Linear interference cancellation in CDMA systems and large deviations of the correlation matrix eigenvalues

Citation for published version (APA):

Document status and date:
Published: 01/01/2003

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.

Download date: 27. May, 2021
Linear Interference Cancellation in CDMA Systems and Large Deviations of the Correlation Matrix Eigenvalues

Anne Fey-den Boer * Remco van der Hofstad* Marten J. Klok†
Department of Mathematics and Computing Science,
Eindhoven University of Technology
P.O.box 513, 5600 MB Eindhoven, the Netherlands

October 14, 2003

Abstract We analytically compute a measure of performance of various linear Parallel Interference Cancellation (PIC) decoding schemes in the infinite stage limit, for moderately loaded CDMA systems without AWGN, or with a sufficiently small amount of AWGN. This measure is the exponential rate of the BEP, which does not involve Gaussian approximations. We obtain these rates using large deviation theory for the eigenvalues of the code correlation matrix. We find that the decorrelator performs best, followed by infinite-stage SD-PIC, which is found to perform better than infinite stage HD-PIC.

1 Introduction

In Code Division Multiple Access (CDMA) communication systems, each user multiplies his data signal by an individual coding sequence. The base station can then distinguish the different messages by taking the inner product of the total signal with each coding sequence. This is called Matched Filter (MF) decoding. An important application is mobile telephony. Since, due to synchronisation problems, it is unfeasable to implement completely orthogonal codes for mobile users, the decoded messages will suffer from Multiple Access Interference (MAI). In practice, pseudorandom codes are used.

Several techniques have been proposed to improve the MF decoding. The Maximum Likelihood detector [1] offers the best quality, but it is too complex to implement real-time. In contrast, we consider versions of linear Parallel Interference Cancellation (PIC) [2, 3, 4, 5]. This is a fast and flexible iterative method which in each stage estimates and subtracts the MAI for all users. The simplest version is called Soft Decision (SD) PIC, while slightly altering the iteration procedure leads to multistage approaches of the decorrelator or the minimum mean squared error (MMSE) detector. To compare the quality of the various decoding techniques, the Bit Error Probability (BEP) is best suited. However, the BEP is not easily calculated. It is therefore important to find useful approximations for the BEP. Mathematically, this is a difficult problem.

Linear PIC procedures have the advantage that they can be expressed in matrix notation. The problem of finding the BEP is then, in the multistage limit, translated into finding the distribution of the eigenvalues of a random matrix, in this case the code correlation matrix. Mathematically, a system with a fixed number of users \( k \) while the code length \( n \) approaches infinity, is most convenient. For this lightly loaded system, we use Large Deviation Theory (LDT) to derive an asymptotic expression for the BEP. LDT provides the decay rate of exponentially small probabilities; we will highlight this theory below.

From a practical point of view, we of course wish to serve many users with a code length as small as possible. At this other end of the spectrum, where \( k \) is proportional to \( n \), a Gaussian approximation of the BEP seems the suitable approach [6, 7]. However, even for this heavily loaded system it is not clear that this approximation is valid. Since the PIC procedure involves addition and subtraction, the resulting variables are no longer independent. We know of similar examples ([8], Section II.F and [9], Section 4.2) where the Gaussian approximation leads to false conclusions.

Simulation is a powerful method for obtaining more information about the BEP. However, in the cases that we have investigated, the BEP is expo-
nentially small, so that a direct Monte Carlo simulation would require enormous amounts of trials. A well-known technique to reduce the simulation time is Importance Sampling (IS). A review on IS was published in [10]. The idea is that one transforms the distribution function from which the random variables are sampled, so that successes (in this case, bit errors) become more likely to occur. Appropriately weighting the output, one obtains an unbiased estimator. For simulating events with exponentially small probabilities, the large deviation rate is a necessary ingredient to obtain the right transformed distribution function.

Most of our large deviation results extend to a moderately loaded system, in our case with $k$ of the order of $n/\log n$. In this regime, multistage linear PIC performs much better than MF.

## 2 System model

The system has $k$ users, who for simplicity are supposed to transmit binary signals. Considering only the first bit, we summarize the data together with the power of each user in a single vector $\mathbf{b}$, where $b_m$ is either valued $+\sqrt{P_m}$ or $-\sqrt{P_m}$, for each $m = 1, \ldots, k$. We can then use a matrix notation for the PIC procedure as follows. We model the $k$ codes as vectors $\mathbf{c}_m$ of length $n$, consisting of random independent bits with distribution $\mathcal{B}(0,1)$, and $\mathcal{B}(1,1)$, respectively. We can choose $\mathbf{c}_m$ as random variables, distributed as $N(0,\sigma^2)$ for all $i = 1, \ldots, n$, where $\sigma$ is a parameter denoting the noise strength. In practice, the AWGN is dealt with by adjusting the powers, so that it is not dominant. This is reflected in the model by the choice of a fixed $\sigma$, even though $n$ may vary.

Decoding for user $m$ is done by taking the inner product with code $\mathbf{c}_m$, and dividing by $n$. This yields an estimate $\hat{b}_m$ for the sent bit $b_m$. In matrix notation, the vector $\mathbf{b}$ is estimated by $\hat{\mathbf{b}} = \frac{1}{n} \mathbf{C} \mathbf{s}$. The bit estimate vector $\hat{\mathbf{b}}$ is given by

$$\hat{\mathbf{b}} = \mathbf{W} \mathbf{b} + \frac{1}{n} \mathbf{C} \mathbf{N},$$

where the matrix $\mathbf{W} = \frac{1}{n} \mathbf{C} \mathbf{C}^T$ contains the inner products, i.e., the cross correlations of the codes.

When this is written as

$$\hat{\mathbf{b}} = \mathbf{b} + (W - I)\mathbf{b} + \frac{1}{n} \mathbf{C} \mathbf{N}, \quad (1)$$

it is clearly seen that the estimated bit vector is a sum of the correct bit vector, MAI, and an AWGN contribution. Decoding the message for addressee $m$ is now done by taking the sign of $\hat{b}_m$. In the rare case that $\hat{b}_m = 0$, the sign will be a random choice with equal probability for + and -. When the sign of $\hat{b}_m$ differs from that of $b_m$, a bit error is made for user $m$. This is the simplest decoding scheme, called Matched Filter, or MF.

PIC schemes deal with the second term, by estimating the MAI. In SD-PIC, this is simply done by subtracting the second term, with $\mathbf{b}$ replaced by $\hat{\mathbf{b}}$. This yields a new and usually better estimate for $\mathbf{b}$. To compare, the Hard Decision (HD) PIC procedure is similar, only each contribution $\hat{b}_m$ is estimated as $\pm \sqrt{P_m}$, thus making it a non-linear procedure.

In the case of multistage PIC, this new estimate is used in a second PIC iteration. We will now write the multistage SD-PIC procedure in matrix notation. We number the successive SD estimates for $\mathbf{b}$ with an index $s$, where $s = 1$ corresponds to (1), the MF decoding. In each new iteration, the latest guess for the MAI is subtracted from (1). The iteration in a recursive form is therefore:

$$\hat{\mathbf{b}}^{(s)} = \hat{\mathbf{b}}^{(1)} - (W - I)\hat{\mathbf{b}}^{(s-1)}. \quad (2)$$

This can be worked out to

$$\hat{\mathbf{b}}^{(s)} = \sum_{\varsigma=0}^{s-1} (I - W)^\varsigma (W \mathbf{b} + \frac{1}{n} \mathbf{C} \mathbf{N}). \quad (3)$$

The rest of the paper will investigate the case when $s \to \infty$; we call this infinite stage SD-PIC. The BEP in the case of infinite stage SD-PIC is strongly related to the eigenvalues of $W$. Note for instance that the series $\sum_{\varsigma=0}^{s-1} (I - W)^\varsigma$ converges to $W^{-1}$, as long as the eigenvalues are between 0 and 2.

To see this more clearly, we omit the noise for a moment and decompose $\mathbf{b}$ in eigenvectors of $W$. We let $\lambda_j$ be the $j^{th}$ eigenvalue of $W$, and $\mathbf{w}_j$ the corresponding eigenvector. We can choose $\mathbf{w}_j$ to be an orthogonal basis, since $W$ is a symmetric matrix, so that

$$\mathbf{b} = \sum_{j=1}^{k} \beta_j \mathbf{w}_j,$$

and

$$\hat{\mathbf{b}}^{(s)} = \sum_{\varsigma=0}^{s-1} \sum_{j=1}^{k} (1 - \lambda_j)^\varsigma \lambda_j \beta_j \mathbf{w}_j$$

for the sent bit $b_m$. This is the simplest decoding scheme, called Matched Filter, or MF.
\[ \sum_{j=1}^{k} (1 - (1 - \lambda_j)^s) \beta_j w_j. \quad (4) \]

From this expression it is clear that when all eigenvalues are in the interval (0, 2) and in the absence of noise, as \( s \to \infty \) the data is estimated as \( \sum_{j=1}^{k} \beta_j w_j = b \), which is the correct value. When one or more eigenvalues are zero and in absence of noise, a data bit could be estimated as 0, in which case the system will guess the sign. When an eigenvalue is larger than 2, the factor \((1 - (1 - \lambda_j)^s)\) becomes very large and alternating with \(s\), which will also lead to errors. This effect is related to the ping-pong effect studied in [11].

The authors use a similar expression as above to investigate the convergence rate and behaviour of the iteration as a function of the number of stages.

The interval for admissible eigenvalues can be widened by adding a weight \( \mu < 1 \) to \( W \) in the iteration (2). Sometimes the weight varies from stage to stage [3], but more commonly a fixed weight is chosen. This weight will appear in front of the \( W \)'s in 3, and in front of the \( \lambda_j \)'s in (4).

The interval of eigenvalues for which (4) converges is then widened to \( (0, 2/\mu) \). When the weight is chosen as \( \mu \leq 2/\lambda_{\text{max}} \), where \( \lambda_{\text{max}} \) is the largest eigenvalue, the interval becomes \( (0, \infty) \), i.e., \( W \) is invertible. The iteration then approaches the decorrelator system, in which \( W^{-1} \) is directly calculated:

\[ \hat{b}_{\text{dec}} = W^{-1} \hat{b} \]

When noise is included in the model, it is enhanced by the PIC procedure [2], so that the variance is no longer \( \sigma^2 \). Particularly, small eigenvalues of \( W \) lead to noise enhancement, so that this BEP source also depends on the eigenvalues.

To see this more clearly, we suppose for a moment that \( W^{-1} \) exists. We calculate the covariance matrix of \( \tilde{N} = W^{-1} \frac{1}{n} CN \) for fixed \( C \) as follows:

\[
\begin{align*}
\text{cov}(\tilde{N}) &= \text{cov}(W^{-1} \frac{1}{n} CN) \\
&= (W^{-1} \frac{1}{n} C \text{cov}(N)(W^{-1} \frac{1}{n} C)^T \\
&= W^{-1} \frac{1}{n} C \sigma^2 I \frac{1}{n} C^T W^{-1} - \frac{1}{n} \sigma^2 W^{-1}.
\end{align*}
\]

Since a bit error for user \( m \) is made when the noise exceeds \( \sqrt{P_m} \), for fixed \( C \) we have

\[ P(\tilde{N}_m < -\sqrt{P_m} | C) = Q\left( \sqrt{\frac{nP_m}{\sigma^2 W_{m,m}}} \right), \]

where \( W^{-1}_{m,m} \) is the \( m \)th diagonal element of \( W^{-1} \), so that for random \( C \)

\[ P(\tilde{N}_m < -\sqrt{P_m}) = E \left[ Q\left( \sqrt{\frac{nP_m}{\sigma^2 W_{m,m}}} \right) \right]. \quad (5) \]

Here, \( Q(x) \) is the error function for the standard normal distribution: \( Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}dt. \) It is difficult to calculate \( W_{m,m}^{-1} \) directly. We can however demonstrate that for \( k > 1 \) we have \( W_{m,m}^{-1} > 1 \), so that indeed the noise is always enhanced. For this we decompose \( W \) in its basis of eigenvectors. We introduce \( V \) as the matrix in which the normalized eigenvectors \( w_j \) of \( W \) are columns. Since \( W \) is symmetric, it has an orthonormal basis of eigenvectors, so that \( V^{-1} = V^T \). Then we have \( W = V \Lambda V^T \), where \( \Lambda \) is the diagonal matrix of eigenvalues. Since eigenvectors of \( W \) are eigenvectors of \( W^{-1} \), we also have \( W^{-1} = V \Lambda^{-1} V^T \). We can now write

\[ W_{m,m}^{-1} = \sum_{j=1}^{k} \frac{1}{\lambda_j} (w_j)_m^2. \]

Since \( \sum_{j=1}^{k} (w_j)_m^2 = \| e_m \|^2 = 1 \), we can use Jensen’s inequality:

\[ \sum_{j=1}^{k} \frac{1}{\lambda_j} (w_j)_m^2 > \frac{1}{\sum_{j=1}^{k} \lambda_j (w_j)_m^2} = 1/W_{m,m} = 1. \]

On the other hand, we have \( W_{m,m}^{-1} \leq 1/\lambda_{\text{min}} \), where \( \lambda_{\text{min}} \) is the smallest eigenvalue, so that indeed the smallest eigenvalue dominates the noise enhancement.

We finally note that another system, the MMSE detector, can also be iteratively approached with a PIC system [7]. The MMSE detector is designed to minimize errors due to MAI as well as due to noise. In the future, we intend to include this system in our study as well.

### 3 Large deviations

We know that as \( n \) increases with \( k \) fixed, the matrix \( W \) approaches the identity matrix almost surely. Therefore, the eigenvalues \( \lambda_j \) with \( j = 1, \ldots, k \), will all converge to 1. If \( k \to \infty \) as well, holding \( \delta = k/n < 1 \) constant, it is known that the smallest eigenvalue converges to \((\sqrt{\delta} - 1)^2 \) [12], and the largest eigenvalue to \((\sqrt{\delta} + 1)^2 \) [13].

Thus for vanishing \( \delta \), the deviations \( \lambda_{\text{min}} = 0 \) and \( \lambda_{\text{max}} \geq 2 \) are rare events, and can be treated with large deviation theory.

We will first explain the essence of this theory [14]. Suppose \( P_n \) is a sequence of probabilities and \( P_n \to 0 \) as \( n \to \infty \). A natural example is the probability that the mean of \( n \) independent variables will be some (fixed) value other than the
expectation. Typically, this decay is exponential in $n$, so that it makes sense to define the exponential rate as

$$ I = - \lim_{n \to \infty} \frac{1}{n} \log P_n. $$

For finite, but not too small $n$, we can read this as

$$ P_n = e^{-nI(1+o(1))}. $$

Thus, the probability of a rare event is mainly characterized by the exponential rate $I$. Note that a large $I$ means a small probability. The $o(1)$ term contains the second order asymptotics, which give information on how fast the exponential curve is approached.

In this paper, we will use Cramér’s Theorem several times, to calculate the above explained rate. This theorem provides an expression for a large deviation of the mean of $n$ i.i.d. random variables $X_i$. If the moment generating function $\phi(t) = \mathbb{E}e^{tX_1} < \infty$ for all $t \in \mathbb{R}$, and $\mathbb{E}X_1 < a$, then

$$ I_a = - \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\frac{1}{n} \sum_{i=1}^{n} X_i \geq a) = \sup_{t \geq 0} (at - \log \phi(t)) $$

(6)

We will also use the 'largest exponent wins' principle ([14], Chapter I). This principle is about the rate of the sum of two (or more) probabilities. Because of the minus sign, a smaller $I$ means a larger exponent, and thus a larger probability. Thus, if we have

$$ P_{1,n} \sim e^{-nI_1} \quad \text{and} \quad P_{2,n} \sim e^{-nI_2} $$

then

$$ - \lim_{n \to \infty} \frac{1}{n} \log(P_{1,n} + P_{2,n}) \sim \min\{I_1, I_2\}. $$

(7)

In words, the principle states that as $n \to \infty$, the smallest exponent (i.e. the largest rate) will become negligible.

Large deviation theory is closely related to importance sampling for simulating events, in our case bit errors, with exponentially small probabilities. Without IS, the number of samples required for an accurate estimate of the BEP would increase exponentially with $n$. The specific transformation for the distribution function is called tilting. With the tilted distribution, the exponential decay is divided out, and only the second order asymptotics remain. This is of course only the case if the right amount of tilting is chosen, in other words, if the exponential rate is known. Otherwise no significant reduction of the required number of samples is obtained. The tilted distribution $\tilde{f}(x)$ of $X$ is given by:

$$ \tilde{f}(x) = \frac{e^{t^* x}}{\mathbb{E}(e^{t^* X})} f(x), $$

where $t^*$ is the maximizer of (6). We see that in addition to finding the supremum, also the original distribution $f(x)$ must be known. In this paper we derive large deviation rates for eigenvalues of the correlation matrix $W$. If the eigenvalue distribution is known, as in the case when $W$ is a Wishart matrix [15], one can directly simulate the eigenvalues, instead of having to simulate $W$. Unfortunately, for our $W$ this distribution is not known. We calculated the rates using a random variable $S_{X,i}^2$ (see Section 5), but we do not know the distribution of this variable explicitly.

### 4 Results

We have succeeded in analytically deriving lower bounds of the large deviation rates for $P_{\max}(\alpha) = \mathbb{P}(\lambda_{\max} \geq \alpha)$, with $\alpha \geq 1$, $P_{\min}(\alpha) = \mathbb{P}(\lambda_{\min} \leq \alpha)$, with $\alpha \in (0, 1]$ and $P_{\min}(0) = \mathbb{P}(\lambda_{\min} = 0)$. Outside these ranges the rates are of course $\infty$, since the corresponding probability is zero.

**Theorem 4.1** Denoting with $I_{\min}(\alpha)$ the rate of $P_{\min}(\alpha)$, and with $I_{\max}(\alpha)$ the rate of $P_{\max}(\alpha)$,

- $I_{\min}(\alpha) = \log 2$, $\alpha = 0$,
- $I_{\min}(\alpha) \geq \frac{1}{2}(-\alpha + \log 2)$, $0 < \alpha \leq 1/2$,
- $I_{\min}(\alpha) \geq \frac{1}{2}(\alpha - 1 - \log \alpha)$, $1/2 \leq \alpha \leq 1$,
- $I_{\max}(\alpha) \geq \frac{1}{2}(\alpha - 1 - \log \alpha)$, $\alpha \geq 1$.

These rates are valid for fixed $k$, and for $k \to \infty$ such that $k \leq n/\log n$.

With these, we can straightforwardly calculate upper bounds for the BEP. As explained in the system model, for the decorrelator a bit error can only occur when $\lambda_{\min} = 0$. Therefore, $I_{\text{dec}} \geq I_{\min}(0)$. For infinite stage SD-PIC, we can also have a bit error when $\lambda_{\max} \geq 2$. Using (7), we conclude that $I_{\text{SD}} \geq \min\{I_{\max}(2), I_{\min}(0)\}$. Inserting the expressions from Theorem 4.1, we find

**Corollary 4.2** The rates for the bit error probability of the infinite stage SD-PIC procedure and the decorrelator, for a lightly loaded $(k \text{ fixed})$ or moderately loaded system $(k \to \infty$ such that $k \leq n/\log n$ and without AWGN), are respectively:

$$ I_{\text{SD}} \geq 1/2 - 1/2 \log 2 = 0.153 \ldots $$

$$ I_{\text{dec}} \geq \log 2 = 0.693 \ldots $$
Theorem 4.3 In the presence of a small amount of noise, the above bounds change into

\[ I_{SD} \geq \frac{1}{2} - \frac{1}{2} \log 2 = 0.153 \ldots \]

\[ I_{dec} \geq \frac{1}{2} \log 2 = 0.346 \ldots \]

What do these numbers mean? For example, when we consider the decorrelator with \( k = 8 \) and \( n = 16 \) in the absence of noise, our first order approximation is \( \text{BEP} \leq e^{-nI} = e^{-16 \log 2} \approx 1.5 \cdot 10^{-5} \). In this simple case, we can even calculate the BEP (see the proof of Theorem 4.1 (a)) as \( 2 \cdot 28(1/2)^{10} = 8.5 \cdot 10^{-4} \), plus some negligible terms. In the presence of a small amount of noise, we approximate \( \text{BEP} \leq 3.9 \cdot 10^{-3} \). Of course these numbers will be more accurate for larger \( n \), as long as \( k \) increases at most proportional to \( n/\log n \). In comparison, a simulation of ([5], Fig. 5) with the same \( k \) and \( n \), and a small amount of noise, gives a BEP of approximately \( 2.2 \cdot 10^{-3} \). We conclude that the results agree.

Comparison of decoding systems The rates are a powerful tool for comparing different decoding systems. For instance, our result immediately shows that for lightly and moderately loaded systems, SD-PIC is superior to HD-PIC in the infinite stage limit. The rate for infinite stage HD-PIC without noise has been derived in [8] and approaches \( \frac{1}{2} \log 2 - \frac{1}{3} = 0.096 \ldots \) as the number of users increases. This result extends to the moderately loaded system, in this case with \( k \leq \delta n \), with small \( \delta \).

Furthermore, as noted above, the largest admissible value for \( \lambda_{\text{max}} \) can be increased using weighted SD-PIC. When this value is increased to 3, already the BEP rate becomes dominated, in the sense of (7), by the event \( \lambda_{\text{min}} = 0 \), so that the BEP will asymptotically be the same as for the decorrelator. To check this, insert \( \alpha = 3 \) in \( I_{\text{max}}(\alpha) \) in Theorem 4.1. To compare the above BEP rates with that of the MF system, we note that (see [9])

\[ I_{\text{MF}} = \frac{k - 2}{2} \log \left( \frac{k - 2}{k - 1} \right) + \frac{k}{2} \log \left( \frac{k}{k - 1} \right) \]

This function decreases with \( k \), \( (I_{\text{MF}} \sim \frac{1}{2k} \) for \( k \to \infty \)) and is already below the values in Corollary 4.2 for \( k \geq 5 \).

In [9], numerical results are given for 1-stage SD-PIC. It is argued that this rate function depends on \( k \) as \( I_{1-\text{SD}} \sim \frac{1}{2\sqrt{k}} \) for large \( k \), and is below our value for infinite stage SD-PIC for \( k \geq 9 \).

5 Proofs

Proof of Theorem 4.1; the case that \( \alpha = 0 \). Since for eigenvectors \( w_j \) of \( W \),

\[ \lambda_j = \langle w_j, W w_j \rangle = \frac{1}{n} \langle w_j, C C^T w_j \rangle = \frac{1}{n} \| C^T w_j \|^2, \]

we must have in this case that \( \| C^T w_j \|^2 = 0 \).

The most probable realisation of this is to have two columns of \( C^T \) equal or opposite in sign. In that case \( w_j \) has only two nonzero components. As the columns have length \( n \), and we have \( k \) columns to choose from, the probability of this is bounded by \( 2(k(k-1)/2)^{2-n} \). When we calculate the large deviation rate, only the factor \( 2^{-n} \) contributes. The other factors disappear as we let \( n \to \infty \). We now see why we need not consider the possibility that \( w_j \) has more nonzero components. Then more columns of \( C^T \) need to be fixed, which will have rate \( 2 \log 2 \) or larger. By (7), these contributions too will disappear.

The cases that \( \alpha \neq 0 \). We will use Cramér’s Theorem. As noted above, for any vector \( x \) with length \( k \), we have

\[ \langle x, W x \rangle = \frac{1}{n} \| C^T x \|^2 = \frac{1}{n} \sum_{i=1}^{k} \left( \sum_{m=1}^{k} x_m C_{mi} \right)^2. \]

We define

\[ S_{x,i} = \sum_{m=1}^{k} x_m C_{mi}, \]

so that

\[ \langle x, W x \rangle = \frac{1}{n} \sum_{i=1}^{n} S_{x,i}^2. \] (8)

Recalling the definition of \( C \), we see that \( \{ S_{x,i} \}_{i=1}^{n} \) is an i.i.d. sequence of random variables with mean 0 and variance \( \| x \|^2 \). Next, we define the function \( I(\alpha) \) by

\[ I(\alpha) = \sup_{x: \| x \|^2 = 1} \frac{\alpha}{2} - \log E(e^\alpha S_{x,i}^2). \]

This is the exponential rate function for the mean of \( S_{x,i}^2 \), maximized over \( x \). We can now state two propositions from which Theorem 4.1 follows for nonzero \( \alpha \). We comment on the discontinuity of \( I_{\text{min}}(\alpha) \) at \( \alpha = 0 \) in the proof of Proposition 5.2.

Proposition 5.1

(a) For all \( \alpha \geq 1 \) and fixed \( k \)

\[ \lim_{n \to \infty} - \frac{1}{n} \log P_{\text{max}}(\alpha) = I(\alpha), \]

(b) For all \( \alpha \leq 1 \) and fixed \( k \)

\[ \lim_{n \to \infty} - \frac{1}{n} \log P_{\text{min}}(\alpha) = I(\alpha), \]
(c) $a-b$ remain valid for $k_n \to \infty$ such that $k_n \log k_n = \mathcal{O}(n)$.

**Proposition 5.2**

(a) For all $\alpha \geq 1/2$

\[ I(\alpha) \geq \frac{1}{2}(\alpha - 1 - \log \alpha). \]

(b) For all $0 < \alpha \leq 1/2$

\[ I(\alpha) \geq \frac{1}{2}(-\alpha + \log 2) \]

**Proof of Propositions 5.1(a-b).**

We will prove part (a) and (b) simultaneously.

We write

\[ P_{\text{min}}(\alpha) = \mathbb{P}(\exists x : \|x\|_2 = 1, \langle x, Wx \rangle \leq \alpha), \]

\[ P_{\text{max}}(\alpha) = \mathbb{P}(\exists x : \|x\|_2 = 1, \langle x, Wx \rangle \geq \alpha). \]

The proof for $\lambda_{\text{max}}$ will be identical to the one for $\lambda_{\text{min}}$, so we will focus on the latter.

The above is a union of events, since to investigate if there are any $x$ with norm 1 such that $\langle x, Wx \rangle \leq \alpha$, we have to check them one by one. We will approach this probability from below by considering only one $x$, and from above by summing over all $x$. Since there are uncountably many, we will do this approximately by summing over a discrete number of vectors. The main difficulty of the proof lies in demonstrating that this approximation is allowed.

The smaller probability will give an upper bound for the rate, and vice versa.

Thus to obtain the upper bound, we use that for any $x'$ with $\|x'\|_2 = 1$, including the $x$ that appears in $I(\alpha)$,

\[ \mathbb{P}(\exists x : \langle x, Wx \rangle \leq \alpha) \geq \mathbb{P}(\langle x', Wx' \rangle \leq \alpha), \quad (9) \]

Now insert (8). Since $x'$ is fixed, the $S_{x',i}$ are i.i.d. variables, and Cramér’s Theorem directly gives $I(\alpha)$, with the supremum replaced by $x'$. We then maximize over $x'$ to get the upper bound.

To obtain the lower bound, we wish to sum over all possible $x$. We approximate the sphere $\|x\|_2 = 1$ by a discrete set of vectors $x^{(i)}$ with $\|x^{(i)}\|_2 = 1$, such that the distance between two of these vectors is at most $d$, and

\[ \|\langle x, Wx \rangle - \langle x^{(i)}, Wx^{(i)} \rangle\| = \]

\[ \|\langle x-x^{(i)}, Wx \rangle - \langle x^{(i)}, W(x-x^{(i)}) \rangle\| = \]

\[ \|\langle x, W(x-x^{(i)}) \rangle - \langle x^{(i)}, W(x-x^{(i)}) \rangle\| \leq (\|x\| + \|x^{(i)}\|) \cdot \|W\| \cdot \|x-x^{(i)}\| \leq 2\lambda_{\text{max}}d \]

Below, we use that $\lambda_{\text{max}} \leq k$, the trace of $W$.

It follows that

\[ P_{\text{min}}(\alpha) \leq \mathbb{P}(\exists x^{(i)} : \langle x^{(i)}, Wx^{(i)} \rangle \leq \alpha - 2dk) \]

\[ \leq \sum_i \mathbb{P}(\langle x^{(i)}, Wx^{(i)} \rangle \leq \alpha - 2dk) \]

\[ \leq N_d \sup_{x^{(i)}} \mathbb{P}(\langle x^{(i)}, Wx^{(i)} \rangle \leq \alpha - 2dk), \quad (10) \]

with $N_d$ the number of vectors needed to cover the sphere. A simple overestimation of this number is obtained by inflating the sphere into a unit cube around the origin. We then lay a grid on this cube with grid length $\frac{1}{L}$. In this case,

\[ d \leq \frac{\sqrt{k}}{L}, \quad \text{and} \quad N_d = L^{k-1}. \quad (11) \]

We need to choose $L$ such that $2dk$ and $\frac{1}{L} \log N_d$ vanish when $n \to \infty$. When $k$ is fixed, this is easily achieved, for example by choosing $L = n$. When the two terms have vanished, we have in fact the same expression as in (9), and the lower bound follows.

The proof of Proposition 5.1(c) is similar to the above, but involves a more elaborate estimation of $N_d$. It is therefore given in the appendix.

**Proof of Proposition 5.2.**

To evaluate $I(\alpha)$, we first replace

\[ \mathbb{E}(e^{\sqrt{t}S_{x,i}}) \leq \frac{1}{\sqrt{1-t\|x\|^2}}. \quad (12) \]

If $S_{x,i}$ would have a normal distribution, this would be an equality. When inserting $x' = \frac{1}{\sqrt{k}}(1, 1, \cdots)$ in (9), by the central limit theorem $S_{x',i} = \frac{1}{\sqrt{k}} \sum_{m=1}^{k} C_{mi}$ converges to a standard normal distribution as $k \to \infty$. For any $k$ and $x$, the bound (12) is valid for all $-1 \leq t\|x\|^2 \leq 1$ ([8], Section IV). We use the above expression to arrive at:

\[ I(\alpha) \geq \frac{1}{t} \frac{\alpha}{2} - \log \frac{1}{\sqrt{1-t\|x\|^2}}. \]

Note that since $\|x\| = 1$, the bound is independent of $x$. Maximizing over $t$ is straightforward, and yields $t = 1 - \frac{1}{\alpha}$. When $\alpha < 1/2$, $t$ becomes smaller than -1, where (12) is not valid. In that case, we insert $t = -1$ as a lower bound for the maximum.

Inserting these $t$ in $I(\alpha)$ gives the desired expressions.
Theorem 4.1 (a), $I_{\min}(0) \geq \log 2$, whereas Proposition 5.2 only predicts $\lim_{n \to 0} I_{\min}(\alpha) \geq 1/2 \log 2$.

We can use (9) again with $x = \frac{1}{\sqrt{2}} (1, 1, 0, \cdots)$. For this vector, $E(e^{\frac{1}{2}S_{x}^{2}}) = 1/2(e^1 + 1)$, and calculating the according rate gives $I(\alpha) \leq 1/2(\alpha \log_2 \alpha + (2 - \alpha) \log_2(2 - \alpha))$, with which $\lim_{\alpha \to 0} I_{\min}(\alpha) \leq \log 2$.

**Proof of Theorem 4.3** We wish to obtain a rate for (5), the probability that the additive noise causes a bit error, in the case $W^{-1}$ exists. First, we note that

$$\text{BEP}_{\text{noise}} = \mathbb{E} \left[ Q \left( \frac{n P_m}{\sigma^2 W_{m,m}} \right) \right] \leq \mathbb{E} \left[ Q \left( \sqrt{\frac{n P_m \lambda_{\min}}{\sigma^2}} \right) \right].$$

We wish to obtain a rate $I_{\text{noise}}$ for this expression using Varadhan’s Lemma ([14], chapter III). This lemma states that if the random variable $x$ obeys a large deviation principle on $X$ with rate function $I(\alpha)$, and $F(x)$ is a continuous function with $F(x) < \infty$ on $X$, then

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}[e^{n F(X_n)}] = \sup_{a \in X} [F(a) - I(\alpha)],$$

We rewrite (13) in this form. First, we use that $Q(x) le q e^{-x^2/2}$ for all $x \geq 0$. This will make a negligible difference, since factors that are sub-exponential in $n$ have no influence on the rate.

Since $F(x)$ in this case is new $F(\lambda_{\min}) = -\frac{P_m \lambda_{\min}}{2 \sigma^2}$, we need to show that $\lambda_{\min}$ obeys a large deviation principle. Our lower bounds for the rate function $I_{\min}(\alpha)$ can be shown to have the right properties.

We insert these in (14), to obtain an upper bound for the supremum. For large enough $P_m / \sigma^2$, the first term will dominate, and the supremum will be reached at $\alpha = 0$. Recall that we can not use our result for $\alpha = 0$, since in that case $W^{-1}$ would not exist. We therefore find

$$I_{\text{noise}} \geq \frac{1}{2} \log 2,$$

Comparing this rate with the rates in Corollary 4.2, and selecting the smaller value in accordance with (7), Theorem 4.3 follows.

### 6 Conclusions

Large deviation theory has been shown to be a powerful tool for comparing decoding techniques for lightly and moderately loaded CDMA systems. We have compared MF, infinite stage HD-PIC and SD-PIC, and the decorrelator without noise. Of these, the decorrelator comes out most favorably. In the infinite stage limit, SD-PIC performs better than HD-PIC, but both do better than the simple MF decoder. A small amount of noise does not change this order.

We reached all these results analytically. Simulations would provide useful complementary information, such as how many stages are necessary before the above results apply. The large deviation rates form an important first step towards simulations using importance sampling.

Interesting remaining problems are to implement these simulations, to include the MMSE detector in our model, and to obtain more analytic results on the influence of AWGN noise.

### Appendix

This appendix gives the proof of Proposition 5.1(c). We only need to show that the lower bound remains valid. As in the proof of Propositions 5.1(a-b), we wish to show that the terms $\frac{1}{n} \log N_d$ and $2d \lambda_{\max}$ vanish when we take the logarithm of (10), divide by $n$ and let $n \to \infty$, but this time we wish to let $k_n \to \infty$ as well, for $k_n$ as large as possible.

The overestimation (11) can be improved using an upper bound for the number $N_L$ of spheres of radius $1/L$ needed to cover a $k$-dimensional sphere of radius $1$ [16],

$$N_L < k (k \log k + k \log \log k + 5k) L^k \equiv f(k) L^k.\tag{15}$$

This bound is valid for $\frac{k}{\log L} \leq L < k$. Since we use small spheres this time, $d \leq 1/L$. We can also improve the upper bound $\lambda_{\max} \leq k$. We write:

$$P_{\min}(\alpha) \leq \mathbb{P}(\lambda_{\min} \leq \alpha \wedge \lambda_{\max} \leq \Omega) + P_{\max}(\Omega)$$

where

$$P_{\max}(\alpha) = \mathbb{P}(\alpha \leq \lambda_{\max} \leq \Omega) + P_{\max}(\Omega).\tag{16}$$

$\Omega$ should be thought of as large, but fixed. The idea is that the first terms of these expressions will yield the rate $I_\alpha$, plus an undesired term $O(\frac{k \log k}{n})$. The terms $P_{\max}(\Omega)$ will give another undesired term $O(\frac{k \log k}{n})$, but this time minus a term $O(\Omega)$. This means that when $\frac{k \log k}{n}$ is a constant, we can choose $\Omega$ large enough to vanquish the undesired terms.

First, we calculate the second term of (16), using (10). This is allowed since at that stage in the proof of Proposition 5.1(a-b), the restriction that $k$ is fixed, is not yet used.

$$P_{\max}(\Omega) \leq N_L \sup_{x} \mathbb{P}(x, W x) \geq \Omega - 2k/L.$$
Inserting (15), and choosing $L = \frac{1}{2} k$ (since $L = k$ is not allowed), this becomes

$$P_{\max}(\Omega) \leq f(k)(k/2)^k \sup_{x} \mathbb{P}(\langle x, Wx \rangle \geq \Omega - 2).$$

Using Proposition 5.2, we find

$$P_{\max}(\Omega) \leq f(k)(k/2)^k e^{-\frac{1}{2} n (\Omega - 5 - \log(\Omega - 4))}.$$  

This probability has is exponentially small with rate $\frac{k \log k}{n} - \frac{1}{2} (\Omega - 5 - \log(\Omega - 4)) + o(\frac{k \log k}{n})$, as long as $\frac{k \log k}{n} < \frac{1}{2} (\Omega - 5 - \log(\Omega - 4))$.

Next, we investigate the first term of (16). In this term, we can use $\Omega$ as upper bound of $\lambda_{\max}$. Therefore, again starting with (10),

$$I_{\min}(\alpha) \geq - \lim_{n \to \infty} \frac{1}{n} \left( \log N_L + \sup_{x} \mathbb{P}(\langle x, Wx \rangle \leq \alpha + 2 \Omega/L) \right).$$

For $\lambda_{\max}$, we get a similar expression. Inserting (15), we need to choose $L$ again. For any $k/\log k \leq L < k$, the terms $2\Omega/L$ vanish, and the term $\frac{1}{n} \log N_L$ contains a dominating term $\frac{k \log k}{n}$, so that the rate of the first term of (16) is $I(\alpha) + \frac{k \log k}{n} + o(\frac{k \log k}{n}).$

Putting the two rates together and applying the largest-exponent-wins principle (7), we find that Propositions 5.1(a-b) remain valid when $I(\alpha) < \frac{1}{2} (\Omega - 5 - \log(\Omega - 4)) - 2 \frac{k \log k}{n}$. Therefore, $k_n \log k_n = O(n)$ is allowed. \hfill \Box

References


