecture⁵ that, at least asymptotically, $E\{\prod_k \mathbf{x}_{k'}^H (\mathbf{Q}_{k'}^H \mathbf{Q}_{k'})^{j_k} \mathbf{x}_{k'}\} = \prod_k E\{\mathbf{x}_{k'}^H (\mathbf{Q}_{k'}^H \mathbf{Q}_{k'})^{j_k} \mathbf{x}_{k'}\}$, with $j_k \in \mathbb{N}$ the large-system approximation of the weights for the multistage Wiener filter for the MIMO channel model (1.9) is a straightforward extension of the results in [1] and [6]. The weights for estimating the signal transmitted by antenna p of user k can be obtained from (1.23) by substituting m_d , $d = 1, \dots, 2D$, with $\rho_d(P_{k,p})$. The multistage Wiener filter performs marginally better than the receiver discussed in Section 1.8, when applied in systems with received power imbalances among users [6] (see Figure 1.7). The two receivers coincide for equal received powers at the receiver front-end [6]. However, the multistage Wiener filter in \mathbf{Q} requires the inversion of the equation system (1.23) for each transmitting antenna of each user. This drawback makes it less attractive for implementation. Often, an alternative linear multistage receiver is considered in literature: the *multistage Wiener filter in $\mathbf{Q}_{k'}$* . For each transmitted signal k' , it is given by:

$$\tilde{\mathbf{d}}[m] = \sum_{k=0}^{D-1} w_k (\mathbf{Q}_{k'}[m]^H \mathbf{Q}_{k'}[m])^k \mathbf{x}_{k'}[m], \quad (1.30)$$

where $\mathbf{x}_{k'}[m]$ is the matched filter output corresponding to the signal k' . The weights satisfy

$$\begin{bmatrix} \tau_1 \\ \vdots \\ \tau_D \end{bmatrix} = \begin{bmatrix} \tau_2 + \sigma^2 \tau_1 + \tau_1^2 & \cdots & \tau_{D+1} + \sigma^2 \tau_D + \tau_1 \tau_D \\ \vdots & \ddots & \vdots \\ \tau_{D+1} + \sigma^2 \tau_D + \tau_D \tau_1 & \cdots & \tau_{2D} + \sigma^2 \tau_{2D-1} + \tau_D^2 \end{bmatrix} \begin{bmatrix} w_0 \\ \vdots \\ w_{D-1} \end{bmatrix}$$

with $\tau_1 = P_{k,p}$ and $\tau_d = P_{k,p} \beta \mu_{d-1}$. This detector has been erroneously considered completely equivalent to the multistage Wiener filter in \mathbf{Q} and widely studied since it is easier to analyze. The two receivers are equivalent from the performance point of view only. However, their implementations differ considerably. The subspace basis of the multistage Wiener filter in \mathbf{Q} can be computed jointly for all the transmitted signals, as shown in Figure 1.5, and it has linear complexity order per signal. The subspace basis of the multistage Wiener filter in $\mathbf{Q}_{k'}$ has to be computed again for each transmitted

⁵This conjecture is supported by simulations (see Figure 1.7).

signal and the filter has a quadratic complexity order per signal, as the linear MMSE receiver.

The convergence rate of the multistage receiver output SINR toward to the full rank LMMSE receiver output SINR is studied in [7]. A unified framework for the asymptotic performance analysis of multistage detectors (1.29) with any kind of weighting in CDMA systems is provided in [8]. The weighting approximation for large systems was proposed in [9] for communication over the AWGN channel. Afterwards, much effort has been devoted to extend these results to more realistic scenarios. We can distinguish two large groups: one assumes the spreading matrix to be a completely random matrix without additional constraints in order to model systems with pseudo-noise scrambling sequences. The second group focuses on unitary random matrices to model systems with orthogonal spreading sequences. The extension of large-system weighting to CDMA systems in uplink with flat fading in [8] belongs to the first group. An approximation of the widely linear MMSE receiver for BPSK modulation in flat fading channels is proposed in [10]. CDMA systems in uplink with multipath fading channels are studied in [11] while the multistage Wiener filter in \mathbf{Q}_k for the downlink is analyzed in [12]. MMSE analysis of systems with large unitary random matrices has been introduced first in [13]. This result allows the extension of asymptotic analysis based on random matrices to OFDM and multi-carrier CDMA systems. Large systems weighting for CDMA systems in down-link with multipath fading channels and unitary random spreading matrices has been proposed in [12]. In [1] it is conjectured the extension of the resource pooling result for scenarios with microdiversity to scenarios with macrodiversity, in which the receive antennas are not co-located. The resource pooling result by Hanly and Tse [1] has been generalized to CDMA systems with correlated antennas in [14]. The general result includes, as special cases, the microdiversity and macrodiversity scenarios in [1] and provides a rigorous proof for macrodiversity. In addition, it allows for arbitrary distributions of the fading coefficients including line-of-sight scenarios.

Generalized resource pooling: Let $\mathbf{h}_{k,p} = [h_{k,p,1}, \dots, h_{k,p,N_R}]^T$ be the known vector of channel coefficients between the p -th antenna of user k and the receive antennas. Let $\mathbf{h}_{k,p}$ be a realization of an N_R -dimensional random column vector \mathbf{h} such that its distribution converges almost surely, as $K \rightarrow \infty$ to a limit distribution function $F_{\mathbf{h}}(\mathbf{h})$ with upper bounded support. Then, as $N, K \rightarrow \infty$ with $\frac{K}{N} \rightarrow \alpha$ and N_R and N_T fixed, the SINR at the receiver output for the p^{th} antenna signal of user k conditioned on $\mathbf{h}_{k,p}$ converges almost surely to the deterministic constant

$$\lim_{N, K \rightarrow \infty} \text{SINR}_{k,p} \stackrel{\text{a.s.}}{=} \mathbf{h}_{k,p}^H \mathbf{C} \mathbf{h}_{k,p} \quad (1.31)$$

where \mathbf{C} is the unique positive definite $N_R \times N_R$ matrix solution to the fixed point equation

$$\mathbf{C}^{-1} = \sigma^2 \mathbf{I} + \alpha N_T \int \frac{\mathbf{h}\mathbf{h}^H}{1 + \mathbf{h}^H \mathbf{C} \mathbf{h}} dF_{\mathbf{h}}(\mathbf{h}). \quad (1.32)$$

The linear multistage approach is extended to estimation of multipath fading channels for synchronous CDMA [15]. The effects of imperfect channel knowledge on the performance of linear multistage receivers is analyzed in [15].

In the previous works only synchronous systems have been considered. The impact of asynchronicity on multistage receivers for CDMA systems in flat fading channels has been studied in [17] and [5]. A slightly modified version of the receiver structure in Figure 1.5 has been proposed. In contrast to the linear MMSE receiver which suffers from performance degradation due to truncation effects in asynchronous systems, the proposed implementation of multistage detectors is not affected from truncation effects [17]. Chip asynchronicity is taken into account in [5] and an algorithm to determine the large system weighting is given.

Acknowledgments

The authors thank Stephen Hanly for clarifying some aspects of resource pooling.

REFERENCES

1. S. V. Hanly and D. N. C. Tse. Resource pooling and effective bandwidth in CDMA networks with multiuser receivers and spatial diversity. *IEEE Trans. Inform. Theory*, 47(4):1328–1351, May 2001.
2. S. Verdú. *Multiuser Detection*. Cambridge University Press, New York, 1998.
3. M. Honig and W. Xiao. Performance of reduced-rank linear interference suppression. *IEEE Trans. Inform. Theory*, 47(5):1928–1946, July 2001.
4. S. Moshavi, E. G. Kanterakis, and D. L. Schilling. Multistage linear receivers for DS-CDMA systems. *Int'l. J. Wireless Inform. Networks*, 3(1):1–17, Jan. 1996.
5. L. Cottatellucci and R. R. Müller. Multistage detectors for asynchronous CDMA. In *Proc. of Int'l. Zurich Seminar on Commun.*, Zurich, Switzerland, Feb. 2004.

6. L. Cottatellucci and R. R. Müller. Multiuser interference mitigation with multistage detectors: Design and analysis for unequal powers. In *Proc. of Asilomar Conf. Signals, Syst., and Comp.*, Pacific Grove, CA, U.S.A., Nov. 2002.
7. P. Loubaton and W. Hachem. Asymptotic analysis of reduced rank Wiener filters. In *Proc. of IEEE Inform. Theory Workshop*, Paris, France, Apr. 2003.
8. L. Cottatellucci and R. R. Müller. Asymptotic design and analysis of multistage detectors with unequal powers. In *Proc. of Inform. Theory Workshop*, Bangalore, India, Oct. 2002.
9. R. R. Müller and S. Verdú. Design and analysis of low-complexity interference mitigation on vector channels. *IEEE J. Sel. Areas in Commun.*, 19(8):1429–1441, Aug. 2001.
10. A. M. Tulino and S. Verdú. Asymptotic analysis of improved linear receivers for BPSK-CDMA subject to fading. *IEEE J. Sel. Areas in Commun.*, 19(8):1544–1555, Aug. 2001.
11. L. Cottatellucci and R. R. Müller. A systematic approach to multistage detectors in multipath fading channels. Submitted to *IEEE Trans. Inform. Theory*, Feb. 2003.
12. W. Hachem. Simple polynomial detectors for CDMA downlink transmissions on frequency-selective channels. *IEEE Trans. Inform. Theory*, 50(1):164–172, Jan. 2004.
13. M. Debbah, W. Hachem, P. Loubaton, and M. de Courville. MMSE analysis of certain large isometric random precoded systems. *IEEE Trans. Inform. Theory*, 49(5), May 2003.
14. L. Cottatellucci and R. R. Müller. A generalized resource pooling result for correlated antennas or asynchronous CDMA systems. Submitted to *IEEE Int'l. Symp. Inform. Theory & Appl.*, Parma, Italy, Oct. 2004.
15. L. Cottatellucci and R. R. Müller. Asymptotic analysis of multistage receivers for multipath fading channels. In *Proc. of IEEE Int'l. Symp. Inform. Theory*, Yokohama, Japan, June/July 2003.
16. J. Evans and D. N. C. Tse. Large system performance of linear multiuser receivers in multipath fading channels. *IEEE Trans. Inform. Theory*, 46(6):2059–2078, Sep. 2000.
17. L. Cottatellucci, M. Debbah, and R. R. Müller. Asymptotic analysis of linear detectors for asynchronous CDMA systems. In *Proc. of IEEE Int'l. Symp. Inform. Theory*, Chicago, IL, U.S.A., June/July 2004.