Mikko Vehkaperä

# Statistical Physics Approach to Design and Analysis of Multiuser Systems Under Channel Uncertainty

Thesis for the degree of Philosophiae Doctor

Trondheim, August 2010

Norwegian University of Science and Technology
Faculty of Information Technology, Mathematics
and Electrical Engineering
Department of Electronics and Telecommunications

NTNU
Norwegian University of
Science and Technology

**NTNU**
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology, Mathematics and Electrical Engineering
Department of Electronics and Telecommunications



ISBN 978-82-471-2256-3 (printed ver.)
ISBN 978-82-471-2258-7 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2010:145

Printed by NTNU-trykk

# Abstract

Code-division multiple-access (CDMA) systems with random spreading and channel uncertainty at the receiver are studied. Frequency selective single antenna, as well as, narrowband multiple antenna channels are considered. Rayleigh fading is assumed in all cases. General Bayesian approach is used to derive both iterative and non-iterative estimators whose performance is obtained in the large system limit via the replica method from statistical physics.

The effect of spatial correlation on the performance of a multiple antenna CDMA system operating in a flat-fading channel is studied. Per-antenna spreading (PAS) with random signature sequences and spatial multiplexing is used at the transmitter. Non-iterative multiuser detectors (MUDs) using imperfect channel state information (CSI) are derived. Training symbol based channel estimators having mismatched a priori knowledge about the antenna correlation are considered. Both the channel estimator and the MUD are shown to admit a simple single-user characterization in the large system limit. By using the decoupled channel model, the ergodic spectral efficiency with single-user decoding and quarternary phase shift keying (QPSK) constrained modulation is derived. In contrast to the case of perfect CSI where transmit correlation has no effect on the ergodic system performance with random PAS, the results show that with channel estimation the ergodic capacity can improve significantly as the correlations between the transmit antennas increase. This requires that the channel estimator knows the correct long term spatial correlation in advance, while no information is required at the transmitter.

Iterative multiuser receivers for randomly spread CDMA over a frequency selective Rayleigh fading channel are analyzed. General Bayesian approach for iterative channel estimation and data detection and decoding is proposed. Both linear and non-linear iterative schemes are considered with soft or hard information feedback. The equivalent single-user representation of the system is derived in the large system limit via the replica method. The decoupled single-user channel, and density evolution with Gaussian approximation are used to obtain the spectral effi-

ciency and bit error rate (BER) of the system using bit-interleaved coded modulation (BICM) and Gray encoded QPSK mapping. The results indicate that in the large system limit and under certain threshold loads, near single-user BER performance with perfect CSI can be achieved by using a vanishing training overhead. This requires, however, an iterative receiver using soft feedbacks only. For relatively slowly time-varying multipath fading channels, the iterative linear minimum mean square error (LMMSE) based channel estimator is also shown to be near optimal in terms of maximizing the spectral efficiency of the system when combined with iterative LMMSE or maximum a posteriori multiuser detectors.

A novel training method based on probability biased signaling is proposed. By assuming an entropy maximizing biasing scheme and standard BICM, it is shown via numerical examples that the proposed training method can offer superior performance over the conventional training symbol based approach when combined with iterative receivers.

# *Preface*

This dissertation is submitted in partial fulfillment of the requirements for the degree of Philosophiae doctor (PhD) at the Department of Electronics and Telecommunications, Norwegian University of Science and Technology (NTNU). The research was carried out in the period from October 2006 to April 2010 and was funded by the Norwegian Research Council (NFR) under grant 171133/V30 through the project Statistical Physics of Advanced Multiuser Communications (SPAM). The advisor for the thesis has been Prof. Ralf R. Müller who is with the Department of Electronics and Telecommunications at NTNU.

## Acknowledgment

I am deeply grateful to my advisor Prof. Ralf R. Müller who gave me the opportunity to work on the topics covered in this thesis. His impeccable insight and intuition (that turned out to be correct time and time again) has been the greatest source of inspiration during this work. I have also great memories of our discussions on more mundane topics and trips to lands far far away.

I am indebted to my co-authors Prof. Toshiyuki Tanaka from the Graduate School of Informatics, Kyoto University and Prof. Keigo Takeuchi from the Department of Communications Engineering and Informatics, University of Electro-Communications, Tokyo. Working with them has been a privilege and their help and guidance in the replica analysis invaluable to my research. I would like to thank Prof. Tanaka also for arranging my visit to Kyoto University.

I would like to express my gratitude to the adjudication committee of the thesis; Prof. Sergio Verdú from Princeton University, Prof. Giuseppe Caire from University of Southern California and Prof. Asle Sudbø from NTNU for carefully reviewing the present dissertation.

My sincere appreciation goes to all my colleagues in the Signal Processing Group at NTNU who created such a friendly and inspiring atmosphere for research

work. Special thanks goes to Mrs. Kirsten Marie Ekseth, who has saved me from many troubles with bureaucracy. I would also like to thank my colleagues Dr. Rodrigo Vicente de Miguel de Juan, Dr. Pål Anders Floor, Dr. Kimmo Kansanen and Dr. Pierluigi Salvo Rossi with whom I had many great discussion both technical and philosophical.

Finally, I would like to express my gratitude to my mother Leena, father Reijo, and sister Kaisa for all the encouragement and support during this work. The greatest and warmest thank, however, belong to Maija for her love, support and understanding during the preparation of this thesis.

Helsinki, Finland, July 2010
Mikko Vehkaperä

# *Contents*

# Abbreviations

| | |
|---|---|
| 3G | third generation |
| APP | a posteriori probability |
| AWGN | additive white Gaussian noise |
| BEP | bit error probability |
| BER | bit error rate |
| BICM | bit-interleaved coded modulation |
| BPSK | binary phase shift keying |
| CE | channel estimation |
| CSI | channel state information |
| DS-CDMA | direct sequence code division multiple access |
| ECC | error control code |
| EXIT | extrinsic information transfer |
| GPME | generalized posterior mean estimator |
| GSM | Global System for Mobile Communications |
| IID | independent identically distributed |
| IO | individually optimum |
| ISI | intersymbol interference |
| LHS | left hand side |
| LMMSE | linear minimum mean square error |
| MAI | multiple access interference |
| MAP | maximum a posteriori |
| MC | multicarrier |
| ME | multiuser efficiency |
| MIMO | multiple input multiple output |
| ML | maximum likelihood |
| MUD | multiuser detection |
| MUDD | multiuser detection and decoding |
| PAS | per-antenna spreading |

| | |
|---|---|
| PDP | power delay profile |
| PIC | parallel interference cancellation |
| PSK | phase shift keying |
| PUS | per-user spreading |
| QPSK | quarternary phase shift keying |
| RHS | right hand side |
| RMT | random matrix theory |
| RS | replica symmetry |
| RSB | replica symmetry breaking |
| RV | random vector |
| S-K | Sherrington-Kirkpatrick |
| SIC | serial interference cancellation |
| SIMO | single input multiple output |
| SINR | signal to interference and noise ratio |
| SNR | signal to noise ratio |
| SUMF | single-user matched filter |
| UMTS | Universal Mobile Telecommunication System |
| ZF | zero-forcing |

# List of Figures

# Chapter 1

# Introduction

The main emphasis in wireless communication systems has been gradually shifting during the recent years from relaying voice calls to providing mobile access to web-based services, such as, high quality multimedia streaming. This has increased the data rate demands by several orders of magnitude, and future broadband mobile communication systems are envisioned to provide throughputs of up to 1 Gbps. In order to achieve such goals, broad bandwidth and physical layer techniques with high spectral efficiency are required — not forgetting the sophisticated higher layer scheduling and resource allocation methods that maximize the long-term throughput. The present dissertation concentrates on analyzing some of the approaches in the former category.

The rest of this chapter is organized as follows. A very general overview of the topics discussed in this thesis is first given in Section 1.1. The presentation and references there are selected to be accessible to as wide audience as possible. A more detailed and technical literature review of the relevant areas is presented in Section 1.2. Finally, Section 1.3 discusses the aims of the research work and provides the outline for the rest of the dissertation.

## 1.1 Background

The current state-of-the art cellular system in Europe is the third generation (3G) Universal Mobile Telecommunication System (UMTS) [1, 2], based on wideband direct sequence code division multiple access (DS-CDMA) technology [3–6]. Although the original specification of the UMTS network contained very basic physical layer techniques, several technological evolutions have added new features to it, such as multi-antenna transmission [7–9] and more sophisticated multiuser de-

tection (MUD) [4, 5, 10]. Both of the proposed schemes lead to increased user data rates as well as cell capacity, and further progress is on-going for still new techniques that can help to achieve the goals set for future wireless systems.

### 1.1.1 Multiple Access

Let us consider the UMTS system mentioned above. The signals traveling between the user terminal and the base station can be divided into two rough categories:

- Signals emanating from the base station (downlink transmission);
- Signals emanating from the user terminal (uplink transmission).

In a multiuser CDMA system, the downlink transmission corresponds to a one-to-many scenario, that is, the transmitted signal has one physical source and several geographically dispersed receivers that are operating independently of each other. The signal transmitted by the base station may contain information that is shared by all users, or what is more common, each user wishes to decode its own unique message embedded in the transmission. The uplink transmission is in a sense the mirror image of the above, i.e., a many-to-one scenario. In this case, several independent user terminals are trying to send their messages to a common destination, that has to decode each of them from what ever it receives through the channel. In wireless systems one of the problems encountered is that the radio waves transmitted at the same time and on the same frequency band interfere with each other in the propagation medium. Therefore, some form of signal processing is again needed at the receiver to reliably separate the messages coming from different users. In information theory, the one-to-many scenario is called the broadcast channel, whereas the many-to-one is the multiple access channel [11]. In this thesis we consider only the uplink, or multiple access case, where several simultaneous transmissions emanate from the non-cooperative user terminals and the base station receives a corrupted superposition of these signals.

In the following, the only multiple access technique that will be discussed in detail is CDMA. More precisely, we consider a form of non-orthogonal code division multiple access where the signature sequences of the users are drawn according to a predefined probability distribution [3, 5]. This is in contrast to multiple access techniques such as time division multiple access and frequency division multiple access [6], where the degrees of freedom are typically allocated to the users in a deterministic manner to guarantee orthogonality. The disadvantage of non-orthogonal multiple access schemes like the randomly spread CDMA is that they require relatively sophisticated signal processing at the receiver in order to operate efficiently.

On the other hand, since the number of users that can access the system simultaneously is not strictly bounded by the number of orthogonal dimensions present in the transmission, there is a greater flexibility in controlling the user loads within a cell. For our purposes, however, the most important benefit of random spreading is that well developed mathematical tools to analyze the performance of such systems exists [12–15].

### 1.1.2 Iterative Processing

One of the paradigms of modern signal processing are the iterative, or so-called turbo, algorithms [16]. The basic motivation behind the iterative algorithms is summarized in the following quote of Viterbi [17]:

> *Never discard information prematurely that may be useful in making a decision until after all decisions related to that information have been completed.*

More precisely, we would like two or more subsystems that are capable of operating indepently to exchange information between each other in a manner that makes iterative refinement of their initial outputs possible. As it turns out, the most important condition for the successful execution of an iterative algorithm is to make sure that the information received by a subsystems is extrinsic to it.

The iterative principle has been applied with a great success, e.g., to intersymbol interference (ISI) cancellation, iterative multiuser detection and decoding (MUDD), iterative channel estimation (CE), decoding of error control codes (ECCs), and so on [16, 18–24]. Typically two different information exchange, or feedback, strategies called "soft" and "hard" are identified. The difference between the two is that in the former, the feedback contains all the information that is obtained by the signal processing block, whereas intermediate quantization is performed in the latter before the other subsystem are allowed to use it. It is commonly accepted that the former guarantees a better performance since the intermediate quantization leads to an inevitable loss of information, but the latter might be easier to implement in a practical system.

In this thesis, the emphasis is on the iterative MUDD and channel estimation methods. Both soft and hard feedback schemes are considered. The estimation algorithms are derived from the factor graph [21–26] representation of the system. One of the advantages of this approach is that the iterative process can then be analyzed by using, for example, density evolution or extrinsic information transfer

(EXIT) charts [21–24]. We shall concentrate on tracking the information exchange between different blocks by using density evolution with some simplifications, as will be explained in the later chapters.

### 1.1.3 Multiple Input Multiple Output Channels

The use of multiple antennas at the receiver for diversity and array gain in single-user systems has been in the telecommunication engineers' tool bags for a long time [6, 8, 9]. One of the examples is the uplink of Global System for Mobile Communications (GSM), where the receiving base station is equipped with multiple antennas to mitigate the effects of low transmit power of the user terminal. Multiple antennas at the receiver can also be used for interference avoidance, or even as a multiple access scheme in multiuser systems [8, 9]. Recently the possibility for diversity and array gains was extended to the case of multiple transmit antennas as well, by using space-time codes [27] and transmit beamforming [8, 9], respectively. One of the drawbacks of transmit beamforming, however, is that it requires (some form of) channel state information (CSI) at both the transmitter and the receiver.

For future communication systems where the design goal is to enable high data rates, the greatest promise of multiple antennas is not the diversity, but the possibility to multiplex several simultaneous transmissions in space. This requires an antenna array at both ends, creating a so-called multiple input multiple output (MIMO) channel between the transmitter and the receiver. The significance of the MIMO channel is that if the environment is rich scattering, spatial multiplexing can provide a dramatic increase in capacity compared to an equivalent single-antenna system. This is usually called degree-of-freedom, or spatial multiplexing, gain. For a MIMO system with $M$ transmit antennas and $N$ receive antennas, the capacity increases roughly linearly in $\min\{M, N\}$ if the system is operating over a fast Rayleigh fading channel that is perfectly known at the receiver [7–9]. This is a significant improvement over the logarithmic increase in spectral efficiency as a function of signal to noise ratio (SNR). Furthermore, to achieve this, there is no need for transmitter CSI and linear minimum mean square error (LMMSE) filter with successive interference cancellation (SIC) suffices at the receiver.

From a practical point of view, one caveat in the above discussion is that the capacity achieving signaling scheme is Gaussian. Such a continuous modulation over an uncountable signal set is not a feasible choice for a real-life system. Unfortunately, the LMMSE estimator with SIC is not an optimum decoding method for modulation schemes used in practice, such as phase shift keying (PSK), even when

combined with optimum codes. The situation is even worse for off-the-shelf ECCs that tend to suffer from error propagation if naïve interference cancellation is used. Nevertheless, multiple antennas combined with the aforementioned iterative algorithms and efficient ECCs provide significant benefits for practical systems as well. Research efforts to find new and improved MIMO schemes that can be realized in the future wireless systems are naturally still going on.

## 1.2 Review of Earlier and Parallel Work

In this section we provide references to more technical literature regarding the topics relevant for the future discussion. The following is not meant to be by any means a comprehensive review of earlier work and emphasis is put on literature related to iterative signal processing and multiuser detection. Due to their more comprehensive nature, journal articles are in general preferred to papers published in conference proceedings. Interested reader will find the first published results in the cited article.

### 1.2.1 A Brief History of Multiuser Detection

The simplest non-trivial data detector for a DS-CDMA system is the single-user matched filter (SUMF), or the conventional detector. It is well known that the SUMF is optimal for single-user communication and for equivalent cases, such as, synchronous narrowband multiuser systems with orthogonal signature sequences [4–6, 10]. Unfortunately, for modern wideband CDMA systems, such as the UMTS, several problems arise with the use of the conventional detector. First, even if the transmissions are synchronous, multipath propagation tends to destroy the orthogonality of the signature waveforms in which case the SUMF becomes highly suboptimal. Second, unless strict uplink power control is employed, the near-far problem arises due to highly unequal received signal powers between the users. At its simplest, this causes the signals of the users who are far from the base station to drown under the transmissions of the users who are near the base station. In addition to strict power control, techniques such as handover are employed in UMTS network to mitigate this problem [1, 2]. All of this adds up to the fact that in practical scenarios the conventional detector tends to suffers from the near-far problem and severe multiple access interference (MAI) [4, 5, 10], even if orthogonal spreading waveforms are used.

The first published proposal for improved detection in multiuser systems using non-orthogonal spreading sequences and operating in additive white Gaussian noise

(AWGN) channels was made by Schneider [28]. Due to an error in the derivation, however, instead of being the optimum detector, the resulting MUD was a linear estimator known nowadays as the decorrelating or zero-forcing (ZF) detector. The breakthrough in multiuser detection came somewhat later in the seminal work of Verdú [29–32] (see also [5]), where the concept of multiuser efficiency (ME) was introduced. Oftentimes the ME is considered in the high SNR region where it can be defined implicitly as the constant $\eta_{\sigma^2\to 0} \in [0, 1]$ that satisfies (see [5])

$$\lim_{\sigma^2\to 0} \frac{\varepsilon_{\mathsf{mu}}(1/\sigma^2)}{\varepsilon_{\mathsf{su}}(\eta_{\sigma^2\to 0}/\sigma^2)} = 1. \tag{1.1}$$

The functions $\varepsilon_{\mathsf{su}}(\overline{\mathsf{snr}}) \leq \varepsilon_{\mathsf{mu}}(\overline{\mathsf{snr}})$ above are the average bit error rates (BERs) of the equivalent single-user and multiuser systems in a channel that has the average received SNR of $\overline{\mathsf{snr}} = 1/\sigma^2$. Hence, the asymptotic ME $\eta_{\sigma^2\to 0}$ quantifies the loss in effective SNR due to MAI and describes the interference suppression capabilities of the MUD at high SNR. Interestingly, in [5, 29–32] it was found that in contrast to the conventional detector for which the asymptotic ME $\eta_{\sigma^2\to 0} = 0$ (the system is multiple access limited), the optimum MUD has a non-zero asymptotic ME $\eta_{\sigma^2\to 0} > 0$ for all finite user loads. Furthermore, under certain conditions even single-user performance, i.e., $\eta_{\sigma^2\to 0} = 1$ can be achieved. Unfortunately, the optimum receiver was also found to be non-linear with its complexity increasing exponentially in the number of users.

The interest in detection algorithms that would strike a balance between the performance and complexity lead to the study of low complexity sub-optimum linear MUDs [33–36]. The results showed that the decorrelator and the LMMSE multiuser detectors exhibit similar near-far resistance, i.e., the worst case asymptotic ME, as the optimum MUD but with significantly lower complexity. In addition to the linear detectors, several schemes combining linear MUD with serial [37, 38] and parallel [39, 40] interference cancellation or sequential decoding [41] were soon proposed and analyzed. For a comprehensive overview on the early literature on MUD, see for example, [4, 5] and the references therein.

The common feature in the aforementioned studies was that the performance of the systems under consideration depended on the selection of the deterministic spreading sequences assigned to the users. An alternative approach, first proposed for the study of conventional detectors [42] (see also [3]), is to use random signature sequences instead of deterministic ones and average the performance over the selection of the spreading codes. Random spreading combined with the large system

analysis[1] was used to obtain the high SNR performance of multiuser detectors in [5, 43]. Rather surprisingly, it was found that the optimum MUD achieves asymptotic ME $\eta_{\sigma^2 \to 0} = 1$ with probability one under random spreading. This is not true for the decorrelating or the LMMSE detector for any positive user load $\alpha$, since for these MUDs the asymptotic ME is given by $\eta_{\sigma^2 \to 0} = 1 - \alpha$, where $0 \leq \alpha \leq 1$ [5, 43].

By using methods similar to [5, 43], upper and lower bounds for the sum-capacity of the CDMA channel with joint decoding were derived in [44]. The results showed that random spreading is on average near optimum for large heavily loaded systems. The average near-far resistance of the LMMSE detector [45], and the capacity of decorrelating detector with and without decision feedback [46] of randomly spread DS-CDMA were also considered. It was, however, the introduction of the random matrix theory (RMT) to telecommunications engineering that finally shifted the paradigm in multiuser detection to the randomly spread CDMA systems [12]. We shall investigate this topic more thoroughly in the next section.

### 1.2.2 Large System Analysis: Random Matrix Theory and the Replica Method

One can argue that with the large system analysis of randomly spread CDMA by Verdú [5], (with Shamai) [47, 48] and Tse & Hanly [49, 50], not forgetting the early studies on MIMO systems by Telatar [51, 52] and Foschini & Gans [53, 54], the telecommunications engineering entered the random matrix theory era. The contribution of [47, 48] was the information theoretic analysis of optimum joint decoding as well as linear multiuser detection when combined with capacity achieving Gaussian codes. In [49, 50], the concepts effective interference, effective bandwidth and resource pooling were introduced, allowing for a surprisingly simple characterization of the performance of linear MUDs in fading channels. These results were soon refined to show that the limiting distribution of the signal to interference and noise ratio (SINR) and MAI of the linear detectors was in fact Gaussian [55–57]. Further extensions related to multiuser communications included, for example, symbol-asynchronous CDMA [58] and analysis of DS-CDMA in multipath fading channel when linear MUD and channel estimation is employed at the receiver [59]. The latter of these was in fact one of the main motivations for the research work presented in this thesis. Some other avenues where RMT has found applications in-

[1] Large system analysis refers to the case when the number of users $K$ and the length of the spreading sequence $L$ are allowed to grow without bound with a finite and fixed ratio $\alpha = K/L$.

clude multicell environments [60], precoded transmission [61], multicarrier CDMA (MC-CDMA) [62], multi-antenna systems [51–54, 63–66] and design of novel receivers [67–71] for DS-CDMA, to name just a few. For a comprehensive overview, see [12] and the references therein.

A typical application of random matrix theory from the telecommunications engineering point of view is when we are interested in some scalar parameter (say output SINR of the multiuser detector or the uncoded bit error rate of the system) that is a function of the eigenvalues and eigenvectors of some random matrix. For finite systems these parameters are in general random variables but, under certain conditions, they converge to deterministic values when the dimensions of the random matrix are allowed to grow without bound[2] [12, 73]. Unfortunately, not all problems encountered in telecommunications fall into this category. For example, when the problem to be considered contains a combinatorial optimization problem, the tools of RMT tend to fall short and different approach is needed. As we shall see next, one of these approaches comes from theoretical physics — the original source of random matrix theory as well.

The seminal paper of Tanaka [74] (see also [75] and [76]) was the first one to report the large system performance of randomly spread CDMA with binary phase shift keying (BPSK) and non-linear Bayesian optimum receiver. At the same time, a new mathematical tool called the replica method from statistical physics [13–15] was introduced to the information theory society at large[3]. Tanaka's original result was extended to arbitrary input constellations and fading channels with unequal received power distribution by Guo & Verdú [85] (see also [86]). Some of the concepts that were implicitly introduced in [74], namely the decoupling principle later generalized by Montanari [87] and Tanaka & Nakamura [88], and generalized posterior mean estimation (GPME), were also further developed in [85] (see also [89]).

Soon the replica method was applied to various problems in communications that had so-far evaded analytical treatment. Examples of these included analysis of multicarrier CDMA with non-linear MUD over frequency selective fading [73,

[2]The eigenvalue spectrum of certain random matrices can be described also for finite dimensional cases and these results have their applications in the analysis of wireless systems as well. However, the finite dimensional results tend to be more cumbersome to use and somewhat limited in scope compared to their asymptotic counterparts. For an overview, see [12, 72] and the references therein.

[3]Tanaka's paper was not the only one utilizing the replica method to solve a problem related to telecommunications. For example, regular and irregular low density parity check codes [77, 78] were analyzed in [79, 80], and the parallel concatenated turbo codes [81] were considered in [82, 83]. These papers, however, were published in physics journals and written to an audience that was already familiar with the method. A more thorough survey can be found, for example, in [14, 15, 84].

90, 91], comparison of space-time spreading schemes for general MIMO CDMA [92, 93], capacity and bit error probability (BEP) analysis of MIMO channels with binary inputs [94], study of general vector channels [95], analysis of quadratic optimization problem arising from vector precoding [96] and sum-rate analysis of multiuser MIMO with spatial correlation [97–99]. Slightly different application of the replica method was also used to determine the moments of mutual information of a slow fading correlated Rayleigh fading MIMO channel with Gaussian inputs [100, 101]. It should be remarked, however, that although the replica method is a standard tool in statistical physics, some of the steps in the "replica trick" are lacking formal justification and present an open problem in mathematical physics. For an overview of the topic, see for example [102–104]. Some recent developments related to the Tanaka's original result [74], can be found in [105–107]. Another recent paper [108] considers the first order replica symmetry breaking in the original vector precoding problem investigated in [96].

A common theme in all of the aforementioned studies apart from [59, 90] was that the channel state information was assumed to be perfect at the receiver. In practice this is not a very realistic assumption since for coherent communication the channel must be estimated by some means at the receiver with a finite accuracy. Another technique that is of practical importance and has not yet been discussed are the iterative MUDD and channel estimation schemes. These topics will be the main focus of the next section.

### 1.2.3 Design and Analysis of Iterative Receivers

The optimum receiver for coded CDMA is the maximum likelihood (ML) decoder that simultaneously resolves the messages of all users. For the symbol synchronous CDMA with Gaussian codes, significantly less complex LMMSE data estimator followed by successive interference cancellation can be used alternatively without any loss in the maximum sum-rate [109]. It is important, however, to make a distinction between the interference cancellation methods [37–40] discussed earlier, and the post-decoding IC discussed in this section. The former use uncoded symbols in an effort to remove the MAI, whereas here the IC is performed after decoding (and re-encoding) the ECC. If interference cancellation is omitted in the synchronous case, irreducible loss in capacity is experienced due to the separation of detection and decoding [46–48, 85, 110, 111]. For binary code books, the loss in spectral efficiency with LMMSE estimation is more severe than in the Gaussian case [110]. This is not surprising since LMMSE estimator is optimal for AWGN

channel with Gaussian inputs, whereas for the binary case the optimum MUD is the non-linear maximum a posteriori (MAP) detector [30, 74]. The latter is also commonly known as the individually optimum (IO) MUD in the literature. Adding SIC to the LMMSE-based MUD front-end mitigates the loss in spectral efficiently noticeably but is still a highly suboptimal decoding strategy for discrete channel inputs. It should be remarked though, that smart rate and power control alleviates the performance degradation significantly [112]. In order to achieve the jointly decoded capacity one must, however, combine the Bayesian optimum MAP detector with SIC [85, 110].

With capacity achieving codes the optimum decoding strategy for synchronous randomly spread CDMA is thus the combination of optimum MUD front-end and SIC, regardless of the channel inputs [85, 110]. In practice, however, the decoded signals are not error-free due to delay constraints, code construction and lack of flexibility in choosing the code rate. Straightforward application of SIC in such case can in fact lead to performance loss compared to the case without IC due to error propagation. To prevent this happening, intuitively one should somehow take into account the uncertainty of the feedback symbols when performing the interference cancellation. Such a reasoning combined with the lessons learned from decoding of turbo codes was used to derive algorithms for the iterative multiuser detection and decoding of CDMA transmissions [113–118], decoding of spatially multiplexed transmissions [119, 120], and iterative ISI suppression [121–123]. The performance of these proposals, however, was studied via rather time consuming Monte Carlo simulations (for a further review, see [16]).

In this thesis, the factor-graph [21–26, 124] based iterative multiuser detection and decoding framework proposed by Boutros & Caire [125] (see also related results [126–128]) is endorsed. A notable benefit of this approach is that it provides a formal framework for interference cancellation based receiver design that can be analyzed via density evolution[4] [21, 23, 24, 132, 133]. In addition to iterative MUDD, this methodology has also been used, e.g., in the design and analysis of MIMO systems [134]. Recently, alternative approaches to the derivation of iterative data detection and decoding algorithms in coded systems have also been proposed, for example, based on divergence minimization [135] and variational inference with mean-field approximation [136]. The evaluation of the derived algorithms in these studies was, however, carried out by using computer simulations.

Apart from two exceptions [137, 138] that will be discussed in more detail later,

[4]Performance analysis similar to [125] were also performed in [129–131] by using a combination of RMT and central limit theorem.

the iterative schemes presented so-far have either been analyzed with the assumption of perfect CSI at the receiver, or studied via numerical Monte Carlo simulations. The effects of imperfect CSI on DS-CDMA and MC-CDMA systems with non-iterative MUD were considered by using RMT and the replica method in [59] and [90], respectively. Other analytical treatments on the same topic include multiple access [139] and MIMO [140] channels, where lower bounds for mutual information were found by assuming a pilot-aided LMMSE channel estimator. Slightly different approach was used to study the optimality (or the lack of it) of Gaussian code books and nearest-neighbor decoding with different levels of channel side information in single [141] and multi-antenna [142] channels. Albeit several studies have shown via numerical simulations that iterative channel estimation can reduce the training overhead and improve the reliability of the CSI significantly [143–152], the only effort to analyze the performance of such a receiver has been made to our knowledge by Li, Betz & Poor in [138]. Indeed, mathematical analysis of iterative systems with imperfect CSI is the main topic of the present dissertation.

## 1.3 Aim and Outline of the Thesis

Channel estimation is an integral part of practical wireless systems. So-far, however, it has received somewhat lesser amount of interest in the analysis of multiuser CDMA channels. The purpose of the present thesis is to address this issue with a methodology that is general enough to be extended in future for further cases of interest as well. The main topics covered in the dissertation are the asymptotic replica analysis of:

1. Multi-antenna DS-CDMA systems in spatially correlated channels using linear channel estimation and multiuser detection (Chapter 3);
2. Single-antenna DS-CDMA systems operating in multipath fading channels and employing iterative channel estimation, detection, and decoding (Chapter 4).

As it turns out, channel estimation can indeed have a highly non-trivial impact on the system performance. The rest of this monograph is organized as follows.

- Chapter 2 introduces the notation used in the rest of the thesis and describes the system model for both of the aforementioned cases. The key assumptions made in the analysis are presented. Some background information on the mathematical methods employed later is given.
- Chapter 3 considers spatially correlated MIMO DS-CDMA systems. Different channel estimation and data detection algorithms are derived as specific

instances of a general Bayesian inference problem. In addition to optimum non-iterative pilot-assisted LMMSE channel estimator, two covariance mismatched linear MMSE estimators and an ML channel estimator are introduced. The MUDs considered in this chapter include the non-linear MAP detector, the linear MMSE and decorrelating detectors as well as the SUMF. The performance of the system is analyzed with the help of the decoupling principle, obtained via an application of the replica method. This chapter extends the previous results on multi-antenna CDMA systems [50, 92, 153–155] to the case of spatially correlated MIMO channels with CSI mismatch at the receiver.

- Chapter 4 derives a class of iterative channel estimators and MUDDs whose performance is analyzed later in the thesis. Specific examples arising from the general approach are presented, covering all the usual iterative MUDDs, as well as, iterative LMMSE and ML based channel estimators. The decoupling results for the iterative receiver are presented and performance analysis carried out. This chapter provides to our knowledge the first proposal for a systematic way of analyzing iterative receivers and contains as special cases the results reported in [137] and [138]. Indeed, the analysis covered in [138] were approximate whereas exact large system results are provided in this chapter.
- Chapter 5 provides the conclusions and discussion on the obtained results. Some future topics for further research are sketched. Most of the proofs are relegated to Appendices A – F.

# Chapter 2

# Preliminaries

This chapter provides necessary background information for the following analysis. We start by introducing the notation used for the rest of the thesis in Section 2.1. The discrete time signal models for the systems studied in the later chapters are given in Section 2.2. Some notes on the employed coding methods follow in Section 2.3, and a novel transmission scheme based on probability biased signaling is introduced in Section 2.4. A brief review of density evolution is carried out in Section 2.5, and discussion about the connection between the statistical physics and information processing can be found in Section 2.6.

## 2.1 Notation

Calligraphic symbols denote for sets and boldface symbols for (column) vectors and matrices. The transpose, complex conjugate and complex conjugate transpose of a matrix $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ are $\boldsymbol{A}^{\mathsf{T}}$, $\boldsymbol{A}^{*}$ and $\boldsymbol{A}^{\mathsf{H}}$, respectively. For matrix $\boldsymbol{A} = [\boldsymbol{a}_1\ \boldsymbol{a}_2\ \cdots\ \boldsymbol{a}_N] \in \mathbb{C}^{M\times N}$, we define $\underline{\boldsymbol{A}} = \mathrm{vec}(\boldsymbol{A}) = [\boldsymbol{a}_1^{\mathsf{T}}\ \boldsymbol{a}_2^{\mathsf{T}}\ \cdots\ \boldsymbol{a}_N^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{C}^{MN}$. Given a vector $\boldsymbol{a} \in \mathbb{C}^{M}$, and a sequence of matrices $(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_M)$, $\boldsymbol{A}_i \in \mathbb{C}^{M_i\times M_i}$, we let $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{a}) \in \mathbb{C}^{M\times M}$ be a diagonal matrix defined by the vector $\boldsymbol{a}$, and $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_M) \in \mathbb{C}^{\sum_i M_i\times\sum_i M_i}$ a block diagonal matrix formed from $(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_M)$. Operator $\otimes$ is the Kronecker product and for positive definite matrix $\boldsymbol{A}$ we write in shorthand $\boldsymbol{A} > 0$. We also denote $\vec{\boldsymbol{x}}$ for a $1 \times N$ row vector and $\boldsymbol{e}_M = [1\ 1\ \cdots\ 1]^{\mathsf{T}} \in \mathbb{R}^M$ for the vector of $M$ ones. Operators $\Re\{\cdot\}$ and $\Im\{\cdot\}$ return the real and imaginary part of the argument, respectively.

Throughout the thesis, we write $\boldsymbol{x} \sim \mathbb{P}$ and $\tilde{\boldsymbol{x}} \sim \mathbb{Q}$ for a random vector (RV) drawn according to the true $\mathbb{P}$ and postulated $\mathbb{Q}$ probability distribution, respectively. The postulated RVs are denoted by the same symbol as the true one with the

tilde on top. One can think of the postulated distributions as the receiver's, possibly mismatched, information about the random variables in the system. When the distribution of the random variable $\boldsymbol{x}$ depends on the iteration index $\ell = 1, 2, \ldots$ (see Chapter 4), we write $\mathbb{P}^{(\ell)}(\boldsymbol{x})$ and omit the index $\ell$ otherwise. The mean and covariance of $\mathbb{P}^{(\ell)}(\boldsymbol{x})$ are $\boldsymbol{\mu}_{\boldsymbol{x}}^{(\ell)}$ and $\boldsymbol{\Omega}_{\boldsymbol{x}}^{(\ell)}$, and the corresponding mean and covariance of the postulated RV $\tilde{\boldsymbol{x}}$ read $\tilde{\boldsymbol{\mu}}_{\boldsymbol{x}}^{(\ell)}$ and $\tilde{\boldsymbol{\Omega}}_{\boldsymbol{x}}^{(\ell)}$. The posterior mean estimate of a RV $\boldsymbol{x}$, and the related error covariance matrix, are denoted by $\langle \tilde{\boldsymbol{x}} \rangle_{(\ell)}$ and $\boldsymbol{\Omega}_{\Delta \boldsymbol{x}}^{(\ell)}$, respectively, unless stated otherwise. We also use $\mathbb{P}$ and $\mathbb{Q}$ to denote true and postulated probabilities (in case of discrete RVs) and densities (for absolutely continuous RVs) when applicable. The suitable interpretation should be clear from the context.

If $\boldsymbol{x}$ is a proper complex Gaussian random vector [156] with mean $\boldsymbol{\mu}_{\boldsymbol{x}} = \mathbb{E}\{\boldsymbol{x}\}$ and covariance matrix $\boldsymbol{\Omega}_{\boldsymbol{x}} = \mathbb{E}\{(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}})(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}})^{\mathsf{H}}\}$, we write in shorthand $\boldsymbol{x} \sim \mathsf{CN}(\boldsymbol{\mu}_{\boldsymbol{x}}; \boldsymbol{\Omega}_{\boldsymbol{x}})$ or $\mathbb{P}(\boldsymbol{x}) = \mathsf{CN}(\boldsymbol{\mu}_{\boldsymbol{x}}; \boldsymbol{\Omega}_{\boldsymbol{x}})$. $\Pr(\cdot)$ denotes for the probability of the argument. The Dirac measure concentrated at $\boldsymbol{x} \in \mathbb{C}^M$ is defined as $\delta_{\boldsymbol{x}}(\mathcal{A}) = 1$ when $\boldsymbol{x} \in \mathcal{A}$ and $\delta_{\boldsymbol{x}}(\mathcal{A}) = 0$ otherwise, satisfying $\int f(\boldsymbol{y}) \delta_{\boldsymbol{x}}(\mathrm{d}\boldsymbol{y}) = f(\boldsymbol{x})$ for any continuous[1] $f : \mathbb{C}^M \to \mathbb{R}$. The indicator function is defined as $\delta_{\boldsymbol{x}}(\mathcal{A}) = \mathbb{1}_{\mathcal{A}}(x)$. All integrals should be considered as Lebesque-Stieltjes integrals over the entire support of the kernel unless stated otherwise.

## 2.2 System Model

In this section, the discrete time signal models for the systems considered in the present dissertation are outlined. We start by considering a MIMO DS-CDMA system operating over a flat fading channel in Section 2.2.1. Single-antenna DS-CDMA system in a multipath fading channel is introduced in Section 2.2.2. The assumptions made in the channel models and the connections between the two systems are briefly discussed. Before proceeding to the details of the system models, two remarks are made:

1. We use the same notation for the variables of both systems. This should cause no confusion since the analysis is carried out in a separate chapter for both of them.
2. Throughout the dissertation the transmitter is assumed to have no information about the channel conditions.

[1] Note that formally $f$ should have a compact support. We can always make it so by letting the range to be the set of extended real numbers while treating the axes of the complex plane $\mathbb{C}$ as extended real lines.

### 2.2.1 MIMO DS-CDMA in Flat Fading Channel

Let us first consider a synchronous uplink MIMO DS-CDMA system operating over a narrowband block fading channel [157, 158] with a fixed coherence time of $T_{\mathsf{coh}}$ symbols. For simplicity, let the mobile terminals of all users $k = 1, \ldots, K$ have $M$ antennas while the receiver is equipped with $N$ antennas. By the assumption of narrowband transmission, the channel is modeled by a single fading tap and hence there is no intersymbol interference (ISI) in the system.

Consider the fading block $c = 1, 2, \ldots, C$ and time instant $t = 1, \ldots, T_{\mathsf{coh}}$. The discrete time received vector after matched filtering and sampling is given for the chip index $l = 1, \ldots, L$

$$\boldsymbol{y}_{l,t}[c] = \begin{cases} \dfrac{1}{\sqrt{L}} \displaystyle\sum_{k=1}^{K} \boldsymbol{H}_k[c]\boldsymbol{P}_{k,t}[c]\boldsymbol{s}_{k,l,t} + \boldsymbol{w}_{l,t}[c] \in \mathbb{C}^N, & t \in \mathcal{T}, \\ \dfrac{1}{\sqrt{L}} \displaystyle\sum_{k=1}^{K} \boldsymbol{H}_k[c]\boldsymbol{X}_{k,t}[c]\boldsymbol{s}_{k,l,t} + \boldsymbol{w}_{l,t}[c] \in \mathbb{C}^N, & t \in \mathcal{D}, \end{cases} \tag{2.1}$$

where

$$\mathcal{T} = \{1, \ldots, \tau_{\mathsf{tr}}\} \quad \text{and} \quad \mathcal{D} = \{\tau_{\mathsf{tr}} + 1, \ldots, T_{\mathsf{coh}}\}, \tag{2.2}$$

contain the time indices related to the training and data transmission phases, respectively. The corresponding diagonal matrices

$$\boldsymbol{P}_{k,t}[c] = \mathrm{diag}(\boldsymbol{p}_{k,t}[c]), \qquad t \in \mathcal{T}, \tag{2.3}$$

$$\boldsymbol{X}_{k,t}[c] = \mathrm{diag}(\boldsymbol{x}_{k,t}[c]), \qquad t \in \mathcal{D}, \tag{2.4}$$

contain the pilot $\boldsymbol{p}_{k,t}[c] = \left[p_{k,t,1}[c] \ \cdots \ p_{k,t,M}[c]\right]^{\mathsf{T}}$ and information bearing vectors $\boldsymbol{x}_{k,t}[c] = \left[x_{k,t,1}[c] \ \cdots \ x_{k,t,M}[c]\right]^{\mathsf{T}}$ sent by the user $k \in \mathcal{K} = \{1, \ldots, K\}$ during $\mathcal{T}$ and $\mathcal{D}$, respectively. For future reference, the number of data vectors transmitted by each user during one fading block is denoted by $\tau_{\mathsf{d}} = |\mathcal{D}|$. Note that since we consider spatial multiplexing at the transmitter, the elements of $\boldsymbol{p}_{k,t}[c]$ and $\boldsymbol{x}_{k,t}[c]$ are assumed to be independent in the following.

The set containing the (vectorized) MIMO channels of all users during the $c$th fading block is given by

$$\mathcal{H}_c = \left\{ \underline{\boldsymbol{H}}_k[c] = \mathrm{vec}(\boldsymbol{H}_k[c]) = \left[\boldsymbol{h}_{k,1}^{\mathsf{T}}[c] \ \cdots \ \boldsymbol{h}_{k,M}^{\mathsf{T}}[c]\right]^{\mathsf{T}} \in \mathbb{C}^{NM} \mid \forall k \in \mathcal{K} \right\}, \tag{2.5}$$

where $\boldsymbol{h}_{k,m}[c] \in \mathbb{C}^N$ is the channel vector between the $k$th user's $m$th transmit antenna and the $N$ receiving antennas. The spreading sequence at time instant $t =$

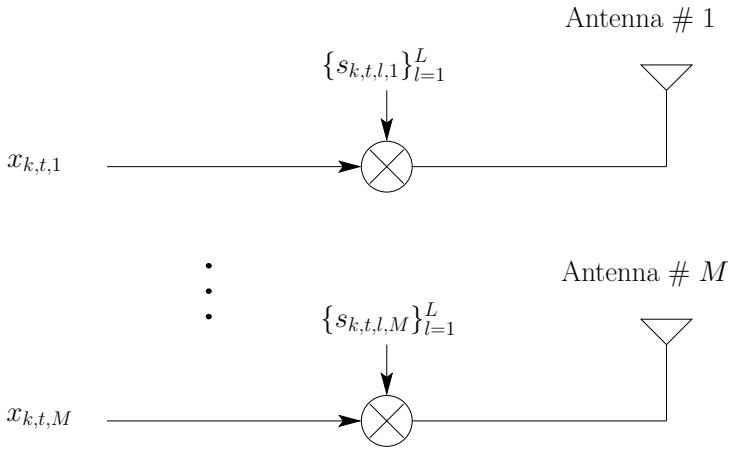



**Figure 2.1.** Per-antenna spreading scheme with spatial multiplexing.

$1, \ldots, T_{\mathsf{coh}}$ for the user $k \in \mathcal{K}$ and chip index $l = 1, \ldots, L$ is written as $\boldsymbol{s}_{k,l,t} \in \mathbb{C}^M$, and $\boldsymbol{w}_{l,t}[c] \sim \mathsf{CN}(\mathbf{0}; \sigma^2 \boldsymbol{I}_N)$ represents the additive white Gaussian noise at the receiver.

For the following development, let us write

$$\mathcal{P} = \{\boldsymbol{p}_{k,t}[c] \in \mathcal{M}^M \mid \forall k \in \mathcal{K}, t \in \mathcal{T}\}, \tag{2.6}$$

$$\mathcal{X}_t = \{\boldsymbol{x}_{k,t}[c] \in \mathcal{M}^M \mid \forall k \in \mathcal{K}\}, \qquad t \in \mathcal{D}, \tag{2.7}$$

for the set of all training symbols (known at the receiver) and for the set of all data symbols transmitted during the $t$th time slot, respectively. We also assume that the RVs in $\mathcal{P}$ and $\{\mathcal{X}_t \mid \forall t \in \mathcal{D}\}$ are independent identically distributed (IID) with their elements uniformly drawn from the quarternary phase shift keying (QPSK) signal set

$$\mathcal{M} = \left\{ \pm \frac{1}{\sqrt{2}} \pm \frac{\mathrm{j}}{\sqrt{2}} \right\}. \tag{2.8}$$

For notational convenience, we write $\mathcal{Y}_t = \{\boldsymbol{y}_{l,t} \,|\, l = 1, \cdots, L\}$ for signals received during the $t$th time slot and $\mathcal{Y}_{\mathcal{T}} = \{\mathcal{Y}_t \mid t \in \mathcal{T}\}$ for the set of vectors received when the pilot symbols $\mathcal{P}$ were transmitted.

For the CDMA we consider a random per-antenna spreading (PAS) scheme that assigns all users and transmit antennas a unique signature sequence [92, 153–155, 159]. See Figure 2.1 for an illustration. For the system model (2.1), this translates to the assumption that the RVs in

$$\mathcal{S} = \{\boldsymbol{s}_{k,l,t} = [s_{k,t,l,1} \ \cdots \ s_{k,t,l,M}]^{\mathsf{T}} \mid \forall k, l, t\} \tag{2.9}$$

are perfectly known at the receiver and have IID elements with zero mean and unit variance. This is in contrast to per-user spreading (PUS) [92, 153], where the spreading codes of the users are independent, but same signature sequence is used for all the antennas of a given user. The motivation for concentrating on the PAS scheme here stems from the decoupling results derived in [92] (see also [160]). The analysis revealed that in the large system limit, when the receiver has perfect CSI and the SNR relatively high, the PAS scheme is able to provide a full multiplexing gain regardless of spatial correlation, whereas PUS was limited by the degrees of freedom in the MIMO channel. Therefore, one expects that per-antenna spreading provides higher spectral efficiencies in correlated environments. The actual comparison of the two spreading methods under the assumption of channel estimation is, however, left as future work.

By the assumption of block fading channel, we let the RVs in $\{\underline{\boldsymbol{H}}_k[c] \mid \forall c\}$ be IID for all users $k = 1, \ldots, K$. Furthermore, the channels between different users are assumed to be independent and the RVs $\underline{\boldsymbol{H}}_k[c], c = 1, \ldots, C$, are drawn according to the proper complex Gaussian distribution $\mathbb{P}(\underline{\boldsymbol{H}}_k[c]) = \mathsf{CN}(\mathbf{0}; \boldsymbol{\Omega}_{\underline{\boldsymbol{H}}_k})$, where the spatial correlation is given by the "Kronecker" model[2] $\boldsymbol{\Omega}_{\underline{\boldsymbol{H}}_k} = \boldsymbol{T}_k \otimes \boldsymbol{R}$. Here, $\boldsymbol{T}_k \in \mathbb{C}^{M \times M}$ and $\boldsymbol{R} \in \mathbb{C}^{N \times N}$ are Hermitian positive definite and represent the decoupled transmitter and receiver side covariance matrices, latter of which has been normalized to have diagonal entries of unity. We also assume that $\{\boldsymbol{T}_k\}_{k=1}^K$ are IID and drawn according to a well defined discrete distribution $p_{\mathsf{tx}}$. The average SNR for user $k$ is defined as $\overline{\mathsf{snr}}_k = \bar{t}_k/\sigma^2$, where $\bar{t}_k = \mathrm{tr}(\boldsymbol{T}_k)$.

### 2.2.2 DS-CDMA in Multipath Fading Channel

Consider next a synchronous uplink DS-CDMA system, operating over a block fading multipath channel with a coherence time of $T_{\mathsf{coh}}$ symbols. For the following discussion we make the simplifying assumption that the ISI induced by the multipath fading has negligible effect on the system performance. We therefore omit the equalization from the analysis and assume that the received signal is not corrupted by ISI. This corresponds to a scenario where the delay spread of the channel is small compared to the symbol period or a block transmission with sufficiently long cyclic prefix is used. Another way of looking at the following results is to consider them

[2] Note that this correlation model coincides with the assumption that the channel matrix is drawn according to $\mathbb{P}(\boldsymbol{H}_k[c]) = \mathsf{CN}(\boldsymbol{M} = \mathbf{0}; \boldsymbol{\Omega}_k = \boldsymbol{T}_k, \boldsymbol{\Sigma} = \boldsymbol{R})$, where $\mathsf{CN}(\boldsymbol{M}; \boldsymbol{\Omega}, \boldsymbol{\Sigma})$ is the complex matrix Gaussian distribution (see, e.g., [161]). For theoretical discussion on this correlation model, see e.g., [66, 162]. Some practical considerations and model verification via channel measurements can be found, for example, in [163, 164].

as being an upper bound on the performance of a practical system that may suffer from ISI.

The discrete time model after matched filtering and chip-rate sampling for the $t$th received vector $\boldsymbol{y}_t[c] \in \mathbb{C}^L$ within the fading block $c = 1, \ldots, C$ can be written as

$$\boldsymbol{y}_t[c] = \begin{cases} \dfrac{1}{\sqrt{L}} \displaystyle\sum_{k=1}^{K} \boldsymbol{S}_{k,t} \boldsymbol{h}_k[c] p_{k,t}[c] + \boldsymbol{w}_t[c] \in \mathbb{C}^L, & t \in \mathcal{T}, \\ \dfrac{1}{\sqrt{L}} \displaystyle\sum_{k=1}^{K} \boldsymbol{S}_{k,t} \boldsymbol{h}_k[c] x_{k,t}[c] + \boldsymbol{w}_t[c] \in \mathbb{C}^L, & t \in \mathcal{D}, \end{cases} \tag{2.10}$$

where $\mathcal{T}$ and $\mathcal{D}$ are defined in (2.2) and collect the time indices when the $\tau_{\mathsf{tr}}$ training $\{p_{k,t}[c]\}_{t \in \mathcal{T}}$ and $\tau_{\mathsf{d}} = |\mathcal{D}|$ information symbols $\{x_{k,t}[c]\}_{t \in \mathcal{D}}$ of the user $k = 1, \ldots, K$ are transmitted. The spreading matrix for the $k$th user at time index $t$ is given by $\boldsymbol{S}_{k,t} \in \mathbb{C}^{L \times M}$. The set of all received vectors during the $c$th fading block is written as $\mathcal{Y}_c = \{\boldsymbol{y}_t[c] \mid t = 1, \cdots, T_{\mathsf{coh}}\}$, and similarly $\mathcal{H}_c = \{\boldsymbol{h}_k[c] = \begin{bmatrix} h_{k,1}[c] & \cdots & h_{k,M}[c] \end{bmatrix}^{\mathsf{T}} \in \mathbb{C}^M \mid \forall k\}$ denotes the fading coefficients of all users in the $c$th fading block. For notational simplicity, we let the number of multipaths $M$ and the spreading factor $L$ be the same for all users. The samples $\{\boldsymbol{w}_t[c] \mid \forall t, c\}$ of thermal noise at the receiver are assumed to be IID and drawn according to the complex Gaussian distribution $\mathbb{P}(\boldsymbol{w}_t[c]) = \mathsf{CN}(\mathbf{0}; \sigma^2 \boldsymbol{I}_L)$.

Let us now consider the spreading matrices $\boldsymbol{S}_{k,t} = \begin{bmatrix} \boldsymbol{s}_{k,t,1} & \cdots & \boldsymbol{s}_{k,t,M} \end{bmatrix} \in \mathbb{C}^{L \times M}$, $t = 1, \ldots, T_{\mathsf{coh}}$, where $\boldsymbol{s}_{k,t,m}$ is the spreading sequence corresponding to the $m$th resolvable multipath. As with the case of MIMO DS-CDMA in Section 2.2.1, we assume that due to random spreading $\mathcal{S} = \{\boldsymbol{S}_{k,t} \mid \forall k, t\}$ are IID random matrices. For a fixed time index $t$, however, the spreading sequences $\{\boldsymbol{s}_{k,t,m}\}_{m=1}^{M}$ of the $k$th user are not IID random vectors. In fact, the spreading sequences for each multipath are cyclically shifted replicas of each other. For the following analysis we make the crucial assumption that [59, Theorem 4] holds for our system.

**Assumption 1.** Without loss of generality, the spreading sequences $\{\boldsymbol{s}_{k,t,m}\}_{m=1}^{M}$ can be modified to have IID entries with zero mean, unit variance and finite moments for all $t = 1, \ldots, T_{\mathsf{coh}}$. ♢

Given the Assumption 1 and under the condition that we can neglect the effects of ISI in the analysis, comparing (2.1) and (2.10) reveals that the DS-CDMA system operating over an $M$-path fading channel is equivalent to a MIMO DS-CDMA system with $M$ transmit antennas and one receiving antenna in a flat fading channel, given each transmit antenna has the same data. Thus, we could use (2.1) to represent the MIMO system described in Section 2.2.1, or a single input multiple output

(SIMO) DS-CDMA system in a multipath fading channel by taking into account the distribution of the elements of the transmit vectors. In the following, however, when discussing the multipath fading channels, we limit our scope to the case of single transmit and receive antenna and use (2.10) to describe the system.

For the statistical channel model, we consider the important special case of frequency selective Rayleigh fading. The users are assumed to be well separated in space and the environment rich scattering, so that the fading channels between the users are independent and the multipaths uncorrelated [158]. The channel vectors in $\mathcal{H}_c$ are thus independent with distribution $\mathbb{P}(\boldsymbol{h}_k[c]) = \mathsf{CN}(\mathbf{0}; \boldsymbol{\Omega}_{\boldsymbol{h}_k[c]})$, where $\boldsymbol{\Omega}_{\boldsymbol{h}_k[c]} = \operatorname{diag}(\bar{\boldsymbol{t}}_k)$ and $\bar{\boldsymbol{t}}_k = \begin{bmatrix}\bar{t}_{k,1} & \cdots & \bar{t}_{k,M}\end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^M$ is the power delay profile (PDP) of the $k$th user's channel. For simplicity we let $\{\bar{\boldsymbol{t}}_k\}_{k=1}^K$ be IID and drawn according to a discrete distribution $p_{\mathsf{pdp}}$. The average received signal-to-noise ratio for the user $k$ is defined as in Section 2.2.1 and, thus, $\overline{\mathsf{snr}}_k = \bar{t}_k/\sigma^2$, where $\bar{t}_k = \operatorname{tr}(\boldsymbol{\Omega}_{\boldsymbol{h}_k[c]})$.

## 2.3 Channel Coding

We next take a brief look at the two different coding strategies encountered later in the thesis. Section 2.3.1 discusses capacity achieving signaling under Gaussian and QPSK constrained channel inputs. Section 2.3.2 follows by introducing a simplified coding scheme called bit-interleaved coded modulation (BICM) [165–168] (see also [169]).

### 2.3.1 Capacity Achieving Codes

Consider a single-user system operating over an ergodic Rayleigh fading SIMO channel[3]

$$\boldsymbol{y}_t = \boldsymbol{h}_t x_t + \boldsymbol{w}_t \in \mathbb{C}^N, \tag{2.11}$$

where $\boldsymbol{w}_t \sim \mathsf{CN}(\mathbf{0}; \boldsymbol{D})$ and $\boldsymbol{h}_t \sim \mathsf{CN}(\mathbf{0}; \boldsymbol{\Omega}_{\boldsymbol{h}})$ are independent RVs for all $t = 1, 2, \ldots, T$. Let the channel coefficients $\{\boldsymbol{h}_t\}_{t=1}^T$ be perfectly known at the receiver. Assume that the messages of the user have equal probability and they are mapped before transmission to the code words $\boldsymbol{x} = \begin{bmatrix}x_1 & \cdots & x_T\end{bmatrix}^{\mathsf{T}}$ of a standard random Gaussian code book with rate $R$ [11, 170]. If ML decoding is used at the receiver, all rates (in bits) below

$$\mathsf{C}_{\mathsf{simo}} = \mathbb{E}_{\boldsymbol{h}}\{\log_2(1 + \boldsymbol{h}^{\mathsf{H}}\boldsymbol{D}^{-1}\boldsymbol{h})\}, \qquad \boldsymbol{h} \sim \mathsf{CN}(\mathbf{0}; \boldsymbol{\Omega}_{\boldsymbol{h}}), \tag{2.12}$$

[3]The reason for concentrating on the SIMO channel will become clear later when the multiuser systems are shown to decouple into sets of single-user SIMO channels with colored Gaussian noise.

are achievable with vanishing probability of error as $T$ grows without bound [11, 170]. Conversely, for all $R > \mathsf{C}_{\mathsf{simo}}$ the error probability is bounded away from zero.

The capacity (2.12) of the SIMO channel (2.11) is achieved with Gaussian channel inputs. A natural question to ask from practical point-of-view might be — what is the highest achievable rate of the same channel if we constrain the channel inputs, say to the QPSK constellation (2.8), i.e., $x_t \in \mathcal{M}, t = 1, \ldots, T$. If we are allowed to optimize the coding and modulation mapping jointly, the QPSK constrained capacity, or coded modulation capacity with QPSK mapping [165, 166], is given by

$$\mathsf{C}_{\mathsf{simo}}^{\mathsf{qpsk}} = \log_2 |\mathcal{M}| - N \log_2 e \\ - \frac{1}{|\mathcal{M}|} \sum_{x \in \mathcal{M}} \mathbb{E}_{\boldsymbol{w},\boldsymbol{h}} \Big\{ \log_2 \sum_{\tilde{x} \in \mathcal{M}} \exp\big( - [\boldsymbol{w} + \boldsymbol{h}(x - \tilde{x})]^{\mathsf{H}} \boldsymbol{D}^{-1} [\boldsymbol{w} + \boldsymbol{h}(x - \tilde{x})] \big) \Big\}, \tag{2.13}$$

where $\boldsymbol{w} \sim \mathsf{CN}(\mathbf{0};\, \boldsymbol{D})$ and $\boldsymbol{h}$ is as in (2.12) [171].

**Example 1.** Let $N = 4$ and $\boldsymbol{w} \sim \mathsf{CN}(\mathbf{0};\, \sigma^2 \boldsymbol{I}_4)$. The capacity of the channel (2.11) with Gaussian (2.12) and QPSK signaling (2.13) is plotted in Figure 2.2 for uncorrelated and fully correlated receive antennas. As expected, the QPSK constrained capacity saturates to 2 bits per channel use, whereas the maximum achievable rate with Gaussian signaling keeps on growing with increasing SNR. ♢

### 2.3.2 Bit-Interleaved Coded Modulation

Let us denote the information bits of the user $k = 1, \ldots, K$ by $\boldsymbol{b}_k \in \{0, 1\}^B$, where the elements of $\boldsymbol{b}_k$ are IID and uniformly drawn from the binary alphabet. Encoding the data of the $k$th user with BICM consists of first applying a binary error correction code to the information bits $\boldsymbol{b}_k$, shuffling the coded bits by using a random uniform bit-interleaver and finally employing a memoryless symbol-by-symbol modulation mapping to form the channel inputs [165–168]. A practical benefit of BICM compared the case of coded modulation, which was the capacity achieving scheme discussed in the previous section, is that with BICM one can concentrate on the task of finding efficient binary ECCs independently of the modulation mapping. Separating the ECC and modulation causes a loss in the achievable capacity but the degradation is very minor if Gray mapping is used, especially for lower order constellations [165–167].

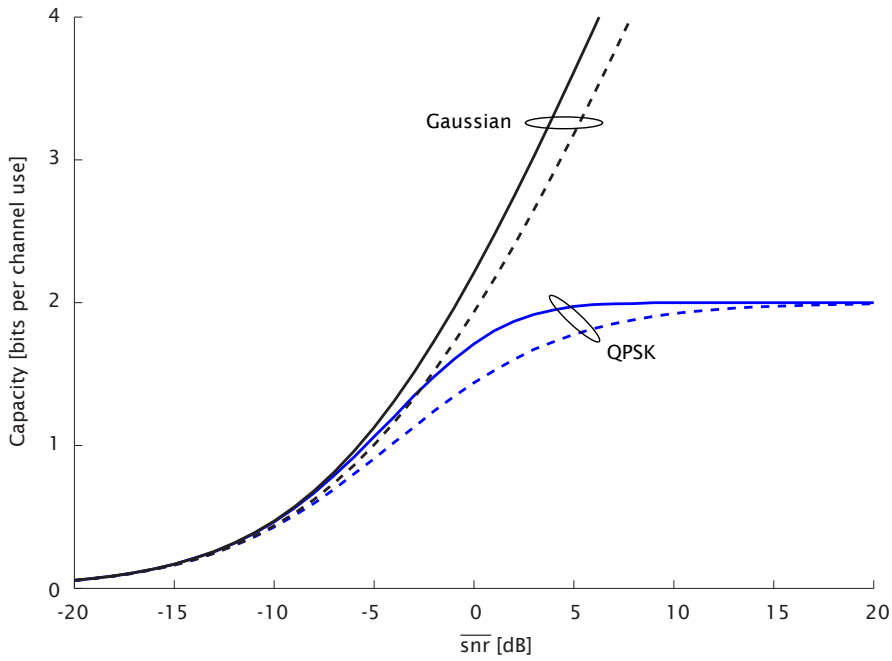



**Figure 2.2.** Capacity (in bits per channel use) of a SIMO system with $N = 4$ receive antennas. Solid lines for uncorrelated antennas and dashed lines for fully correlated ones.

Consider encoding the data $\boldsymbol{b}_k$ of the $k$th user with BICM when the modulation is constrained to the standard QPSK signal set $\mathcal{M}$ given in (2.8). We can write this operation formally as

$$\phi_k : \{0,1\}^B \to \mathcal{M}^T : \boldsymbol{b}_k \mapsto \boldsymbol{x}_k, \tag{2.14}$$

where

$$\boldsymbol{x}_k = \mathrm{vec}\big([\boldsymbol{x}_k[1] \ \cdots \ \boldsymbol{x}_k[C]]\big) \in \mathcal{M}^T, \tag{2.15}$$

with

$$\boldsymbol{x}_k[c] = \big[x_{k,\tau_{\mathsf{tr}}+1}[c] \ \cdots \ x_{k,T_{\mathsf{coh}}}[c]\big]^{\mathsf{T}} \in \mathcal{M}^{\tau_{\mathsf{d}}}, \qquad c = 1, \ldots, C, \tag{2.16}$$

is the code word containing the $T = \tau_{\mathsf{d}} C$ channel coded information symbols of the user $k = 1, \ldots, K$. The rate $R = B/T$ BICM code book of the $k$th user is written as

$$\mathcal{C}_k = \big\{\boldsymbol{x}_k = \phi_k(\boldsymbol{b}_k) \mid \forall \boldsymbol{b} \in \{0,1\}^B\big\}. \tag{2.17}$$

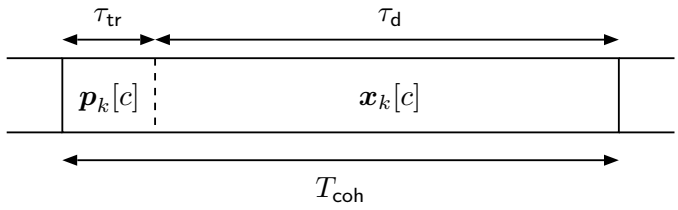



**Figure 2.3.** Frame structure of the considered system.

In the later sections we occasionally use the notation

$$x_{k,t} = \frac{1}{\sqrt{2}}(a_{k,t,1} + \mathrm{j}a_{k,t,2}), \qquad a_{k,t,1}, a_{k,t,2} \in \{\pm 1\}, \tag{2.18}$$

where $a_{k,t,1}$ and $a_{k,t,2}$ are the scaled real and imaginary parts of $x_{k,t}$, respectively. Unless stated otherwise (see next section), we always consider binary linear ECCs so that due to random bit-interleaving and Gray mapping the BICM decouples the real and imaginary parts of the code symbols $\mathcal{X}_c = \{x_{k,t}[c] \mid \forall k, c, t \in \mathcal{D}\}$ in the limit $T = \tau_{\mathsf{d}} C \to \infty$ with $\tau_{\mathsf{d}}$ fixed, that is,

$$\mathbb{P}\left(x_{k,t} = \frac{1}{\sqrt{2}}(a_{k,t,1} + \mathrm{j}a_{k,t,2})\right) = \mathbb{P}(a_{k,t,1})\mathbb{P}(a_{k,t,2}), \tag{2.19}$$

where

$$\mathbb{P}(a_{k,t,q}) = \frac{1}{2}\delta_{a_{k,t,q}}(-1) + \frac{1}{2}\delta_{a_{k,t,q}}(+1), \qquad q = 1, 2. \tag{2.20}$$

**Assumption 2.** In this thesis, the data bits are encoded by using binary trellis codes with trellis termination. All users are assumed to derive the ECC from the same ensemble of binary codes, while the random bit-interleavers are IID for all $k = 1, \ldots, K$. Gray modulation mapping is always employed. $\diamondsuit$

## 2.4 Training via Biased Signaling

In the previous sections we assumed that in addition to the information carrying data symbols, each fading block $c = 1, \ldots, C$ contains also $\tau_{\mathsf{tr}}$ training symbols for all users. These symbols are always perfectly known at the receiver and can be used to perform the initial channel estimation. The frame structure of this transmission scheme is illustrated in Fig. 2.3. This is, however, not the only option and we shall describe in the following a scheme based on probability biased signaling that can be used to initiate the channel estimation.

Let $\Theta = \{\theta_{k,t}[c] \in \mathbb{C} \mid \forall k, t, c\}$ be a set of design variables known to both the transmitter and the receiver. Define the conditional prior distribution of $x_{k,t}[c]$ as

$$\begin{aligned}\mathbb{P}(x_{k,t}[c] \mid \theta_{k,t}[c]) = \mathbb{P}'(\Re\{x_{k,t}[c]\} \mid \Re\{\theta_{k,t}[c]\}) \\ \times \mathbb{P}'(\Im\{x_{k,t}[c]\} \mid \Im\{\theta_{k,t}[c]\}),\end{aligned} \tag{2.21}$$

where

$$\mathbb{P}'(x \mid \theta) = \frac{1+\sqrt{2}\theta}{2}\delta_x(1/\sqrt{2}) + \frac{1-\sqrt{2}\theta}{2}\delta_x(-1/\sqrt{2}). \tag{2.22}$$

Similarly, we let $\mathbb{P}(\theta_{k,t}[c]) = \mathbb{P}'(\Re\{\theta_{k,t}[c]\})\mathbb{P}'(\Im\{\theta_{k,t}[c]\})$ be the prior of $\theta_{k,t}[c]$ with

$$\begin{aligned}\mathbb{P}'(\theta) = \frac{\Delta_{\mathsf{tr}}}{2}\delta_\theta(1/\sqrt{2}) + \frac{\Delta_{\mathsf{tr}}}{2}\delta_\theta(-1/\sqrt{2}) \\ + \frac{1-\Delta_{\mathsf{tr}}}{2}\delta_\theta(\sigma_{\mathsf{bias}}) + \frac{1-\Delta_{\mathsf{tr}}}{2}\delta_\theta(-\sigma_{\mathsf{bias}}),\end{aligned} \tag{2.23}$$

where $\sigma_{\mathsf{bias}} \in [0, 1/\sqrt{2})$ and $\Delta_{\mathsf{tr}} \in [0, 1)$ are fixed design parameters for all $k, t$ and $c$. Thus, $\mathbb{E}\{x_{k,t}[c] \mid \theta_{k,t}[c]\} = \theta_{k,t}[c]$, where $\Re\{\theta_{k,t}[c]\}, \Im\{\theta_{k,t}[c]\} \in [-1/\sqrt{2}, 1/\sqrt{2}]$. Since $\Theta$ is assumed to be known at the receiver, setting $\sigma_{\mathsf{bias}} = 0$ gives the traditional pilot assisted transmission scheme. For large $T_{\mathsf{coh}}$, we may assume without loss of generality that each fading block has then $\tau_{\mathsf{tr}} = \Delta_{\mathsf{tr}} T_{\mathsf{coh}}$ modulated "hard" pilot symbols, denoted as before by $\boldsymbol{p}_k[c] \in \mathcal{M}^{\tau_{\mathsf{tr}}}$, and the number of data symbols $\tau_{\mathsf{d}} = T_{\mathsf{coh}} - \tau_{\mathsf{tr}}$ is fixed for all fading blocks. If, on the other hand, we set $\Delta_{\mathsf{tr}} = 0$, the optimum hyperprior for unconstrained receiver is retrieved (see [172, Prop. 2]).

The total training overhead of the system as a fraction of the total transmission is given by

$$\Delta_{\mathsf{tot}} = \Delta_{\mathsf{tr}} + (1 - \Delta_{\mathsf{tr}})\Delta_{\mathsf{d}} \in [0, 1), \tag{2.24}$$

where we have assumed that

$$\Delta_{\mathsf{d}} = 1 - H\left(\frac{1-\sigma_{\mathsf{bias}}}{2}\right), \tag{2.25}$$

is the amount of pilot information embedded in the data symbols, and

$$H(p) = -p\log_2 p - (1-p)\log_2(1-p), \tag{2.26}$$

is the binary entropy function. This corresponds to an ideal method of signal biasing that incurs no additional overhead by itself. In the numerical examples we also assume that the bit error rate performance of the BICM is not affected by the a priori signal bias.

## 2.5 Density Evolution

In [125, 126], the information exchange in the iterative processing was analyzed with the help of density evolution [21, 23, 24, 132, 133]. The present dissertation follows the same basic approach in the analysis of the iterative MUDD and the iterative CE. Parallel scheduling with extrinsic information is used for multiuser detection and decoding as in [125, 126]. All the results with iterative processing are obtained in the limit of large code word length $T \to \infty$. As remarked in [125, pp. 1780], the order of limits should be taken so that we first let $T \to \infty$ and then take the large system limit $K = \alpha L \to \infty$, implying that for finite user loads $T \gg K$. This is true for any modern wireless system, where the code words are typically several thousand symbols long.

Simplifying the density evolution by treating the outputs of the sum-product decoder as Gaussian random variables with symmetric density is used extensively in the coding theory literature (see e.g., [133, 173, 174]). Although this is an approximation of the true output of a physical system, the error resulting from this simplification is typically small. We make thus make the following assumption for the rest of the thesis.

**Assumption 3.** For all iterations $\ell = 1, 2, \ldots$, the true posterior distribution of the symbol probabilities at the outputs of the sum-product decoders coincides with the ones obtained by using density evolution with Gaussian approximation. $\diamondsuit$

Rest of the details regarding density evolution with Gaussian approximation are postponed to Section 4.3.1.

## 2.6 Statistical Physics and the Replica Method

In this section, we give a very brief description of some of the main concepts in statistical physics. The main focus is on the special set of magnetic materials, spin glasses, whose mathematical models have been recently found to have connections to many problems encountered in engineering and information processing sciences. A mathematical framework proposed for the analysis of disordered spin glasses, the replica method, is briefly discussed in the context of infinite range Sherrington-Kirkpatrick (S-K) model of spin glasses [175] (see also [176] and [177]).

The following should not be taken as a general introduction to statistical mechanics, and the interested reader will find much deeper discussion on the connection between the statistical physics and information processing in the recent books

[14, 15] (see [13, 102–104] for a discussion on the validity of the replica method and the assumption of replica symmetry (RS) in statistical physics applications) and articles [73, 75, 85, 88, 89]. A very nice introduction to statistical physics and information processing in telecommunications can also be found in [178].

### 2.6.1 A Note on Statistical Physics

One of the main goals of statistical physics is to explain how the complex macroscopic (large scale) behavior of homogeneous physical systems arises from its simple microscopic (small scale) structure. The problem is, however, that an effort to give a meticulous description of the interactions between the particles (atoms, molecules, etc.) via (quantum) mechanics does not in general lead to tractable mathematical models. Indeed, one of the key ideas in statistical physics is to use a simplified probabilistic model for the particle interactions so that the resulting system can be analyzed mathematically. To illustrate the concept, we shall consider in the following a simple classical (as opposed to quantum) statistical mechanic system that has connections to the mathematical models found in some engineering disciplines as well.

Let the set $\{1, 2, \ldots, K\}$ denote the for the $K$ "sites" present in the system. Here the term "site" is a placeholder capable of accommodating an arbitrary abstract object used to characterize the microscopic particles of the physical system under consideration. Throughout this section we shall assign the microscopic state variables

$$\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \end{bmatrix}^{\mathsf{T}}, \qquad x_k \in \mathcal{X}, \tag{2.27}$$

to their respective sites and let $\mathcal{X}$ be a finite set consisting of all allowed per-site states. The energy function, or Hamiltonian[4], for a given configuration $\boldsymbol{x} \in \mathcal{X}^K$ is denoted by $E(\boldsymbol{x})$. As mentioned above, the key idea in statistical physics is that the equilibrium interactions between the elements of $\boldsymbol{x}$ in $E(\boldsymbol{x})$ can be described in probabilistic terms, and that this fully characterizes the macroscopic (deterministic) behavior of the system in the thermodynamic limit, i.e., $K \to \infty$.

In order to fix the nomenclature for the following discussion, let us now concentrate on the specific case of magnetic materials. Assume for simplicity that $\mathcal{X} = \{+1, -1\}$, i.e., the sites $\{1, 2, \ldots, K\}$ are associated with the binary microstate variables $x_k \in \{\pm 1\}, k = 1, \ldots, K$. In statistical physics, such $x_k$ are

[4] The Hamiltonian gives the microscopic energy (hence the other name energy function) of a given configuration $\boldsymbol{x} \in \mathcal{X}^K$. For our purposes, it is a real valued function that specifies the microscopic behavior of the physical system of interest entirely. Here we do not dwell on the topic of how to find suitable Hamiltonians for the physical system, but rather assume it has been predefined.

called Ising spins and they represent the spins, or magnetic moments, of the electrons. Let the Hamiltonian $E(\boldsymbol{x})$ be chosen such that it gives a (simplified) description of the microscopic interactions in the desired physical system. The starting point for the analysis of such a system in the thermodynamic equilibrium is the Boltzmann distribution, or the Gibbs measure,

$$\mathbb{P}(\boldsymbol{x}) = \frac{1}{Z}\mathrm{e}^{-\beta E(\boldsymbol{x})}, \qquad \boldsymbol{x} \in \{\pm 1\}^K, \tag{2.28}$$

where $\beta = 1/T > 0$ is the inverse temperature. The special property of (2.28) is that it maximizes the entropy[5]

$$H = -\sum_{\boldsymbol{x}} \mathbb{P}(\boldsymbol{x}) \log \mathbb{P}(\boldsymbol{x}), \tag{2.29}$$

given the average energy

$$\langle E \rangle = \sum_{\boldsymbol{x}} \mathbb{P}(\boldsymbol{x}) E(\boldsymbol{x}), \tag{2.30}$$

is kept constant. Note that this condition arises from the observed behavior of physical systems in nature. The notation used in (2.30) is commonly used for averages that are taken with respect to the Gibbs measure (2.28).

The physical interpretation of the Boltzmann distribution is that if we keep the system at some fixed macrostate (say, constant volume, pressure, etc.), and let it set to equilibrium with an infinite heat bath at temperature $T = 1/\beta$, the probability of observing a particular configuration $\boldsymbol{x} \in \{\pm 1\}^K$ is given by (2.28). The most probable configuration, ground state, is the one that minimizes the Hamiltonian $E(\boldsymbol{x})$ and is consistent with the constraints imposed at the macroscopic level. For very low temperature $\beta \to \infty$, the system is thus found with very high probability in its ground state. The normalization factor

$$Z = \sum_{\boldsymbol{x}} \mathrm{e}^{-\beta E(\boldsymbol{x})}, \tag{2.31}$$

in (2.28) is called the partition function, and it encodes the statistical properties of the system in the thermodynamic equilibrium. In theory, if we know $E(\boldsymbol{x})$ and the configuration space $\mathcal{X}^K$, all important macroscopic quantities (observables) of the related physical system can be calculated from $Z$. Oftentimes, though, the (normalized) thermodynamic quantity (given the limit exists)

$$F = -\lim_{K\to\infty} \frac{1}{\beta K} \log Z, \tag{2.32}$$

[5] We have implicitly made here the so-called ergodicity assumption, i.e., for the observable quantities the time-average equals the average over the probability distribution of the configuration space. For this reason we have also omitted the time dependence in the state variables $\{x_k\}_{k=1}^K$.

called (Helmholtz) free energy is more convenient for deriving the macroscopic variables instead[6]. The problem with (2.32) (or (2.31)) is, however, that except for some special cases (e.g., one and two dimensional Ising models), the computational complexity of calculating the partition function $Z$ directly is prohibitive due to the large number of particles $K$ in the system.

### 2.6.2 Spin Glasses and the Replica Method

One of the cases where the direct computation of the free energy (2.32) is infeasible arises with the aforementioned magnetic materials termed spin glasses. The mathematical model for the spin glasses is chosen here for simplicity to be defined by the Hamiltonian

$$E(\boldsymbol{x}) = -\frac{1}{\sqrt{K}} \sum_{k=1}^{K} \sum_{l=k+1}^{K} J_{l,k} x_k x_l, \tag{2.33}$$

where

$$\mathcal{J}_{\mathsf{s-k}} = \{J_{l,k} \mid \forall k = 1, \ldots, K \wedge l = k+1, \ldots, K\}, \tag{2.34}$$

is a set of $K(K-1)/2$ IID standard Gaussian random variables. In statistical physics terms, the set $\mathcal{J}_{\mathsf{s-k}}$ represents quenched disorder in the spin glass, i.e., it defines a random interaction between the spins that does not evolve with time. The Hamiltonian (2.33) represents a special case of the infinite S-K model of spin glasses without an external field. It should be remarked that the above is not by any means a realistic model for a physical spin glass since all sites in (2.33) are mutually coupled and their geometric locations neglected. Such a simplification is termed mean-field approximation in physics literature and we shall next consider how to obtain the free energy (2.32) for this simplified spin glass model.

In statistical mechanics, it is oftentimes postulated that in the thermodynamic limit $K \to \infty$, the free energy (2.32) converges to its quenched average, i.e.,

$$F = -\lim_{K\to\infty} \frac{1}{\beta K} \log Z = -\lim_{K\to\infty} \frac{1}{\beta K} \mathbb{E}_{\mathcal{J}_{\mathsf{s-k}}}\{\log Z\}, \tag{2.35}$$

where the expectation is with respect to the quenched randomness of the spin glass, namely, the interactions $\mathcal{J}_{\mathsf{s-k}}$. This is called the self-averaging property of the free

[6] In fact, other thermodynamic potentials, such as, Gibbs free energy and enthalpy exist and are better suited for some other cases. For our purposes, however, the free energy is the most convenient choice and the physical quantities of interest (for example, magnetization) can typically be expressed directly in terms of $F$ and its derivatives.

energy and can be proved to be true for the S-K model rigorously (see, e.g., [102–104]). For the rest of this thesis, however, the convergence of the type (2.35) is assumed to exists for the free energies of our interest, and the proof is considered to be out of the scope of the present dissertation.

**Assumption 4.** The thermodynamic limit of the free energy exists and it is self-averaging with respect to the quenched randomness present in the system. ♢

With considerable foresight, we proceed to calculate the RHS of (2.35) and expect this to be a simpler problem than the one encountered in (2.32). Unfortunately, assessing the expectation in (2.35) is still a formidable task. Somewhat simpler expression can be obtained if we define a real valued parameter $n$ and use the identity

$$\frac{\partial}{\partial n}\log\left(\mathbb{E}_{\mathcal{J}_{\mathsf{s-k}}}\{Z^n\}\right) = \mathbb{E}_{\mathcal{J}_{\mathsf{s-k}}}\left\{\frac{\partial Z^n}{\partial n}\right\}\frac{1}{\mathbb{E}_{\mathcal{J}_{\mathsf{s-k}}}\{Z^n\}} = \frac{\mathbb{E}_{\mathcal{J}_{\mathsf{s-k}}}\{Z^n\log(Z)\}}{\mathbb{E}_{\mathcal{J}_{\mathsf{s-k}}}\{Z^n\}}, \tag{2.36}$$

on the right hand side (RHS) of (2.35), i.e.,

$$F = -\lim_{K\to\infty}\frac{1}{\beta K}\mathbb{E}_{\mathcal{J}_{\mathsf{s-k}}}\{\log(Z)\} = -\lim_{K\to\infty}\frac{1}{\beta K}\lim_{n\to 0}\frac{\partial}{\partial n}\log(\mathbb{E}_{\mathcal{J}_{\mathsf{s-k}}}\{Z^n\}). \tag{2.37}$$

Note that given the Assumption 4 holds, the RHS in (2.37) is indeed equal to the free energy in the thermodynamic limit. The problem still persists, however, that evaluating the expectation for a real power $n \in \mathbb{R}$ of the partition function is in practice infeasible.

The basic idea of the replica method is to first calculate the moments of $Z$, i.e., $\mathbb{E}_{\mathcal{J}_{\mathsf{s-k}}}\{Z^n\}$ for an integer $n$ by introducing the statistically identical replicas (hence the name replica method) of the Hamiltonians

$$E(\boldsymbol{x}^{\{a\}}) = -\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\sum_{l=k+1}^{K} J_{l,k}x_k^{\{a\}}x_l^{\{a\}}, \qquad a = 1,\dots,n, \tag{2.38}$$

where $\boldsymbol{x}^{\{a\}} = \begin{bmatrix}x_1^{\{a\}} & x_2^{\{a\}} & \cdots & x_K^{\{a\}}\end{bmatrix}^{\mathsf{T}} \in \mathcal{X}^K$, for all $a = 1,\dots,n$. Then the limit in (2.37) is taken as if $n$ was real valued. In order to help the evaluation of the summations over the replicated spin configurations, it is further postulated that the limits commute and the free energy under the replica method can be written as

$$F_{\mathsf{rm}} = -\lim_{n\to 0}\frac{\partial}{\partial n}\lim_{K\to\infty}\frac{1}{\beta K}\log\left(\mathbb{E}_{\mathcal{J}_{\mathsf{s-k}}}\left\{\prod_{a=1}^{n}\sum_{\boldsymbol{x}^{\{a\}}\in\mathcal{X}^K}\mathrm{e}^{-\beta E(\boldsymbol{x}^{\{a\}})}\right\}\right). \tag{2.39}$$

Now, for the Gaussian interactions $\mathcal{J}_{\mathsf{s}-\mathsf{k}}$, we can first assess the expectation in (2.39) with the help of the Gaussian integral (see (C.22)). Furthermore, in the limit $K \to \infty$, one can next use the results from large deviation theory (for example, [179, 180]) to calculate the summations over the replicated configurations $\{\boldsymbol{x}^{\{a\}}\}_{a=1}^{n}$. See [13, 14] for the actual computations in the S-K model and Appendix C in the context of channel estimation.

Thus far we have simplified the original problem of computing an expectation of a logarithm as encountered in (2.35), to Gaussian integrals and optimization problems to find the saddle-points of exponential functions. Unfortunately, it is not guaranteed that $F_{\mathsf{rm}}$ equals $F$ for general Hamiltonians and quenched disorder. Furthermore, one usually needs to limit the state space of the saddle-point conditions in order to get a closed form solution for the free energy. This can be achieved by defining a correlation matrix $\boldsymbol{Q} = [Q^{\{a,b\}}]_{n\times n}$ with elements

$$Q^{\{a,b\}} = \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}\Big\{x_k^{\{a\}} x_k^{\{b\}}\Big\}, \qquad a,b = 1,2,\ldots,n, \tag{2.40}$$

and imposing symmetry conditions on $\boldsymbol{Q}$. In this thesis, we consider only the replica symmetric saddle-points (see Assumption 7 in Appendix C), which translates for the S-K model as the condition

$$Q^{\{a,b\}} = q_{\mathsf{rs}}, \qquad \forall a \neq b. \tag{2.41}$$

Quite remarkably, obtaining the free energy for the S-K model under the assumption of RS reduces then to solving a fixed point equation and a single integral (cf. [13, Eqs. (3.21) and (3.22)] and [14, Eqs. (2.27) and (2.30)])

$$F_{\mathsf{rm}-\mathsf{rs}} = -\frac{1}{4}\beta(1-q_{\mathsf{rs}})^2 - \frac{1}{\beta}\int \log\left(2\cosh(\beta\nu\sqrt{q_{\mathsf{rs}}})\right)\mathrm{D}\nu, \tag{2.42}$$

$$q_{\mathsf{rs}} = \int \tanh^2(\beta\nu\sqrt{q_{\mathsf{rs}}})\mathrm{D}\nu, \tag{2.43}$$

where we used the short hand notation

$$\mathrm{D}\nu = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}\nu^2}\mathrm{d}\nu, \tag{2.44}$$

for the standard Gaussian measure.

The condition (2.41) may sound intuitively very reasonable since the replicas were introduced merely as a mathematical trick to compute the expectation of a power. Unfortunately (again), it is known that the RS free energy $F_{\mathsf{rm}-\mathsf{rs}}$ is not the

correct solution[7] for the S-K model due to replica symmetry breaking (RSB) [13, 14]. It does, however, give a very good approximation in general, and sometimes even the exact solution (say, for Gaussian spins). Due to the relatively complex Hamiltonians encountered in the latter parts of the dissertation, we have left the investigation of RSB as a future topic.

[7] The RS solution of the free energy gives in fact negative ground state entropy and energy, that is, when $T \to 0$. The correct form of the free energy for the S-K model has been recently proved [104, 181] to be the so-called Parisi formula proposed in [177].

# Chapter 3

# Non-Iterative Receivers for MIMO DS-CDMA in Flat Fading Channels

In this chapter, the performance of multi-antenna DS-CDMA operating over a narrowband Rayleigh fading MIMO channel is examined. The spatial correlation is assumed to be given by the Kronecker model, as discussed in Section 2.2.1. The receiver is composed of a non-iterative channel estimator and MUD — the latter of which is not necessarily linear.

The outline of the chapter is as follows. Section 3.1 derives the pilot-aided channel estimators and non-iterative multiuser detectors suitable for MIMO DS-CDMA from the class of generalized posterior mean estimators. The specific instances that will be considered in detail are in decreasing order of complexity:

- ***Channel estimators:*** linear MMSE, covariance mismatched LMMSE and maximum likelihood estimators;
- ***Data detectors:*** non-linear MAP, LMMSE, decorrelating and conventional detectors.

In Section 3.2, the equivalent single-user representations of the multiuser systems utilizing the components listed above are derived with the help of the replica method. Using this single-user characterization, the large system performance analysis of the multiuser receivers is carried out in Section 3.3. Selected numerical examples are given in Section 3.4.

The channel estimators and multiuser detectors presented in this chapter are studied with less detail than the iterative receivers considered in the next chapter. The proofs follow along the same lines as the ones in Chapter 4, and are mostly omitted. Brief discussion on the diagonalization of the equivalent noise covariance matrices can be found in Appendix A. For notational convenience, we omit the

block index $c$ in the following discussion.

## 3.1 Multiuser Receivers via Bayesian Framework

Here we briefly describe the non-iterative CEs and MUDs for multiuser MIMO DS-CDMA that will be analyzed later in the chapter. The outline is as follows.

- Section 3.1.1 presents a class of linear channel estimators for correlated MIMO channels. Linear MMSE and ML estimators assuming different levels of a priori knowledge about the channel conditions are given.
- Section 3.1.2 derives the MAP detector for the considered MIMO system under the assumption of imperfect CSI. The equivalent detector for the case of perfect CSI is oftentimes called the individually optimum detector in the literature [5, 74, 85].
- Section 3.1.3 introduces a class of linear MUDs derived under the assumption of mismatched channel information. Specific detectors considered include the LMMSE and decorrelating MUDs and the SUMF.

### 3.1.1 Linear Channel Estimation

Consider the set of vectors $\mathcal{Y}_{\mathcal{T}}$ received during the training phase $\mathcal{T}$ in (2.1). For notational convenience, define also a spreading matrix

$$\boldsymbol{S}_{k,t} = [\boldsymbol{s}_{k,1,t} \ \cdots \ \boldsymbol{s}_{k,L,t}]^{\mathsf{T}} \in \mathcal{M}^{L\times M}, \tag{3.1}$$

and a combined data-spreading matrix

$$\boldsymbol{G}_k = \frac{1}{\sqrt{L}}[\boldsymbol{P}_{k,1}\boldsymbol{S}_{k,1}^{\mathsf{T}} \ \cdots \ \boldsymbol{P}_{k,\tau_{\mathsf{tr}}}\boldsymbol{S}_{k,\tau_{\mathsf{tr}}}^{\mathsf{T}}]^{\mathsf{T}} \otimes \boldsymbol{I}_N \in \mathcal{M}^{\tau_{\mathsf{tr}}LN\times MN}. \tag{3.2}$$

Note that (3.2) is perfectly known at the receiver (cf. Section 2.2.1). The information contained in the set $\mathcal{Y}_{\mathcal{T}}$ and the channel model (2.1) during the time $t \in \mathcal{T}$ and chip $l = 1, \ldots, L$ indices can then be written in the vector form

$$\boldsymbol{y}_{\mathcal{T}} = \sum_{k=1}^{K} \boldsymbol{G}_k \underline{\boldsymbol{H}}_k + \boldsymbol{w}_{\mathcal{T}} \in \mathbb{C}^{\tau_{\mathsf{tr}}LN}, \tag{3.3}$$

where the complex Gaussian random vector $\boldsymbol{w}_{\mathcal{T}} \sim \mathsf{CN}(\boldsymbol{0}; \sigma^2\boldsymbol{I})$ represents the samples of thermal noise during the training phase.

Recall the notation introduced in Section 2.1. In order to derive a channel estimator based on the knowledge of $\boldsymbol{y}_{\mathcal{T}}$ and $\{\boldsymbol{G}_k\}_{k\in\mathcal{K}}$, let us postulate a new channel model related to (3.3) as

$$\tilde{\boldsymbol{y}}_{\mathcal{T}} = \sum_{k=1}^{K} \boldsymbol{G}_k \underline{\tilde{\boldsymbol{H}}}_k + \tilde{\boldsymbol{w}}_{\mathcal{T}} \in \mathbb{C}^{\tau_{\mathrm{tr}} LN}. \tag{3.4}$$

We let $\tilde{\sigma}^2$ be the postulated variance of the Gaussian noise $\tilde{\boldsymbol{w}}_{\mathcal{T}} \sim \mathsf{CN}(\boldsymbol{0}; \tilde{\sigma}^2 \boldsymbol{I})$, and assign prior distributions (to be defined later) $\mathbb{Q}(\underline{\tilde{\boldsymbol{H}}}_k)$ to the postulated channel coefficients $\underline{\tilde{\boldsymbol{H}}}_k \in \mathbb{C}^{MN}\ \forall k = 1, \ldots, K$. The covariance matrix of the postulated channel is written as $\tilde{\boldsymbol{\Omega}}_{\underline{\boldsymbol{H}}_k} = \tilde{\boldsymbol{T}}_k \otimes \tilde{\boldsymbol{R}}$, where $\tilde{\boldsymbol{T}}_k$ and $\tilde{\boldsymbol{R}}$ are the postulated covariance matrices related to the transmitter and receiver side spatial correlation. For later use we denote $\tilde{\mathcal{H}} = \{\underline{\tilde{\boldsymbol{H}}}_k \mid k = 1, \ldots, K\}$ for the set of postulated channels of all users and let the postulated channels between the users be independent, i.e., $\mathbb{Q}(\tilde{\mathcal{H}}) = \prod_{k=1}^{K} \mathbb{Q}(\underline{\tilde{\boldsymbol{H}}}_k)$.

We can now interpret (3.4) as the receiver's knowledge about the true channel model (3.3). All estimation algorithms are then based on the postulated information (3.4), and the prior probabilities associated with the RVs in it. In general, if the two system models (3.3) and (3.4) with the accompanying prior probabilities are not the same, the resulting estimator can be thought of to be a mismatched solution to a Bayesian inference problem. The mismatch may arise from a limited knowledge about the parameters involved, or from a conscious choice. Albeit the latter may seem like a strange position to take at first, it makes sense from the point of view of system design when the limited resources prevent employing the optimum strategies (see Sections 3.1.2 and 3.1.3). This method of deriving the desired estimators and detectors is used for the rest of the thesis.

Denote $\mathcal{I} = \{\mathcal{S}, \mathcal{P}, \mathcal{Y}_{\mathcal{T}}\}$, and let

$$\mathbb{Q}(\underline{\tilde{\boldsymbol{H}}}_\xi \mid \mathcal{I}) = \frac{\mathbb{Q}(\underline{\tilde{\boldsymbol{H}}}_\xi)\mathbb{E}_{\tilde{\mathcal{H}}\setminus\underline{\tilde{\boldsymbol{H}}}_\xi}\{\mathbb{Q}(\tilde{\boldsymbol{y}}_{\mathcal{T}} = \boldsymbol{y}_{\mathcal{T}} \mid \mathcal{I}, \tilde{\mathcal{H}})\}}{\mathbb{E}_{\tilde{\mathcal{H}}}\{\mathbb{Q}(\tilde{\boldsymbol{y}}_{\mathcal{T}} = \boldsymbol{y}_{\mathcal{T}} \mid \mathcal{I}, \tilde{\mathcal{H}})\}}, \tag{3.5}$$

be the postulated a posteriori probability (APP) of the channel coefficients of the user $\xi \in \mathcal{K}$, given the information $\mathcal{I}$ and channel model (3.4). The resulting GPME [85, 89] reads

$$\langle \underline{\tilde{\boldsymbol{H}}}_\xi \rangle = \int \underline{\tilde{\boldsymbol{H}}}_\xi \mathrm{d}\mathbb{Q}(\underline{\tilde{\boldsymbol{H}}}_\xi \mid \mathcal{I}) \in \mathbb{C}^{MN}, \tag{3.6}$$

where $\langle \underline{\tilde{\boldsymbol{H}}}_\xi \rangle = \mathrm{vec}([\langle \tilde{\boldsymbol{h}}_{\xi,1} \rangle \ \cdots \ \langle \tilde{\boldsymbol{h}}_{\xi,M} \rangle])$, and $\{\langle \tilde{\boldsymbol{h}}_{\xi,m} \rangle\}_{m=1}^{M}$ are the estimates of $\{\boldsymbol{h}_{\xi,m}\}_{m=1}^{M}$. If we postulate Gaussian priors $\mathbb{Q}(\underline{\tilde{\boldsymbol{H}}}_k) = \mathsf{CN}(\boldsymbol{0}; \tilde{\boldsymbol{\Omega}}_{\underline{\boldsymbol{H}}_k})\ \forall k \in \mathcal{K}$,

the integrals in (3.5) – (3.6) can be calculated with the help of (C.22), which yields after some algebra[1] a linear estimator

$$\langle \underline{\tilde{\boldsymbol{H}}}_\xi \rangle = \tilde{\boldsymbol{\Omega}}_{\underline{\tilde{\boldsymbol{H}}}_\xi} \boldsymbol{G}_\xi^{\mathsf{H}} \left( \sum_{k=1}^{K} \boldsymbol{G}_k \tilde{\boldsymbol{\Omega}}_{\underline{\boldsymbol{H}}_k} \boldsymbol{G}_k^{\mathsf{H}} + \tilde{\sigma}^2 \boldsymbol{I}_{\tau_{\mathsf{tr}} LN} \right)^{-1} \boldsymbol{y}_{\mathcal{T}}, \tag{3.7}$$

that is parametrized by:

1. $\{\tilde{\boldsymbol{\Omega}}_{\underline{\boldsymbol{H}}_k}\}_{k=1}^K$, the postulated covariance matrices of the MIMO channels;
2. $\tilde{\sigma}^2$, the postulated noise variance.

When the GPME (3.6) is considered in the following, the underlying assumption is always that Gaussian priors leading to (3.7) are postulated. For later use, we let $\Delta\underline{\boldsymbol{H}}_k = \underline{\boldsymbol{H}}_k - \langle \underline{\tilde{\boldsymbol{H}}}_k \rangle$ be the error of the channel estimates, and $\boldsymbol{\Omega}_{\Delta\underline{\boldsymbol{H}}_k} = \mathbb{E}\{\Delta\underline{\boldsymbol{H}}_k \Delta\underline{\boldsymbol{H}}_k^{\mathsf{H}}\}$ the corresponding covariance matrix. The error covariance estimate obtained by the CE is denoted by $\tilde{\boldsymbol{\Omega}}_{\Delta\underline{\boldsymbol{H}}_k}$. Note that $\tilde{\boldsymbol{\Omega}}_{\Delta\underline{\boldsymbol{H}}_k}$ can be different from $\boldsymbol{\Omega}_{\Delta\underline{\boldsymbol{H}}_k}$, in which case the MUD is misinformed about the error statistics of the channel estimates.

**Example 2.** Let $\tilde{\sigma}^2 = \sigma^2$ and $\tilde{\boldsymbol{\Omega}}_{\underline{\boldsymbol{H}}_k} = \boldsymbol{\Omega}_{\underline{\boldsymbol{H}}_k}$. The resulting estimator is the optimum non-iterative pilot assisted MMSE channel estimator for the channel model (3.3). ♢

**Example 3.** Let $\tilde{\sigma}^2 = \sigma^2$ and assume that the CE knows the diagonals $\{T_{k,m,m}\}_{m=1}^M$ of $\boldsymbol{T}_k$, but neglects the correlations between the transmit antennas. We define two mismatched CEs based on their knowledge about the receive correlation:

- Type-1: $\tilde{\boldsymbol{R}} = \boldsymbol{R}$
- Type-2: $\tilde{\boldsymbol{R}} = \boldsymbol{I}_N$

The resulting CEs are called covariance mismatched LMMSE channel estimators of Type-1 and Type-2 for the rest of the chapter. ♢

**Example 4.** Let $\tilde{\boldsymbol{\Omega}}_{\underline{\boldsymbol{H}}_k} = \boldsymbol{I}_{MN}$ and $\tilde{\sigma}^2 \to 0$. The GPME (3.6) reduces then to the ML channel estimator for the MIMO channel (3.3). ♢

**Remark 1.** In the following we assume that the posterior distribution (3.5) of the channel estimates satisfies $\mathbb{Q}(\tilde{\mathcal{H}} \mid \mathcal{I}) = \prod_{k=1}^K \mathbb{Q}(\underline{\tilde{\boldsymbol{H}}}_k \mid \mathcal{I})$, although this may not strictly hold due to joint estimation over the users. ♢

[1] The matrix identity $\boldsymbol{A} - \boldsymbol{A}\boldsymbol{U}(\boldsymbol{C} + \boldsymbol{V}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{U})^{-1}\boldsymbol{V}^{\mathsf{H}}\boldsymbol{A} = (\boldsymbol{A}^{-1} + \boldsymbol{U}\boldsymbol{C}^{-1}\boldsymbol{V}^{\mathsf{H}})^{-1}$, where the inverses are assumed to exist, is very helpful here (see, e.g. [182] and the references therein).

### 3.1.2 Non-Linear MAP Detector with CSI Mismatch

The first MUD to be considered for the MIMO DS-CDMA system is the non-linear MAP detector. We start by re-writing (2.1) in the form

$$\boldsymbol{y}_{l,t} = \sum_{k=1}^{K} \boldsymbol{G}_{k,l,t}\underline{\boldsymbol{H}}_k + \boldsymbol{w}_{l,t} \in \mathbb{C}^N, \tag{3.8}$$

where shorthand notation

$$\boldsymbol{G}_{k,l,t} = \frac{1}{\sqrt{L}}\big(\boldsymbol{X}_{k,t}\boldsymbol{s}_{k,l,t} \otimes \boldsymbol{I}_N\big)^{\mathsf{T}} \in \mathbb{C}^{N\times MN}, \tag{3.9}$$

was used for convenience. Suppose that before MUD, the CSI is obtained by the LMMSE channel estimator of Example 2 in the form

$$\mathbb{Q}(\underline{\tilde{\boldsymbol{H}}}_k \mid \mathcal{I}) = \mathsf{CN}\big(\langle\underline{\tilde{\boldsymbol{H}}}_k\rangle; \tilde{\boldsymbol{\Omega}}_{\Delta\underline{\boldsymbol{H}}_k}\big). \tag{3.10}$$

We then postulate a new channel model related to (3.8)

$$\tilde{\boldsymbol{y}}_{l,t} = \sum_{k=1}^{K} \tilde{\boldsymbol{G}}_{k,l,t}\underline{\tilde{\boldsymbol{H}}}_k + \tilde{\boldsymbol{w}}_{l,t} \in \mathbb{C}^N, \tag{3.11}$$

where the matrix $\tilde{\boldsymbol{G}}_{k,l,t} = \frac{1}{\sqrt{L}}\big(\tilde{\boldsymbol{X}}_{k,t}\boldsymbol{s}_{k,l,t} \otimes \boldsymbol{I}_N\big)^{\mathsf{T}}$, contains the spreading sequence $\boldsymbol{s}_{k,l,t}$ and the postulated data $\tilde{\boldsymbol{X}}_{k,t} = \mathrm{diag}(\tilde{\boldsymbol{x}}_{k,t})$. We also denote for later use $\tilde{\mathcal{X}}_t = \{\tilde{\boldsymbol{x}}_{k,t} \mid k = 1,\ldots,K\}$ for the set of all transmitted vectors during the $t$th index in the postulated channel model (3.11).

Now, assign the true prior probabilities $\mathbb{Q}(\tilde{\boldsymbol{x}}_{k,t} = \boldsymbol{x}_{k,t}) = \mathbb{P}(\boldsymbol{x}_{k,t})\ \forall k \in \mathcal{K}$ to the data symbols and let the noise be zero-mean complex Gaussian $\tilde{\boldsymbol{w}}_{l,t} \sim \mathsf{CN}(\boldsymbol{0}; \tilde{\sigma}^2\boldsymbol{I})$ with the correct variance $\tilde{\sigma}^2 = \sigma^2$. The channel estimates are introduced to the system model (3.11) by taking the conditional expectation over the posterior probabilities (3.5), resulting to

$$\begin{aligned}&\mathbb{Q}(\tilde{\boldsymbol{y}}_{l,t} = \boldsymbol{y}_{l,t} \mid \mathcal{I}, \tilde{\mathcal{X}}_t, \mathcal{Y}_t)\\ &\quad= \int \mathbb{Q}(\tilde{\boldsymbol{y}}_{l,t} = \boldsymbol{y}_{l,t} \mid \mathcal{I}, \tilde{\mathcal{X}}_t, \mathcal{Y}_t, \{\underline{\tilde{\boldsymbol{H}}}_k\}_{k=1}^K) \prod_{k=1}^{K} \mathrm{d}\mathbb{Q}(\underline{\tilde{\boldsymbol{H}}}_k \mid \mathcal{I}), \quad l = 1,\ldots,L.\end{aligned} \tag{3.12}$$

Note that for Rayleigh fading channel the integrals can be calculated in closed form and (3.12) is a conditional Gaussian distribution. The postulated APPs of the transmitted symbols $\boldsymbol{x}_{\xi,t}$ are then given by

$$\mathbb{Q}(\tilde{\boldsymbol{x}}_{\xi,t} \mid \mathcal{I}, \mathcal{Y}_t) = \frac{\mathbb{Q}(\tilde{\boldsymbol{x}}_{\xi,t})\mathbb{E}_{\tilde{\mathcal{X}}_t\setminus\tilde{\boldsymbol{x}}_{\xi,t}}\Big\{\prod_{l=1}^{L}\mathbb{Q}(\tilde{\boldsymbol{y}}_{l,t} = \boldsymbol{y}_{l,t} \mid \mathcal{I}, \tilde{\mathcal{X}}_t, \mathcal{Y}_t)\Big\}}{\mathbb{E}_{\tilde{\mathcal{X}}_t}\Big\{\prod_{l=1}^{L}\mathbb{Q}(\tilde{\boldsymbol{y}}_{l,t} = \boldsymbol{y}_{l,t} \mid \mathcal{I}, \tilde{\mathcal{X}}_t, \mathcal{Y}_t)\Big\}}, \tag{3.13}$$

where the expectations are with respect to the postulated a priori probabilities of the data symbols. The posterior mean estimate reads

$$\langle \tilde{\boldsymbol{x}}_{\xi,t} \rangle = \sum_{\tilde{\boldsymbol{x}}_{\xi,t} \in \mathcal{M}^M} \tilde{\boldsymbol{x}}_{\xi,t} \mathbb{Q}(\tilde{\boldsymbol{x}}_{\xi,t} \mid \mathcal{I}, \mathcal{Y}_t). \tag{3.14}$$

### 3.1.3 Linear Multiuser Detection with CSI Mismatch

Having derived the optimum non-linear MUD in the previous section, we now turn to the computationally less complex linear detectors. If the channel would be perfectly known at the receiver, one could simply proceed by assigning Gaussian priors $\mathbb{Q}(\tilde{\boldsymbol{x}}_{k,t}) = \mathsf{CN}(\boldsymbol{0};\, \boldsymbol{I})$ to all users in (3.11) (see, e.g., [74, 85, 89]). When the channel knowledge contains uncertainty, however, such an approach does not directly apply.

Consider the data transmission phase $t \in \mathcal{D}$ in (2.1). Let the channel estimates $\{\langle \tilde{\boldsymbol{h}}_{k,m} \rangle \mid \forall k, m\}$ and the error covariance matrices $\{\tilde{\boldsymbol{\Omega}}_{\Delta \underline{\boldsymbol{H}}_k} \mid \forall k\}$ obtained by the channel estimator as described in Section 3.1.1 be available to the MUD. Define an error term

$$\Delta \underline{\boldsymbol{V}}_{k,t} = [\Delta \boldsymbol{v}_{k,t,1}^{\mathsf{T}} \cdots \Delta \boldsymbol{v}_{k,t,M}^{\mathsf{T}}]^{\mathsf{T}} = (\boldsymbol{X}_{k,t} \otimes \boldsymbol{I}_N) \Delta \underline{\boldsymbol{H}}_k \in \mathbb{C}^{MN} \tag{3.15}$$

where $\Delta \boldsymbol{v}_{k,t,m} \in \mathbb{C}^N \ \forall m$, and write (3.6) as

$$\langle \tilde{\boldsymbol{H}}_k \rangle_{\mathsf{d}} = \begin{bmatrix} \langle \tilde{\boldsymbol{h}}_{k,1} \rangle & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \langle \tilde{\boldsymbol{h}}_{k,M} \rangle \end{bmatrix} \in \mathbb{C}^{MN \times M}. \tag{3.16}$$

Using (3.1), (3.15) and (3.16), an equivalent re-presentation for the received signal $\mathcal{Y}_t$ in (2.1) during the data transmission phase $t \in \mathcal{D}$ is obtained

$$\boldsymbol{y}_t = \frac{1}{\sqrt{L}} \sum_{k=1}^{K} (\boldsymbol{S}_{k,t} \otimes \boldsymbol{I}_N) \langle \tilde{\boldsymbol{H}}_k \rangle_{\mathsf{d}} \boldsymbol{x}_{k,t} + (\boldsymbol{S}_{k,t} \otimes \boldsymbol{I}_N) \Delta \underline{\boldsymbol{V}}_{k,t} + \boldsymbol{w}_t \in \mathbb{C}^{LN}. \tag{3.17}$$

Note that so-far we have not changed the channel model, and writing out (3.17) returns the same dependence between the received and transmitted vectors as in (2.1).

The receiver has knowledge of $\{\boldsymbol{S}_{k,t}\}$ and $\{\langle \tilde{\boldsymbol{H}}_k \rangle_{\mathsf{d}}\}$, and the (possibly mismatched) statistics of the channel estimation errors $\{\Delta \underline{\boldsymbol{H}}_k\}$. Therefore, if we are interested in estimating $\boldsymbol{X}_{\xi,t}$, for the user $\xi \in \mathcal{K}$, (3.15) contains a multiplicative error term and postulating Gaussian prior for the data does not yield a linear MUD as

desired. It is worth pointing out though that if the estimation error $\Delta\underline{\boldsymbol{H}}_k$ is complex Gaussian and QPSK signaling is used, the true error term (3.15) is indeed complex Gaussian. This is in fact what happens with the MAP detector discussed in previous section. Intuitively, we could then argue that (3.15) contains no information from the receiver's point of view and treat $\Delta\underline{\boldsymbol{V}}_{k,t}$ as additional complex Gaussian noise in the channel. However, this does not hold formally if we postulate Gaussian prior for the data.

Let $\xi \in \mathcal{K}$ be the user of interest and denote for notational convenience $\mathcal{I}_t = \{\mathcal{P}, \mathcal{S}, \mathcal{Y}_t, \{\langle\tilde{\underline{\boldsymbol{H}}}_k\rangle\}_{k\in\mathcal{K}}\}$. Let

$$\tilde{\boldsymbol{y}}_t = \frac{1}{\sqrt{L}}\sum_{k=1}^{K}(\boldsymbol{S}_{k,t}\otimes\boldsymbol{I}_N)\langle\tilde{\boldsymbol{H}}_k\rangle_{\mathsf{d}}\tilde{\boldsymbol{x}}_{k,t} + (\boldsymbol{S}_{k,t}\otimes\boldsymbol{I}_N)\Delta\tilde{\underline{\boldsymbol{V}}}_{k,t} + \tilde{\boldsymbol{w}}_t \in \mathbb{C}^{LN}, \quad (3.18)$$

be the receiver's knowledge of (3.17), where $\tilde{\boldsymbol{w}}_t \sim \mathsf{CN}(\boldsymbol{0};\, \tilde{\sigma}^2\boldsymbol{I}_{LN})$ and the interfering users have postulated Gaussian priors $\mathbb{Q}(\tilde{\boldsymbol{x}}_{j,t}) = \mathsf{CN}(\boldsymbol{0};\, \boldsymbol{I})\ \forall j \in \mathcal{K}\setminus\xi$. Furthermore, let $\Delta\tilde{\underline{\boldsymbol{V}}}_{k,t} = \mathrm{vec}([\Delta\tilde{\boldsymbol{v}}_{k,t,1}\ \cdots\ \Delta\tilde{\boldsymbol{v}}_{k,t,M}]) \in \mathbb{C}^{MN}\ \forall k \in \mathcal{K}$ with the postulated distribution $\mathbb{Q}(\Delta\tilde{\underline{\boldsymbol{V}}}_{k,t} \mid \mathcal{I}_t) = \mathsf{CN}(\boldsymbol{0};\, \tilde{\boldsymbol{\Omega}}_{\Delta\underline{\boldsymbol{V}}_{k,t}})$ represent the receiver's knowledge about the error term (3.15). The postulated APP of the $\xi$th user's data symbols transmitted during the time index $t \in \mathcal{D}$ reads then

$$\begin{aligned}&\mathbb{Q}(\tilde{\boldsymbol{x}}_{\xi,t} \mid \mathcal{I}_t)\\ &= \frac{\mathbb{Q}(\tilde{\boldsymbol{x}}_{\xi,t})\mathbb{E}_{\tilde{\mathcal{X}}_t\setminus\tilde{\boldsymbol{x}}_{\xi,t}}\left\{\mathbb{E}_{\{\Delta\tilde{\underline{\boldsymbol{V}}}_{k,t}\}_{k\in\mathcal{K}}}\{\mathbb{Q}(\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t \mid \mathcal{I}_t, \tilde{\mathcal{X}}_t, \{\Delta\tilde{\underline{\boldsymbol{V}}}_{k,t}\}_{k\in\mathcal{K}},)\}\right\}}{\mathbb{E}_{\tilde{\mathcal{X}}_t}\left\{\mathbb{E}_{\{\Delta\tilde{\underline{\boldsymbol{V}}}_{k,t}\}_{k\in\mathcal{K}}}\{\mathbb{Q}(\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t \mid \mathbb{Q}(\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t \mid \mathcal{I}_t, \tilde{\mathcal{X}}_t, \{\Delta\tilde{\underline{\boldsymbol{V}}}_{k,t}\}_{k\in\mathcal{K}},)\}\right\}},\end{aligned} \quad (3.19)$$

where $\mathbb{Q}(\tilde{\boldsymbol{x}}_{\xi,t})$ is the postulated prior for the user of interest. Plugging (3.19) for the posterior distribution in

$$\langle\tilde{\boldsymbol{x}}_{\xi,t}\rangle = \int \tilde{\boldsymbol{x}}_{\xi,t}\mathrm{d}\mathbb{Q}(\tilde{\boldsymbol{x}}_{\xi,t} \mid \mathcal{I}_t), \quad (3.20)$$

gives the desired GPME[2] that is parametrized by:

1. $\mathbb{Q}(\tilde{\boldsymbol{x}}_{\xi,t})$, the a priori probability of the transmit symbols;
2. $\tilde{\boldsymbol{\Omega}}_{\Delta\underline{\boldsymbol{V}}_{k,t}}$, the postulated covariance of the error term arising from the channel estimation errors;
3. $\tilde{\sigma}^2$, the postulated noise variance.

[2] See the Remark 4 in Section 4.1.5 for discussion on how to treat posterior mean estimates when belief propagation based channel decoder is used.

Note that if we postulate $\mathbb{Q}(\tilde{\boldsymbol{x}}_{\xi,t}) = \mathsf{CN}(\boldsymbol{0};\, \boldsymbol{I})$, the GPME (3.20) reduces to the familiar linear form

$$\langle \boldsymbol{x}_{\xi,t} \rangle = \frac{1}{\sqrt{L}} \langle \boldsymbol{H}_{\xi} \rangle_{\mathsf{d}}^{\mathsf{H}} (\boldsymbol{S}_{\xi,t}^{\mathsf{H}} \otimes \boldsymbol{I}_N) \times \left( \tilde{\sigma}^2 \boldsymbol{I}_{LN} + \frac{1}{L} \sum_{k=1}^{K} (\boldsymbol{S}_{k,t} \otimes \boldsymbol{I}_N)(\langle \underline{\tilde{\boldsymbol{H}}}_k \rangle \langle \underline{\tilde{\boldsymbol{H}}}_k \rangle^{\mathsf{H}} + \tilde{\boldsymbol{\Omega}}_{\Delta \underline{\boldsymbol{V}}_{k,t}})(\boldsymbol{S}_{k,t}^{\mathsf{H}} \otimes \boldsymbol{I}_N) \right)^{-1} \boldsymbol{y}_t. \tag{3.21}$$

**Example 5.** Let us denote $\mathbb{E}^{\mathsf{d}}\{\,\cdot\,\} = \mathbb{E}\{\,\cdot \mid \mathcal{P}, \mathcal{S}, \{\langle \underline{\tilde{\boldsymbol{H}}}_k \rangle\}_{k \in \mathcal{K}}\}$ and assume that $\mathbb{E}^{\mathsf{d}}\{\underline{\boldsymbol{H}}_k \underline{\boldsymbol{H}}_k^{\mathsf{H}}\} = \langle \underline{\tilde{\boldsymbol{H}}}_k \rangle \langle \underline{\tilde{\boldsymbol{H}}}_k \rangle^{\mathsf{H}} + \tilde{\boldsymbol{\Omega}}_{\Delta \underline{\boldsymbol{H}}_k}$. The LMMSE estimator

$$\langle \boldsymbol{x}_{k,t} \rangle = \mathbb{E}^{\mathsf{d}}\{\boldsymbol{x}_{k,t} \boldsymbol{y}_t^{\mathsf{H}}\} \mathbb{E}^{\mathsf{d}}\{\boldsymbol{y}_t \boldsymbol{y}_t^{\mathsf{H}}\}^{-1} \boldsymbol{y}_t, \tag{3.22}$$

equals then (3.21) with $\tilde{\sigma}^2 = \sigma^2$ and $\tilde{\boldsymbol{\Omega}}_{\Delta \underline{\boldsymbol{V}}_{k,t}} = \tilde{\boldsymbol{\Omega}}_{\Delta \underline{\boldsymbol{H}}_k}$. This is akin to the linear MMSE data estimator studied for the single-antenna multipath DS-CDMA systems by Evans & Tse [59]. ♢

**Example 6.** If we set $\tilde{\boldsymbol{\Omega}}_{\Delta \underline{\boldsymbol{V}}_{k,t}} = \boldsymbol{0}$ in (3.21) and then let $\tilde{\sigma}^2 \to \infty$ or $\tilde{\sigma}^2 \to 0$, we get the single-user matched filter and decorrelator that assumes perfect CSI, respectively. ♢

## 3.2 Decoupling Results

In this section, the decoupling of the multiuser MIMO DS-CDMA system described in Sections 2.2.1 and 3.1 is presented. We assume for simplicity that $p_{\mathsf{tx}}$ contains a single mass point and the MIMO channels of the users are therefore IID. The decoupling of the multiuser channel is obtained via an application of the replica method by using the same methodology as in [85, 89, 92]. An analogous case can be found in [183, Sec. V]. In deriving the decoupling results we have assumed that the assumptions made in the replica analysis are valid and replica symmetry holds. Note that the replica method relies on the large system limit where $K = \alpha L \to \infty$ with fixed system load $0 < \alpha < \infty$.

### 3.2.1 Linear Channel Estimation

Consider a set of single-user channels

$$\boldsymbol{z}_{k,m} = \boldsymbol{h}_{k,m} + \boldsymbol{w}_{k,m} \in \mathbb{C}^N, \qquad \boldsymbol{w}_{k,m} \sim \mathsf{CN}(\boldsymbol{0};\, \boldsymbol{C}), \tag{3.23}$$

$$\tilde{\boldsymbol{z}}_{k,m} = \tilde{\boldsymbol{h}}_{k,m} + \tilde{\boldsymbol{w}}_{k,m} \in \mathbb{C}^N, \qquad \tilde{\boldsymbol{w}}_{k,m} \sim \mathsf{CN}(\boldsymbol{0};\, \tilde{\boldsymbol{C}}), \tag{3.24}$$

where $m = 1, \ldots, M$, and

$$\underline{\boldsymbol{H}}_k = \mathrm{vec}\big([\boldsymbol{h}_{k,1} \ \cdots \ \boldsymbol{h}_{k,M}]\big) \sim \mathsf{CN}(\boldsymbol{0};\, \boldsymbol{\Omega}_{\underline{\boldsymbol{H}}}), \tag{3.25}$$

$$\underline{\tilde{\boldsymbol{H}}}_k = \mathrm{vec}\big([\tilde{\boldsymbol{h}}_{k,1} \ \cdots \ \tilde{\boldsymbol{h}}_{k,M}]\big) \sim \mathsf{CN}(\boldsymbol{0};\, \tilde{\boldsymbol{\Omega}}_{\underline{\boldsymbol{H}}}). \tag{3.26}$$

Let the GPME for the true channel coefficients $\boldsymbol{h}_{k,m}, m = 1, \ldots, M$, based on the knowledge of (3.24) and (3.26), be given by

$$\langle \tilde{\boldsymbol{h}}_{k,m} \rangle_k = \frac{\mathbb{E}_{\{\tilde{\boldsymbol{h}}_{k,i}\}_{i=1}^M} \{ \tilde{\boldsymbol{h}}_{k,m} \prod_{i=1}^M \mathbb{Q}(\tilde{\boldsymbol{z}}_{k,i} \mid \tilde{\boldsymbol{h}}_{k,i}) \}}{\mathbb{E}_{\{\tilde{\boldsymbol{h}}_{k,i}\}_{i=1}^M} \{ \prod_{i=1}^M \mathbb{Q}(\tilde{\boldsymbol{z}}_{k,i} \mid \tilde{\boldsymbol{h}}_{k,i}) \}}. \tag{3.27}$$

Furthermore, let the noise covariances read

$$\boldsymbol{C} = \sigma^2 \boldsymbol{I}_N + \alpha \lim_{k \to \infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \boldsymbol{\Omega}_{\Delta \boldsymbol{h}_{k,m}}(\boldsymbol{C}, \tilde{\boldsymbol{C}}), \tag{3.28}$$

$$\tilde{\boldsymbol{C}} = \tilde{\sigma}^2 \boldsymbol{I}_N + \alpha \lim_{k \to \infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_{k,m}}(\boldsymbol{C}, \tilde{\boldsymbol{C}}), \tag{3.29}$$

where

$$\boldsymbol{\Omega}_{\Delta \boldsymbol{h}_{k,m}}(\boldsymbol{C}, \tilde{\boldsymbol{C}}) = \mathbb{E}\{ (\boldsymbol{h}_{k,m} - \langle \tilde{\boldsymbol{h}}_{k,m} \rangle_k)(\boldsymbol{h}_{k,m} - \langle \tilde{\boldsymbol{h}}_{k,m} \rangle_k)^{\mathsf{H}} \}, \tag{3.30}$$

$$\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_{k,m}}(\boldsymbol{C}, \tilde{\boldsymbol{C}}) = \mathbb{E}\{ (\tilde{\boldsymbol{h}}_{k,m} - \langle \tilde{\boldsymbol{h}}_{k,m} \rangle_k)(\tilde{\boldsymbol{h}}_{k,m} - \langle \tilde{\boldsymbol{h}}_{k,m} \rangle_k)^{\mathsf{H}} \}. \tag{3.31}$$

**Claim 1.** *Conditioned on $\{\mathcal{P}, \mathcal{S}\}$, the joint distribution of the true and postulated inputs and the GPME* (3.6) *of the multiuser system converges in probability to the joint distribution of the true and postulated inputs and the GPME* (3.27) *of the single-user system as $K = \alpha L \to \infty$ with $\alpha$ fixed.*

### 3.2.2 Non-Linear MAP Detector with CSI Mismatch

Consider the set of single-user SIMO channels

$$\boldsymbol{z}_{k,m} = \boldsymbol{h}_{k,m} x_{k,m} + \boldsymbol{w}_{k,m} \in \mathbb{C}^N, \tag{3.32}$$

where $\boldsymbol{w}_{k,m} \sim \mathsf{CN}(\boldsymbol{0};\, \boldsymbol{D}), m = 1, \ldots, M$. Let

$$\tilde{\boldsymbol{z}}_{k,m} = \tilde{\boldsymbol{h}}_{k,m} \tilde{x}_{k,m} + \tilde{\boldsymbol{w}}_{k,m} \in \mathbb{C}^N, \tag{3.33}$$

with $\tilde{\boldsymbol{w}}_{k,m} \sim \mathsf{CN}(\boldsymbol{0};\, \tilde{\boldsymbol{D}})$ be the corresponding channel model assumed by the receiver. Denote

$$\mathcal{J}_{k,m} = \{ \boldsymbol{z}_{k,m}, \mathbb{Q}(\tilde{\boldsymbol{h}}_{k,m} \mid \mathcal{I}) \}, \tag{3.34}$$

where the CSI $\mathbb{Q}(\tilde{\boldsymbol{h}}_{k,m} \mid \mathcal{I})$ is provided by the channel estimator. Let the postulated prior coincide with the true one $\mathbb{Q}(\tilde{x}_{k,t}) = \mathbb{P}(x_{k,t} = \tilde{x}_{k,t})$ (see (2.19) – (2.20)). The single-user posterior mean estimator based on (3.41) is written as

$$\begin{aligned}&\langle \cdots \rangle_{k,m} \\ &\quad = \frac{\sum_{\tilde{x}_{k,m} \in \mathcal{M}} \mathbb{Q}(\tilde{x}_{k,m}) \int \cdots \mathbb{Q}(\tilde{\boldsymbol{z}}_{k,m} = \boldsymbol{z}_{k,m} \mid \tilde{x}_{k,m}, \mathcal{J}_{k,m}) \mathrm{d}\mathbb{Q}(\tilde{\boldsymbol{h}}_{k,m} \mid \mathcal{I})}{\sum_{\tilde{x}_{k,m} \in \mathcal{M}} \mathbb{Q}(\tilde{x}_{k,m}) \int \mathbb{Q}(\tilde{\boldsymbol{z}}_{k,m} = \boldsymbol{z}_{k,m} \mid \tilde{x}_{k,m}, \mathcal{J}_{k,m}) \mathrm{d}\mathbb{Q}(\tilde{\boldsymbol{h}}_{k,m} \mid \mathcal{I})},\end{aligned} \tag{3.35}$$

for all $m = 1, \ldots, M$ and $k = 1, \ldots, K$. We write for notational convenience

$$\mathbb{E}^{\mathsf{d}}_{k,m} = \mathbb{E}\left\{ \cdots \mid \boldsymbol{h}_{k,m}, \mathcal{J}_{k,m} \right\}, \tag{3.36}$$

so that the true $\boldsymbol{D}$ and postulated $\tilde{\boldsymbol{D}}$ noise covariance matrices are given by the solutions to the coupled fixed point equations

$$\boldsymbol{D} = \sigma^2 \boldsymbol{I}_N + \alpha \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \boldsymbol{\Sigma}_{k,m}(\boldsymbol{D}, \tilde{\boldsymbol{D}}), \tag{3.37}$$

$$\tilde{\boldsymbol{D}} = \tilde{\sigma}^2 \boldsymbol{I}_N + \alpha \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \tilde{\boldsymbol{\Sigma}}_{k,m}(\boldsymbol{D}, \tilde{\boldsymbol{D}}), \tag{3.38}$$

respectively, where $\tilde{\sigma}^2 = \sigma^2$ and

$$\begin{aligned}&\boldsymbol{\Sigma}_{k,m}(\boldsymbol{D}, \tilde{\boldsymbol{D}}) \\ &\quad = \mathbb{E}^{\mathsf{d}}_{k,m} \left\{ \left( \boldsymbol{h}_{k,m} x_{k,m} - \langle \tilde{\boldsymbol{h}}_{k,m} \tilde{x}_{k,m} \rangle_{k,m} \right) \left( \boldsymbol{h}_{k,m} x_{k,m} - \langle \tilde{\boldsymbol{h}}_{k,m} \tilde{x}_{k,m} \rangle_{k,m} \right)^{\mathsf{H}} \right\},\end{aligned} \tag{3.39}$$

$$\begin{aligned}&\tilde{\boldsymbol{\Sigma}}_{k,m}(\boldsymbol{D}, \tilde{\boldsymbol{D}}) \\ &\quad = \mathbb{E}^{\mathsf{d}}_{k,m} \left\{ \left( \tilde{\boldsymbol{h}}_{k,m} \tilde{x}_{k,m} - \langle \tilde{\boldsymbol{h}}_{k,m} \tilde{x}_{k,m} \rangle_{k,m} \right) \left( \tilde{\boldsymbol{h}}_{k,m} \tilde{x}_{k,m} - \langle \tilde{\boldsymbol{h}}_{k,m} \tilde{x}_{k,m} \rangle_{k,m} \right)^{\mathsf{H}} \right\}.\end{aligned} \tag{3.40}$$

**Claim 2.** *Conditioned on $\{\mathcal{H}, \mathcal{S}\}$ and the CE output, the joint distribution of the true and postulated inputs and the GPME* (3.14) *of the multiuser system converges in probability to the joint distribution of the true and postulated inputs and the GPME* (3.35) *of the single-user system as $K = \alpha L \to \infty$ with $\alpha$ fixed.*

### 3.2.3 Linear Multiuser Detection with CSI Mismatch

Consider the set of single-user SIMO channels (3.32) in Section 3.2.2 and let

$$\tilde{\boldsymbol{z}}_{k,m} = \langle \tilde{\boldsymbol{h}}_{k,m} \rangle \tilde{x}_{k,m} + \Delta \tilde{\boldsymbol{v}}_{k,m} + \tilde{\boldsymbol{w}}_{k,m} \in \mathbb{C}^N, \qquad m = 1, \ldots, M, \tag{3.41}$$

where $\tilde{\boldsymbol{w}}_{k,m} \sim \mathsf{CN}(\boldsymbol{0};\, \tilde{\boldsymbol{D}})$, be the channel model postulated for the $k$th user at the receiver. The posterior mean estimates $\{\langle \tilde{\boldsymbol{h}}_{k,m} \rangle\}_{m=1}^{M}$ of the channel coefficients $\{\boldsymbol{h}_{k,m}\}_{m=1}^{M}$ are known and $\{\Delta\tilde{\boldsymbol{v}}_{k,m}\}_{m=1}^{M}$ is a set of independent Gaussian RVs with distributions $\mathbb{Q}(\Delta\tilde{\boldsymbol{v}}_{k,m}) = \mathsf{CN}(\boldsymbol{0};\, \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_{k,m}}), m = 1, \ldots, M$, where $\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_{k,m}}$ are the $N \times N$ block diagonals of the matrix $\tilde{\boldsymbol{\Omega}}_{\Delta \underline{\boldsymbol{V}}_{k,t}}$ introduced in Section 3.1.3. Let the single-user GPME based on (3.41) be defined as

$$\langle \cdots \rangle_{k,m} = \frac{\mathbb{E}_{\tilde{x}_{k,m}, \Delta\tilde{\boldsymbol{v}}_{k,m}} \{ \cdots \mathbb{Q}(\tilde{\boldsymbol{z}}_{k,m} = \boldsymbol{z}_{k,m} \mid \tilde{x}_{k,m}, \Delta\tilde{\boldsymbol{v}}_{k,m}, \mathcal{J}_{k,m}) \}}{\mathbb{E}_{\tilde{x}_{k,m}, \Delta\tilde{\boldsymbol{v}}_{k,m}} \{ \mathbb{Q}(\tilde{\boldsymbol{z}}_{k,m} = \boldsymbol{z}_{k,m} \mid \tilde{x}_{k,m}, \Delta\tilde{\boldsymbol{v}}_{k,m}, \mathcal{J}_{k,m}) \}}, \quad (3.42)$$

for all $m = 1, \ldots, M$, $k = 1, \ldots, K$, and denote $\tilde{\boldsymbol{v}}_{k,m} = \langle \boldsymbol{h}_{k,m} \rangle \tilde{x}_{k,m} + \Delta\tilde{\boldsymbol{v}}_{k,m}$ for notational convenience. The true and postulated noise covariance matrices are then given by the solutions to the coupled fixed point equations (3.37) – (3.38) with (3.39) – (3.39) replaced by

$$\begin{aligned} &\boldsymbol{\Sigma}_{k,m}(\boldsymbol{D}, \tilde{\boldsymbol{D}}) \\ &\quad = \mathbb{E}^{\mathsf{d}}_{k,m} \left\{ (\boldsymbol{h}_{k,m} x_{k,m} - \langle \tilde{\boldsymbol{v}}_{k,m} \rangle_{k,m}) \, (\boldsymbol{h}_{k,m} x_{k,m} - \langle \tilde{\boldsymbol{v}}_{k,m} \rangle_{k,m})^{\mathsf{H}} \right\}, \end{aligned} \quad (3.43)$$

$$\begin{aligned} &\tilde{\boldsymbol{\Sigma}}_{k,m}(\boldsymbol{D}, \tilde{\boldsymbol{D}}) \\ &\quad = \mathbb{E}^{\mathsf{d}}_{k,m} \left\{ (\tilde{\boldsymbol{v}}_{k,m} - \langle \tilde{\boldsymbol{v}}_{k,m} \rangle_{k,m}) \, (\tilde{\boldsymbol{v}}_{k,m} - \langle \tilde{\boldsymbol{v}}_{k,m} \rangle_{k,m})^{\mathsf{H}} \right\}. \end{aligned} \quad (3.44)$$

Note that if the a priori probability of the data is Gaussian

$$\tilde{\boldsymbol{x}}_k = [\tilde{x}_{k,1}, \ldots, \tilde{x}_{k,M}]^{\mathsf{T}} \sim \mathsf{CN}(\boldsymbol{0};\, \boldsymbol{I}_M), \quad (3.45)$$

the data estimator based on the GPME (3.42) reduces to $\langle x_{k,m} \rangle_{k,m} = \boldsymbol{m}_{k,m}^{\mathsf{H}} \boldsymbol{z}_{k,m}$, where

$$\boldsymbol{m}_{k,m}^{\mathsf{H}} = \frac{\langle \tilde{\boldsymbol{h}}_{k,m} \rangle^{\mathsf{H}} (\tilde{\boldsymbol{D}} + \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_{k,m}})^{-1}}{1 + \langle \tilde{\boldsymbol{h}}_{k,m} \rangle^{\mathsf{H}} (\tilde{\boldsymbol{D}} + \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_{k,m}})^{-1} \langle \tilde{\boldsymbol{h}}_{k,m} \rangle}. \quad (3.46)$$

**Claim 3.** *Conditioned on $\{\mathcal{H}, \mathcal{S}\}$ and the CE output, the joint distribution of the true and postulated inputs and the GPME* (3.20) *of the multiuser system converges in probability to the joint distribution of the true and postulated inputs and the GPME* (3.42) *of the single-user system as $K = \alpha L \to \infty$ with $\alpha$ fixed.*

## 3.3 Performance of Large MIMO DS-CDMA Systems

In this section the actual performance analysis of the multiuser MIMO DS-CDMA system described in Sections 2.2.1 and 3.1 is carried out. Due to the relatively large amount of different results that will follow, the organization of the next section is provided below:

- Section 3.3.1 considers the linear channel estimators given in Section 3.1.1. The output statistics of the following CEs in increasing order of model mismatch are obtained:
  1. Optimum pilot-aided LMMSE channel estimator;
  2. Covariance mismatched LMMSE channel estimators of Type-1 (neglects transmit correlation) and Type-2 (neglects spatial correlation completely).
  3. Linear maximum likelihood channel estimator.
- Section 3.3.2 concentrates on the analysis of the MUDs introduced in Sections 3.1.2 and 3.1.3. The performance of the following non-iterative multiuser detectors, arranged in decreasing order of complexity, are given:
  1. MAP detector;
  2. LMMSE detector;
  3. Decorrelator;
  4. Conventional detector (SUMF).

It is assumed for the rest of this chapter that the replica symmetric solutions of Claims 1 – 3 are valid, so that we can concentrate on studying the equivalent single-user systems given in the previous section. For simplicity, we assume in the following that the channels between the users are IID and $\boldsymbol{T}_k = \boldsymbol{T}\ \forall k \in \mathcal{K}$. We also drop the user index $k$ and omit the time dependence, writing with a slight abuse of notation, e.g., $x_m = x_{k,t,m}$ for some $k \in \mathcal{K}$ and $t \in \mathcal{D}$.

### 3.3.1 Linear Channel Estimation

In this section we examine the performance of the linear channel estimators described in Section 3.1.1. The first result gives the error covariance estimate obtained by the CE. We assume that this is also the information that the MUD has about the error statistics in the channel estimation.

**Proposition 1.** *Consider the linear channel estimator defined by* (3.7)*. Let* $\tilde{\boldsymbol{\Omega}}_{\underline{\boldsymbol{H}}}$ *be the postulated covariance matrix of the channel* $\underline{\boldsymbol{H}}$*. The error covariance of the channel estimates* $\langle \underline{\tilde{\boldsymbol{H}}} \rangle$*, obtained by the channel estimator and forwarded to the MUD reads then*

$$\begin{aligned}\tilde{\boldsymbol{\Omega}}_{\Delta\underline{\boldsymbol{H}}}(\tilde{\boldsymbol{C}}) &= \mathbb{E}\{(\underline{\tilde{\boldsymbol{H}}} - \langle \underline{\tilde{\boldsymbol{H}}} \rangle)(\underline{\tilde{\boldsymbol{H}}} - \langle \underline{\tilde{\boldsymbol{H}}} \rangle)^{\mathsf{H}}\} \\ &= \tilde{\boldsymbol{\Omega}}_{\underline{\boldsymbol{H}}}(\boldsymbol{I}_M \otimes \tilde{\boldsymbol{C}} + \tau_{\mathsf{tr}}\tilde{\boldsymbol{\Omega}}_{\underline{\boldsymbol{H}}})^{-1}(\boldsymbol{I}_M \otimes \tilde{\boldsymbol{C}}) \in \mathbb{C}^{MN\times MN}, \end{aligned} \tag{3.47}$$

*where* $\tilde{\boldsymbol{C}}$ *is the noise covariance* (3.29) *of the postulated single-user channel* (3.24)*. Let* $\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_m}(\tilde{\boldsymbol{C}})$ *be the estimated error covariance matrices for the transmit antennas*

*$m = 1, \ldots, M$, given by the $M \times M$ block diagonals of* (3.47)*. The matrix $\tilde{\boldsymbol{C}}$ is then given by*

$$\tilde{\boldsymbol{C}} = \tilde{\sigma}^2 \boldsymbol{I}_N + \alpha \sum_{m=1}^{M} \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_m}(\tilde{\boldsymbol{C}}). \tag{3.48}$$

*The error covariance matrices that are solutions to the coupled equations* (3.47) *and* (3.48) *will be denoted by $\tilde{\boldsymbol{\Omega}}_{\Delta \underline{\boldsymbol{H}}}$ and $\{\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_m}\}_{m=1}^{M}$ in the following.*

Proposition 1 holds for all channel estimators introduced in Section 3.1.1 and gives the postulated error covariance obtained by the CE in the form of coupled equations (3.47) and (3.48). However, in order to assess the performance of the system, we need to obtain the true error covariance $\boldsymbol{\Omega}_{\Delta \underline{\boldsymbol{H}}}$ of the channel estimates as well. For the optimum pilot-aided LMMSE channel estimator given in Example 2, the two coincide and the MUD is thus correctly informed about the error statistics in channel estimation.

**Corollary 1.** *Let $\tilde{\sigma}^2 = \sigma^2$ and $\tilde{\boldsymbol{\Omega}}_{\underline{\boldsymbol{H}}} = \boldsymbol{\Omega}_{\underline{\boldsymbol{H}}}$ in* (3.47) *and* (3.48)*. Denote $\Delta \underline{\boldsymbol{H}} = \underline{\boldsymbol{H}} - \langle \tilde{\underline{\boldsymbol{H}}} \rangle$. Then*

$$\tilde{\boldsymbol{\Omega}}_{\Delta \underline{\boldsymbol{H}}} = \boldsymbol{\Omega}_{\Delta \underline{\boldsymbol{H}}} = \mathbb{E}\{\Delta \underline{\boldsymbol{H}} \, \Delta \underline{\boldsymbol{H}}^{\mathsf{H}}\}. \tag{3.49}$$

Consider next the outputs of the two covariance mismatched LMMSE channel estimators described in Example 3. Recall that the Type-1 covariance mismatched estimator neglects the correlation between the transmit antennas, but knows the correlation between the receive antennas. Type-2 estimator neglects the spatial correlation altogether.

**Proposition 2.** *Let $\boldsymbol{R} = \boldsymbol{U} \boldsymbol{\Lambda}_{\boldsymbol{R}} \boldsymbol{U}^{\mathsf{H}}$ and $\tilde{\boldsymbol{R}} = \tilde{\boldsymbol{U}} \tilde{\boldsymbol{\Lambda}}_{\boldsymbol{R}} \tilde{\boldsymbol{U}}^{\mathsf{H}}$, where $\boldsymbol{U}, \tilde{\boldsymbol{U}} \in \mathbb{C}^{N \times N}$ are unitary matrices, and the diagonal matrices $\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{R}}$ and $\boldsymbol{\Lambda}_{\boldsymbol{R}}$ contain the eigenvalues of $\tilde{\boldsymbol{R}}$ and $\boldsymbol{R}$, respectively. For the Type-1 and Type-2 covariance mismatched channel estimators in Example 3, the noise and error covariance matrices $\tilde{\boldsymbol{C}} = \tilde{\boldsymbol{U}} \tilde{\boldsymbol{\Lambda}}_{\boldsymbol{C}} \tilde{\boldsymbol{U}}^{\mathsf{H}}$ and $\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_m} = \tilde{\boldsymbol{U}} \tilde{\boldsymbol{\Lambda}}_{\Delta \boldsymbol{h}_m} \tilde{\boldsymbol{U}}^{\mathsf{H}}$ are given by the solutions to the coupled equations*

$$\tilde{\boldsymbol{\Lambda}}_{\Delta \boldsymbol{h}_m}(\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{C}}) = T_{m,m} \tilde{\boldsymbol{\Lambda}}_{\boldsymbol{R}} \left( \tilde{\boldsymbol{\Lambda}}_{\boldsymbol{C}} + \tau_{\mathsf{tr}} T_{m,m} \tilde{\boldsymbol{\Lambda}}_{\boldsymbol{R}} \right)^{-1} \tilde{\boldsymbol{\Lambda}}_{\boldsymbol{C}}, \tag{3.50}$$

$$\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{C}} = \tilde{\sigma}^2 \boldsymbol{I}_N + \alpha \sum_{m=1}^{M} \tilde{\boldsymbol{\Lambda}}_{\Delta \boldsymbol{h}_m}(\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{C}}). \tag{3.51}$$

*For the Type-1 mismatched estimator, if we let $\Delta \boldsymbol{h}_m = \boldsymbol{h}_m - \langle \tilde{\boldsymbol{h}}_m \rangle$ then*

$$\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_m} = \boldsymbol{\Omega}_{\Delta \boldsymbol{h}_m} = \mathbb{E}\{\Delta \boldsymbol{h}_m \Delta \boldsymbol{h}_m^{\mathsf{H}}\}. \tag{3.52}$$

*For the Type-2 mismatched estimator the postulated error and noise covariance matrices can be written as* $\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_m} = \tilde{\lambda}_{\Delta \boldsymbol{h}_m} \boldsymbol{I}_N$ *and* $\tilde{\boldsymbol{C}} = \tilde{\lambda}_{\boldsymbol{C}} \boldsymbol{I}_N$*, respectively, where* $\tilde{\lambda}_{\Delta \boldsymbol{h}_m}$ *and* $\tilde{\lambda}_{\boldsymbol{C}}$ *are the solutions to the equations*

$$\tilde{\lambda}_{\Delta \boldsymbol{h}_m}(\tilde{\lambda}_{\boldsymbol{C}}) = \frac{T_{m,m} \tilde{\lambda}_{\boldsymbol{C}}}{\tilde{\lambda}_{\boldsymbol{C}} + \tau_{\mathsf{tr}} T_{m,m}}, \tag{3.53}$$

$$\tilde{\lambda}_{\boldsymbol{C}} = \tilde{\sigma}^2 + \alpha \sum_{m=1}^{M} \tilde{\lambda}_{\Delta \boldsymbol{h}_m}(\tilde{\lambda}_{\boldsymbol{C}}). \tag{3.54}$$

*The true error covariance for the Type-2 mismatched estimator is given by* $\boldsymbol{\Omega}_{\Delta \boldsymbol{h}_m} = \boldsymbol{U} \boldsymbol{\Lambda}_{\Delta \boldsymbol{h}_m} \boldsymbol{U}^{\mathsf{H}}$*, where*

$$\boldsymbol{\Lambda}_{\Delta \boldsymbol{h}_m} = \frac{T_{m,m}}{(\tilde{\lambda}_{\boldsymbol{C}} + \tau_{\mathsf{tr}} T_{m,m})^2} (\tilde{\lambda}_{\boldsymbol{C}}^2 \boldsymbol{\Lambda}_{\boldsymbol{R}} + \tau_{\mathsf{tr}} T_{m,m} \boldsymbol{\Lambda}_{\boldsymbol{C}}), \tag{3.55}$$

$$\boldsymbol{\Lambda}_{\boldsymbol{C}} = \tilde{\lambda}_{\boldsymbol{C}} \left( \sigma^2 \boldsymbol{I}_N + \alpha \boldsymbol{\Lambda}_{\boldsymbol{R}} \sum_{m=1}^{M} \frac{\tilde{\lambda}_{\Delta \boldsymbol{h}_m}^2}{T_{m,m}} \right) \Bigg/ \left( \tilde{\sigma}^2 + \alpha \sum_{m=1}^{M} \frac{\tilde{\lambda}_{\Delta \boldsymbol{h}_m}^2}{T_{m,m}} \right). \tag{3.56}$$

Note that (3.55) – (3.56) follows from the fact that the Type-2 mismatched LMMSE estimator postulates $\tilde{\boldsymbol{R}} = \boldsymbol{I}_N$ and, thus, $\boldsymbol{U}$ simultaneously diagonalizes $\boldsymbol{R}$ and $\tilde{\boldsymbol{R}}$. The above result also states that if the LMMSE channel estimator has correct information about the transmitted powers and the spatial correlation on the receiver's side, the error statistics provided to the MUD are correct. Postulating uncorrelated antennas at the transmitter does not cause a mismatch in error statistics. The actual MSE $\boldsymbol{\Omega}_{\Delta \underline{\boldsymbol{H}}}$ obtained by the estimator described in Examples 2 and the Type-1 mismatched estimator of Example 3 are, however, different unless $\boldsymbol{T}$ is diagonal. The following simple example illustrates the effect of transmit correlation on the accuracy of channel estimation.

**Example 7.** Let us consider for simplicity uncorrelated receive antennas $\boldsymbol{R} = \boldsymbol{R} = \boldsymbol{U} = \tilde{\boldsymbol{U}} = \boldsymbol{I}_N$, and equal power transmission $T_{m,m} = \frac{1}{M}$. Let the transmit correlation be modelled as

$$\boldsymbol{T} = \frac{1-\rho}{M} \boldsymbol{I}_M + \frac{\rho}{M} \boldsymbol{e}_M \boldsymbol{e}_M^{\mathsf{T}}, \qquad 0 \leq \rho \leq 1, \tag{3.57}$$

i.e., by the constant correlation model (see, for example, [184, 185]). Using the matrix determinant lemma [186, Theorem 13.3.8], the eigenvalues of $\boldsymbol{T}$ are easily obtained as

$$\lambda_1 = \frac{1}{M}[1 + \rho(M-1)], \tag{3.58}$$

$$\lambda_2 = \frac{1}{M}(1-\rho), \tag{3.59}$$

where $\lambda_2$ has the multiplicity of $M-1$. The noise covariances are given by $\boldsymbol{C} = \tilde{\boldsymbol{C}} = C\boldsymbol{I}_N$, where

$$C = \sigma^2 + \alpha M \Omega_{\Delta \boldsymbol{h}_m}(C), \tag{3.60}$$

is a fixed point equation with

$$\Omega_{\Delta \boldsymbol{h}_m}(C) = \frac{1}{M}\left[\lambda_{\Delta \boldsymbol{h}_1}(C) + (M-1)\lambda_{\Delta \boldsymbol{h}_2}(C)\right], \tag{3.61}$$

$$\lambda_{\Delta \boldsymbol{h}_i}(C) = \frac{C\lambda_i}{C + \tau_{\mathsf{tr}}\lambda_i}, \qquad i = 1, 2. \tag{3.62}$$

Let $M, C, \tau_{\mathsf{tr}}$ be arbitrary and fixed. It is easy to verify that

$$\frac{\partial}{\partial \rho}\Omega_{\Delta \boldsymbol{h}_m}(C) < 0, \qquad \frac{\partial^2}{\partial \rho^2}\Omega_{\Delta \boldsymbol{h}_m}(C) < 0, \qquad 0 < \rho < 1, \tag{3.63}$$

so that for fixed $M, C, \tau_{\mathsf{tr}}$, the MSE is a decreasing concave function of transmit correlation $\rho$. Since $C$ decreases with $\Omega_{\Delta \boldsymbol{h}_m}(C)$, we know that increasing correlation between the transmit antennas helps the channel estimator to obtain lower MSEs.

The two extreme cases of transmit correlation for this model are obtained by setting $\rho = 0$ (uncorrelated transmit antennas) and $\rho = 1$ (fully correlated transmit antennas), which yields

$$\rho = 0: \quad \Omega_{\Delta \boldsymbol{h}_m}(C) = \frac{1}{M}\frac{C}{C + \dfrac{\tau_{\mathsf{tr}}}{M}}, \tag{3.64}$$

$$\rho = 1: \quad \Omega_{\Delta \boldsymbol{h}_m}(C) = \frac{1}{M}\frac{C}{C + \tau_{\mathsf{tr}}}. \tag{3.65}$$

In the limit of extremely correlated transmit antennas, we thus obseve that the effective number of training symbols is increased by a factor of $M$. This is an intuitively pleasing result since there is only one physical channel per receive antenna to estimate, but the estimator still receives $\tau_{\mathsf{tr}}M$ training sequences per receive antenna. Such a simple interpretation, however, cannot be made for $0 < \rho < 1$ or if the receive antennas are correlated as well. ♢

The Type-2 mismatched channel estimator gives the MUD incorrect information about the estimation errors, unless $\tilde{\boldsymbol{R}} = \boldsymbol{R} = \boldsymbol{I}_N$. This affects the analysis of the MUD and makes it in general quite cumbersome. Another property of this estimator is that the estimation errors are correlated with the channel estimates, as discussed below.

**Remark 2.** Let the cross covariance between the channel estimate and the estimation error be $\boldsymbol{\Omega}_{\Delta \boldsymbol{h}_m, \langle \tilde{\boldsymbol{h}}_m \rangle} = \mathbb{E}\{\Delta \boldsymbol{h}_m \langle \tilde{\boldsymbol{h}}_m \rangle^{\mathsf{H}}\}$. In the large system limit, the channel estimates provided by the estimators defined in Examples 2 and 3 have unconditional complex Gaussian distribution $\langle \tilde{\boldsymbol{h}}_m \rangle \sim \mathsf{CN}(\boldsymbol{0}; \boldsymbol{\Omega}_{\langle \tilde{\boldsymbol{h}}_m \rangle})$, where $\boldsymbol{\Omega}_{\langle \tilde{\boldsymbol{h}}_m \rangle} = \boldsymbol{U} \boldsymbol{\Lambda}_{\langle \tilde{\boldsymbol{h}}_m \rangle} \boldsymbol{U}^{\mathsf{H}}$ and

$$\boldsymbol{\Lambda}_{\langle \tilde{\boldsymbol{h}}_m \rangle} = T_{m,m} \boldsymbol{\Lambda}_{\boldsymbol{R}} - 2 \boldsymbol{\Lambda}_{\Delta \boldsymbol{h}_m, \langle \tilde{\boldsymbol{h}}_m \rangle} - \boldsymbol{\Lambda}_{\Delta \boldsymbol{h}_m} \tag{3.66}$$

A little bit of algebra reveals that we have for the channel estimators in Examples 2 and 3 the following eigenvalues for the cross-covariance matrices;

- Optimum LMMSE (Example 2) and Type-1 mismatched (Example 3):

$$\boldsymbol{\Lambda}_{\Delta \boldsymbol{h}_m, \langle \tilde{\boldsymbol{h}}_m \rangle} = \boldsymbol{0}; \tag{3.67}$$

- Type-2 mismatched (Example 3):

$$\boldsymbol{\Lambda}_{\Delta \boldsymbol{h}_m, \langle \tilde{\boldsymbol{h}}_m \rangle} = \frac{\tau_{\mathsf{tr}} T_{m,m}^2}{\left(\tilde{\lambda}_{\boldsymbol{C}} + \tau_{\mathsf{tr}} T_{m,m}\right)^2} (\tilde{\lambda}_{\boldsymbol{C}} \boldsymbol{\Lambda}_{\boldsymbol{R}} - \boldsymbol{\Lambda}_{\boldsymbol{C}}). \tag{3.68}$$

For the Type-2 mismatched channel estimator the channel estimate and the error are therefore correlated unless $\boldsymbol{R} = \boldsymbol{I}_N$. One should not, however, confuse the Type-2 covariance mismatched LMMSE estimator to the linear ML channel estimator in Example 4, for which the error is uncorrelated with the channel coefficients. In fact, we immediately get from (3.55) and (3.68) that

$$\mathbb{E}\{\Delta \boldsymbol{h}_m \boldsymbol{h}_m^{\mathsf{H}}\} = \frac{\tilde{\lambda}_{\boldsymbol{C}} T_{m,m}}{\tilde{\lambda}_{\boldsymbol{C}} + \tau_{\mathsf{tr}} T_{m,m}} \boldsymbol{R}, \tag{3.69}$$

and, thus, the estimation error is correlated also with the channel coefficients for the Type-2 mismatched LMMSE channel estimator. ♢

Finally, let us consider the ML channel estimator described in Example 4. This channel estimator neglects both the spatial correlation and the additive noise in the channel. For simplicity, we consider only the case $\tau_{\mathsf{tr}} > \alpha M$ for this CE.

**Proposition 3.** *For the ML channel estimator in Example 4,* $\mathbb{E}\{\Delta \boldsymbol{h}_m \boldsymbol{h}_m^{\mathsf{H}}\} = \boldsymbol{0}$, $\boldsymbol{C} = C \boldsymbol{I}_N$ *and*

$$C = \frac{\sigma^2}{1 - \alpha M / \tau_{\mathsf{tr}}}, \tag{3.70}$$

$$\boldsymbol{\Omega}_{\Delta \boldsymbol{h}_m} = -\boldsymbol{\Omega}_{\Delta \boldsymbol{h}_m, \langle \tilde{\boldsymbol{h}}_m \rangle} = \frac{1}{\tau_{\mathsf{tr}}} \boldsymbol{C} = \frac{\sigma^2}{\tau_{\mathsf{tr}} - \alpha M} \boldsymbol{I}_N, \tag{3.71}$$

*whenever* $\tau_{\mathsf{tr}} > \alpha M$.

### 3.3.2 Multiuser Detection with Mismatched CSI

Let us first consider the performance of the non-linear MAP detector described in Section 3.1.2. For simplicity, the CSI is assumed to be given by the LMMSE channel estimator given in Example 2 and, thus, the statistics of the channel estimation error are correct. The next result reports the SINR of the "hidden" Gaussian channel (see, e.g., [85, Eq. (183)], [89, Sec. 1.3.1] and the discussion therein), at the output of the non-linear multiuser detector.

**Proposition 4.** *Let $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Lambda}_{\boldsymbol{R}}\boldsymbol{U}^{\mathsf{H}}$, where $\boldsymbol{U}$ is a unitary matrix containing the eigenvectors of $\boldsymbol{R}$. Consider the GPME given by* (3.13) – (3.14)*, and let the channel information be provided by the LMMSE channel estimator of Example 2. Then $\tilde{\boldsymbol{D}} = \boldsymbol{D} = \boldsymbol{U}\boldsymbol{\Lambda}_{\boldsymbol{D}}\boldsymbol{U}^{\mathsf{H}}$ in* (3.37) – (3.38)*. Furthermore, let $\boldsymbol{g} \in \mathbb{R}^N$ be a RV with IID elements $\{g_j\}_{j=1}^N$ drawn according to the exponential distribution $\mathbb{P}(g_j) = 1 - e^{-g_j}, g_j > 0$. Then,*

$$\mathsf{sinr}_m^{\mathsf{mmse}}(\boldsymbol{g}, \boldsymbol{\Lambda}_{\boldsymbol{D}}) = \sum_{n=1}^{N} \frac{g_n \lambda_{\langle \tilde{\boldsymbol{h}}_m \rangle}^{(n)}}{\lambda_{\boldsymbol{D}}^{(n)} + \lambda_{\Delta \boldsymbol{h}_m}^{(n)}}, \tag{3.72}$$

*is the output SINR of the equivalent "hidden" Gaussian channel where the elements of the diagonal matrix $\boldsymbol{\Lambda}_{\boldsymbol{D}} = \mathrm{diag}([\lambda_{\boldsymbol{D}}^{(1)} \cdots \lambda_{\boldsymbol{D}}^{(N)})$ are the solution to the fixed point equations*

$$\lambda_{\boldsymbol{D}}^{(n)} = \sigma^2 + \alpha \sum_{m=1}^{M} \frac{\lambda_{\boldsymbol{D}}^{(n)} \lambda_{\Delta \boldsymbol{v}_m}^{(n)}}{\lambda_{\boldsymbol{D}}^{(n)} + \lambda_{\Delta \boldsymbol{v}_m}^{(n)}} + \left( \frac{\lambda_{\boldsymbol{D}}^{(n)}}{\lambda_{\boldsymbol{D}}^{(n)} + \lambda_{\Delta \boldsymbol{v}_m}^{(n)}} \right)^2 \int_0^{\infty} g_n \lambda_{\langle \tilde{\boldsymbol{h}}_m \rangle}^{(n)}$$
$$\times \left[ 1 - \int_{-\infty}^{\infty} \tanh\left( \mathsf{sinr}_m^{\mathsf{mmse}}(\boldsymbol{g}, \boldsymbol{\Lambda}_{\boldsymbol{D}}) + \nu \sqrt{\mathsf{sinr}_m^{\mathsf{mmse}}(\boldsymbol{g}, \boldsymbol{\Lambda}_{\boldsymbol{D}})} \right) \mathrm{D}\nu \right] \prod_{j=1}^{N} e^{-g_j} \mathrm{d}g_j, \tag{3.73}$$

*and $\mathrm{D}\nu$ is defined in* (2.44)*.*

By Proposition 4, the SINR of the hidden Gaussian channel related to the $m$th transmit antenna after MAP detector has the same distribution as the received SNR of the single-user system

$$\boldsymbol{z}_m = \langle \tilde{\boldsymbol{h}}_m \rangle x_m + \boldsymbol{\zeta}_m \in \mathbb{C}^N, \tag{3.74}$$

where $\boldsymbol{\zeta}_m \sim \mathsf{CN}(\boldsymbol{0};\, \boldsymbol{D} + \boldsymbol{\Omega}_{\Delta \boldsymbol{h}_m})$, $\boldsymbol{D} = \boldsymbol{U}\boldsymbol{\Lambda}_{\boldsymbol{D}}\boldsymbol{U}^{\mathsf{H}}$ and the channel coefficients $\langle \tilde{\boldsymbol{h}}_m \rangle \sim \mathsf{CN}(\boldsymbol{0};\, \boldsymbol{\Omega}_{\langle \tilde{\boldsymbol{h}}_m \rangle})$ are perfectly known at the receiver. For a block fading

channel and sufficiently long code words, each transmit antenna encounters thus an ergodic SIMO channel in time domain. If perfect CSI is available at the receiver, i.e., $\boldsymbol{\Omega}_{\Delta \boldsymbol{h}_m} = \mathbf{0}$, we immediately see that transmit correlation has no effect on the ergodic spectral efficiency of the system, as expected from the earlier results reported in [92, 154]. With estimated channel, however, the covariance of the channel $\boldsymbol{\Omega}_{\langle \tilde{\boldsymbol{h}}_m \rangle}$ for the optimum pilot-assisted LMMSE channel estimator depends on the transmit correlation through (3.47) (see also Example 7). Furthermore, as noted in Remark 2, $\boldsymbol{\Omega}_{\Delta \boldsymbol{h}_m} = T_{m,m}\boldsymbol{R} - \boldsymbol{\Omega}_{\langle \tilde{\boldsymbol{h}}_m \rangle}$ and, thus, the covariance of the noise $\boldsymbol{\zeta}_m$ in (3.74) depends on transmit correlation as well.

Next the noise covariance matrices of the decoupled single-user channels (3.32) – (3.41) related to the linear MUDs discussed in Section 3.1.3 are given.

**Proposition 5.** *Consider the case where the CSI be provided by one of the LMMSE channel estimators described in Examples 2 – 4. Let $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Lambda}_{\boldsymbol{R}}\boldsymbol{U}^{\mathsf{H}}$ as in Proposition 2. The noise covariance of the postulated channel* (3.41) *can then be written as $\tilde{\boldsymbol{D}} = \boldsymbol{U}\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{D}}\boldsymbol{U}^{\mathsf{H}}$, where $\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{D}} = \operatorname{diag}([\tilde{\lambda}_{\boldsymbol{D}}^{(1)} \cdots \tilde{\lambda}_{\boldsymbol{D}}^{(N)}])$. The eigenvalues in $\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{D}}$ are the solutions to the fixed point equation*

$$\begin{aligned}\tilde{\lambda}_{\boldsymbol{D}}^{(n)} = \tilde{\sigma}^2 &+ \alpha \sum_{m=1}^{M} \frac{\tilde{\lambda}_{\boldsymbol{D}}^{(n)} \tilde{\lambda}_{\Delta \boldsymbol{v}_m}^{(n)}}{\tilde{\lambda}_{\boldsymbol{D}}^{(n)} + \tilde{\lambda}_{\Delta \boldsymbol{v}_m}^{(n)}} \\ &+ \left( \frac{\tilde{\lambda}_{\boldsymbol{D}}^{(n)}}{\tilde{\lambda}_{\boldsymbol{D}}^{(n)} + \tilde{\lambda}_{\Delta \boldsymbol{v}_m}^{(n)}} \right)^2 \int_0^\infty \frac{g_n \lambda_{\langle \tilde{\boldsymbol{h}}_m \rangle}^{(n)}}{1 + \sum_{j=1}^{N} \frac{g_j \lambda_{\langle \tilde{\boldsymbol{h}}_m \rangle}^{(j)}}{\tilde{\lambda}_{\boldsymbol{D}}^{(j)} + \tilde{\lambda}_{\Delta \boldsymbol{v}_m}^{(j)}}} \prod_{j=1}^{N} e^{-g_j} \mathrm{d}g_j, \end{aligned} \tag{3.75}$$

*where $\tilde{\boldsymbol{\Lambda}}_{\Delta \boldsymbol{v}_m} = \operatorname{diag}([\tilde{\lambda}_{\Delta \boldsymbol{v}_m}^{(1)} \cdots \tilde{\lambda}_{\Delta \boldsymbol{v}_m}^{(N)}])$ are the eigenvalues of the postulated error covariance matrix $\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_m} = \boldsymbol{U}\tilde{\boldsymbol{\Lambda}}_{\Delta \boldsymbol{v}_m}\boldsymbol{U}^{\mathsf{H}}$. Given $\tilde{\boldsymbol{D}}$, the noise covariance of the decoupled channel* (3.32) *can then be solved from*

$$\begin{aligned}\boldsymbol{D} = \Bigg[ &\sigma^2 \boldsymbol{I} + \alpha \sum_{m=1}^{M} \tilde{\boldsymbol{D}} (\tilde{\boldsymbol{D}} + \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_m})^{-1} \\ &\times \mathbb{E}\{ (\boldsymbol{I} - \langle \tilde{\boldsymbol{h}}_m \rangle \boldsymbol{m}_m^{\mathsf{H}}) (\boldsymbol{h}_m \boldsymbol{h}_m^{\mathsf{H}}) (\boldsymbol{I} - \boldsymbol{m}_m \langle \tilde{\boldsymbol{h}}_m \rangle^{\mathsf{H}}) \} (\tilde{\boldsymbol{D}} + \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_m})^{-1} \tilde{\boldsymbol{D}} \Bigg] \\ \times \Bigg[ &\tilde{\sigma}^2 \boldsymbol{I} + \alpha \sum_{m=1}^{M} \tilde{\boldsymbol{D}} (\tilde{\boldsymbol{D}} + \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_m})^{-1} \\ &\times \mathbb{E}\{ (\boldsymbol{I} - \langle \tilde{\boldsymbol{h}}_m \rangle \boldsymbol{m}_m^{\mathsf{H}}) (\langle \tilde{\boldsymbol{h}}_m \rangle \langle \tilde{\boldsymbol{h}}_m \rangle^{\mathsf{H}} + \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_m}) (\boldsymbol{I} - \boldsymbol{m}_m \langle \tilde{\boldsymbol{h}}_m \rangle^{\mathsf{H}}) \} \\ &\times (\tilde{\boldsymbol{D}} + \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_m})^{-1} \tilde{\boldsymbol{D}} \Bigg]^{-1} \tilde{\boldsymbol{D}}. \end{aligned} \tag{3.76}$$

The general form of Proposition 5 is quite difficult to work with. Luckily, some simple special cases can be obtained. The first examples is the LMMSE MUD of Example 5, given the error statistics from the CE are correct.

**Corollary 2.** *Let the CSI be provided by the LMMSE estimator of Example 2, or the Type-1 mismatched LMMSE estimator of Example 3, so that* $\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_m} = \boldsymbol{\Omega}_{\Delta \boldsymbol{h}_m}$ *and* $\boldsymbol{\Omega}_{\Delta \boldsymbol{h}_m, \langle \tilde{\boldsymbol{h}}_m \rangle} = \mathbf{0}$*. Then, for the LMMSE MUD of Example 5,* $\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_m} = \boldsymbol{\Omega}_{\Delta \boldsymbol{h}_m}$*,* $\tilde{\boldsymbol{D}} = \boldsymbol{D}$*, and the post-detection SINR for the* $m$*th transmit antenna is in* (3.72)*.*

If the CSI is provided by the LMMSE channel estimator of Example 2, or the Type-1 covariance mismatched estimator of Example 3, Corollary 2 tells us that for the non-iterative LMMSE MUD of Example 5 one needs to solve only the fixed point equation (3.75) to get the statistics of the decoupled channels (3.32) – (3.41). For the Type-2 mismatched LMMSE CE, however, $\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_m} \neq \boldsymbol{\Omega}_{\Delta \boldsymbol{h}_m}$ and, thus, $\tilde{\boldsymbol{D}} \neq \boldsymbol{D}$ in general. In this case one needs to first solve (3.75), substitute the solutions to (3.76), and solve it for $\boldsymbol{D}$ while taking into account the notes made in Remark 2.

Recall that the SUMF and decorrelator described in Example 6 discard all statistical information about the channel estimation errors. This simplifies the task of solving (3.76) and we can obtain $\boldsymbol{D}$ in closed form. However, one should remember that the correlation between the channel estimates and the estimation errors is non-zero for the Type-2 mismatched and ML channel estimators. In this thesis, the Type-2 channel estimator is considered only with the SUMF for simplicity.

**Proposition 6.** *Consider the single-user matched filter and decorrelator in Example 6. For the SUMF*

$$\boldsymbol{D} = \sigma^2 \boldsymbol{I}_N + \alpha \bar{t} \boldsymbol{R}, \tag{3.77}$$

*where* $\bar{t} = \mathrm{tr}(\boldsymbol{T})$*, as defined in Section 2.2.1. For the decorrelator, with optimal or Type-1 channel estimator, equal transmit power* $T_{m,m} = \bar{t}/M$ *and uncorrelated receive antennas* $\boldsymbol{R} = \boldsymbol{I}_N$*,* $\boldsymbol{D}$ *is given by*[3]

$$\boldsymbol{D} = \begin{cases} N \dfrac{\sigma^2 + \alpha M \lambda_{\Delta \boldsymbol{h}_m}}{N - \alpha M} \boldsymbol{I}_N, & N > \alpha M, \\ \dfrac{\sigma^2 \alpha M + \lambda_{\langle \tilde{\boldsymbol{h}}_m \rangle} (\alpha M - N)^2 + \lambda_{\Delta \boldsymbol{h}_m} [N + \alpha M (\alpha M - 2)]}{\alpha M - N} \boldsymbol{I}_N, & N < \alpha M. \end{cases} \tag{3.78}$$

[3] Obtaining a closed form solution for matrix $\boldsymbol{D}$ in the general case is difficult and not considered in this thesis.

**Remark 3.** For the matched filter, the noise covariance (3.77) does not depend on channel estimation. The same can be observed in [59, Eq. (15)] for the case of DS-CDMA. For the decorrelator, (3.78) generalizes the previous results [85, 187] to the case of mismatched CSI and multiple antennas. Note that to obtain (3.78), we restricted all transmit antennas to have the same nominal power, but the spatial correlation at the transmitter could still be arbitrary. As with the other MUDs in this chapter, the transmit correlation manifests through the estimation errors and does not affect the equivalent noise covariance of the single-user channel (3.32). ♢

From Propositions 1 – 6 we get the QPSK constrained capacity of correlated MIMO DS-CDMA system using linear channel estimation, linear MUD and separate decoding.

**Proposition 7.** *The per-antenna ergodic spectral efficiency* $\mathsf{C}_m^{\mathsf{qpsk}}$ *(bits) for* $m$*th antenna and all MUDs in this section is given under separate decoding and QPSK signaling by*

$$\mathsf{C}_m^{\mathsf{qpsk}} = \log_2|\mathcal{M}| - N\log_2 e - \frac{1}{|\mathcal{M}|}\sum_{x\in\mathcal{M}} \mathbb{E}\Big\{ \log_2 \sum_{\tilde{x}\in\mathcal{M}} \exp\left(-\hat{\boldsymbol{\mu}}^{\mathsf{H}}\hat{\boldsymbol{\Omega}}^{-1}\hat{\boldsymbol{\mu}}\right) \mid x, \tilde{x}\Big\}, \quad (3.79)$$

*where we denoted for notational convenience*

$$\hat{\boldsymbol{\mu}} = \boldsymbol{\zeta}_m + \langle\tilde{\boldsymbol{h}}_m\rangle(x-\tilde{x}) - \boldsymbol{\Omega}_{\Delta\boldsymbol{h}_m,\langle\tilde{\boldsymbol{h}}_m\rangle}\boldsymbol{\Omega}^{-1}_{\langle\tilde{\boldsymbol{h}}_m\rangle}\langle\tilde{\boldsymbol{h}}_m\rangle\tilde{x}, \quad (3.80)$$

$$\hat{\boldsymbol{\Omega}} = \boldsymbol{D} + \boldsymbol{\Omega}_{\Delta\boldsymbol{h}_m} - \boldsymbol{\Omega}_{\Delta\boldsymbol{h}_m,\langle\tilde{\boldsymbol{h}}_m\rangle}\boldsymbol{\Omega}^{-1}_{\langle\tilde{\boldsymbol{h}}_m\rangle}\boldsymbol{\Omega}^{\mathsf{H}}_{\Delta\boldsymbol{h}_m,\langle\tilde{\boldsymbol{h}}_m\rangle}, \quad (3.81)$$

*and*

$$\begin{bmatrix}\boldsymbol{\zeta}_m \\ \langle\tilde{\boldsymbol{h}}_m\rangle\end{bmatrix} \Bigg| \, x \sim \mathsf{CN}\left(\boldsymbol{0}; \begin{bmatrix}\boldsymbol{\Omega}_{\Delta\boldsymbol{h}_m}+\boldsymbol{D} & \boldsymbol{\Omega}_{\Delta\boldsymbol{h}_m,\langle\tilde{\boldsymbol{h}}_m\rangle}x \\ \boldsymbol{\Omega}^{\mathsf{H}}_{\Delta\boldsymbol{h}_m,\langle\tilde{\boldsymbol{h}}_m\rangle}x^* & \boldsymbol{\Omega}_{\langle\tilde{\boldsymbol{h}}_m\rangle}\end{bmatrix}\right). \quad (3.82)$$

From Remark 2, we get the simplified version of Proposition 7, applicable to the optimum pilot-aided LMMSE channel estimator given in Example 2, and the Type-1 covariance mismatch LMMSE estimator described in Example 3. When the CSI is provided by either of these channel estimators, the per-antenna ergodic capacity $\mathsf{C}_m^{\mathsf{qpsk}}$ (bits) for the $m$th antenna and all MUDs in this section is given under separate decoding and QPSK signaling by (3.79) with (3.80) and (3.81) replaced by

$$\hat{\boldsymbol{\mu}} = \boldsymbol{\zeta}_m + \langle\tilde{\boldsymbol{h}}_m\rangle(x-\tilde{x}), \quad (3.83)$$

$$\hat{\boldsymbol{\Omega}} = \boldsymbol{D} + \boldsymbol{\Omega}_{\Delta\boldsymbol{h}_m}, \quad (3.84)$$

respectively. Furthermore, $\boldsymbol{\zeta}_m \sim \mathsf{CN}(\mathbf{0};\, \boldsymbol{\Omega}_{\Delta \boldsymbol{h}_m} + \boldsymbol{D})$ and $\langle \tilde{\boldsymbol{h}}_m \rangle \sim \mathsf{CN}(\mathbf{0};\, \boldsymbol{\Omega}_{\langle \tilde{\boldsymbol{h}}_m \rangle})$ are then independent RVs. Comparing this to (2.12) in Section 2.3.1, we note that this is just the capacity of a SIMO channel with QPSK inputs, zero-mean proper complex Gaussian additive noise with variance $\boldsymbol{\Omega}_{\Delta \boldsymbol{h}_m} + \boldsymbol{D}$ and Rayleigh fading channel, perfectly known at the receiver and with covariance $\boldsymbol{\Omega}_{\langle \tilde{\boldsymbol{h}}_m \rangle}$.

## 3.4 Numerical Examples and Discussion

Selected numerical examples based on the large system analysis carried out in Sections 3.2 and 3.3 are given below. The main emphasis is on the linear MUDs since they provide more practical means for data detection in MIMO systems than the computationally rather complex non-linear MAP-MUD. The latter is, however, considered briefly as a benchmark to the performance of the linear detectors. The reader is reminded that for all cases we let $K = \alpha L \to \infty$, while keeping the channel load $\alpha$ and the number of antennas at both ends constant.

Let the spectral efficiency for the considered system be defined as

$$\mathsf{C}^{\mathsf{qpsk}} = \alpha \left(1 - \frac{\tau_{\mathsf{tr}}}{T_{\mathsf{coh}}}\right) \sum_{m=1}^{M} \mathsf{C}_m^{\mathsf{qpsk}}, \tag{3.85}$$

where $\mathsf{C}_m^{\mathsf{qpsk}}$ is given in Proposition 7. For simplicity, we assume equal transmit power per antenna and uniform linear arrays at both ends of transmission. The spatial correlation is modelled as in [100], i.e., the elements of the covariance matrices $\boldsymbol{T} = [T_{m,i}] \in \mathbb{R}^{M \times M}$ and $\boldsymbol{R} = [R_{n,j}] \in \mathbb{R}^{N \times N}$ are given by

$$T_{m,i} = \frac{\bar{t}}{M} \int_{-180}^{180} \frac{1}{\sqrt{2\pi\delta_{\mathsf{tx}}}} \exp\left[2\pi \mathrm{j}(m-i)d_\lambda \sin\left(\frac{\varphi\pi}{180}\right) - \frac{\varphi^2}{2\delta_{\mathsf{tx}}^2}\right] \mathrm{d}\varphi, \tag{3.86}$$

$$R_{n,j} = \int_{-180}^{180} \frac{1}{\sqrt{2\pi\delta_{\mathsf{rx}}}} \exp\left[2\pi \mathrm{j}(n-j)d_\lambda \sin\left(\frac{\varphi\pi}{180}\right) - \frac{\varphi^2}{2\delta_{\mathsf{rx}}^2}\right] \mathrm{d}\varphi, \tag{3.87}$$

where $\delta_{\mathsf{tx}}$ and $\delta_{\mathsf{rx}}$ are the angular spread at the transmitter and the receiver side, respectively, given in degrees. In the following we let the nearest neighbor antenna spacing be $d_\lambda = 1$ (wavelengths) at both the transmitter and the receiver. Therefore, angular spread is the only free parameter that determines the spatial correlation in the following discussion.

In Figure 3.1, the normalized MSEs for the channel estimators introduced in Examples 2 – 4 are given as a function of transmit correlation. The $N = 4$ receive

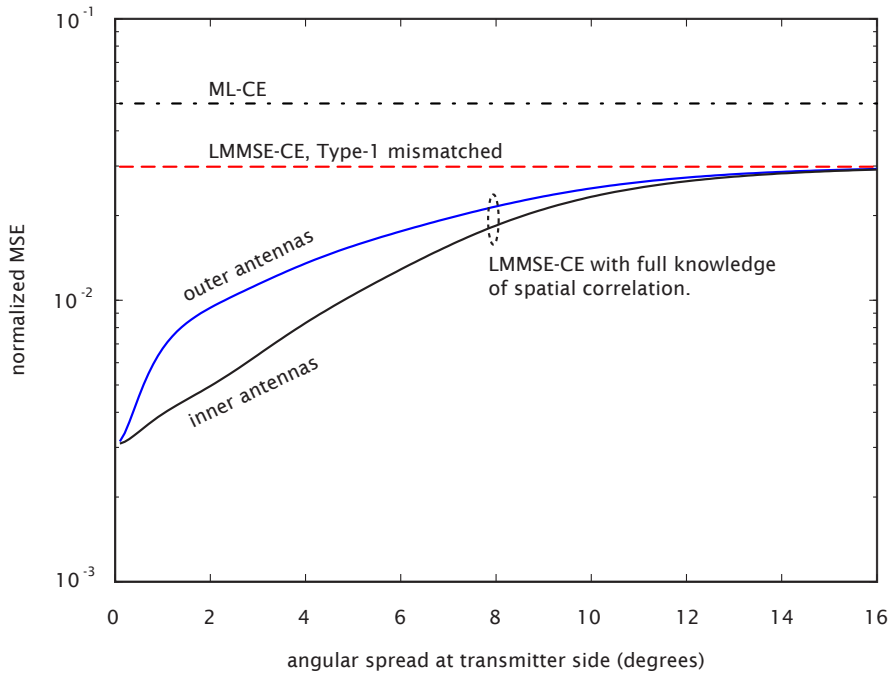



**Figure 3.1.** Normalized MSE vs. angular spread at the transmitter $\delta_{\mathsf{tx}} \in [0.1, 16]$. Uncorrelated receive antennas, equal transmit power per antenna, $4 \times 4$ MIMO channel. User load $\alpha = 2$, number of pilots per fading block $\tau_{\mathsf{tr}} = 10$ and average SNR per receive antenna $\overline{\mathsf{snr}} = 10$ dB.

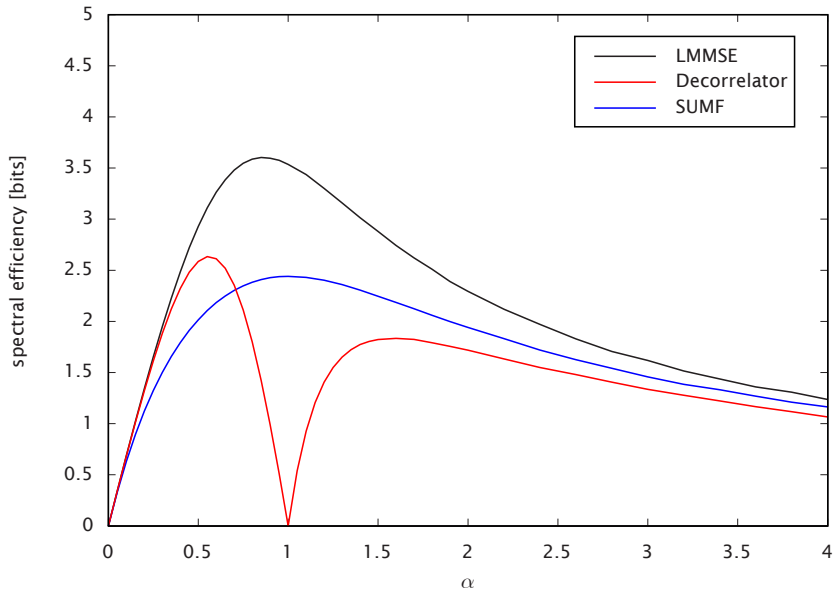
antennas are assumed to be uncorrelated and, thus, the Type-1 and Type-2 mismatched LMMSE estimators have equal performance. For the optimum LMMSE-CE we have plotted separately the MSEs for the transmit antennas at the edges and in the middle of the linear array with $M = 4$ elements. As expected, when the angular spread is high (transmit correlation low), the performance of the optimum (Example 2) and covariance mismatched (Example 3) LMMSE estimators are equal. For high spatial correlation at the transmitter side, however, significantly lower MSE can be obtained if the transmit covariance matrix is known at the channel estimator. The ML channel estimator (Example 4) that neglects both spatial correlation and additive noise provides the worst performance, as expected.

The spectral efficiency of a MIMO DS-CDMA system using $M = N = 4$ antennas at both ends and $\tau_{\mathsf{tr}} = 4$ or $\tau_{\mathsf{tr}} = 10$ training symbols per fading block is plotted in Figures 3.2a and 3.2b, respectively. The channel coherence time is set to $T_{\mathsf{coh}} = 50$ symbols and the loss in system throughput due to pilot symbols

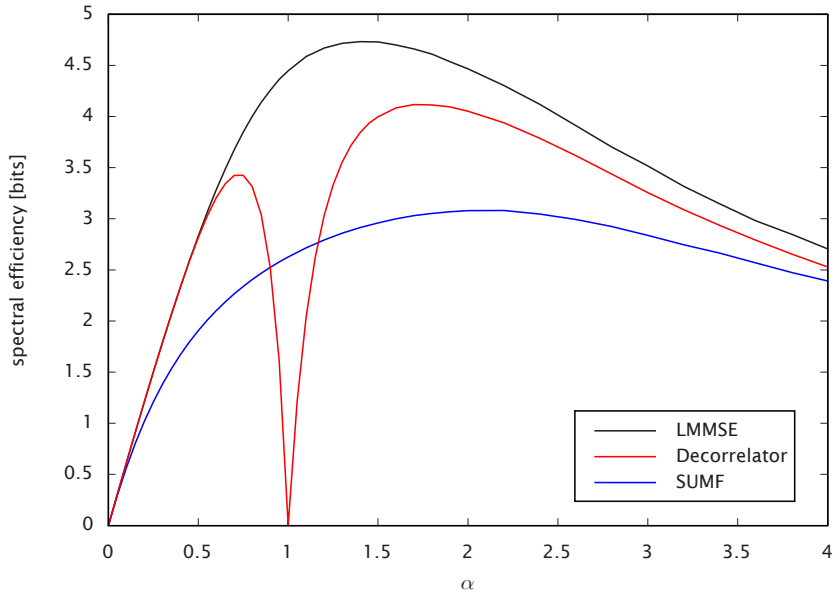
is taken into account in the plots. In all cases the LMMSE channel estimator of Example 2 is used. For all MUDs, using ten training symbols instead of four results to better spectral efficiency, even though the effective bandwidth is reduced due to transmission of known pilots. This is an example of the trade-off between the loss incurred by the seriousness of the CSI mismatch and the reduction in bandwidth due to training sequences. It is interesting to note that the MUDs respond differently to the level of channel uncertainty. Most serious loss in spectral efficiency due to lack of precise CSI is experienced by the decorrelator. In fact, if the channel information is not accurate enough, there is virtually no benefit of using the decorrelator instead of the simple SUMF. At high loads, the SUMF even offers better per-user rates than the decorrelator when $\tau_{\mathsf{tr}} = 4$ pilot symbols are used.

We next consider the effects of spatial correlation on the system throughput. In Figure 3.3, the spectral efficiency $\mathsf{C}^{\mathsf{qpsk}}$ of the linear MUDs presented in Examples 5 and 6 with the LMMSE channel estimator of Example 2 is plotted as a function of the angular spread at the transmitter side $\delta_{\mathsf{tx}}$ (in degrees). The receive antennas are assumed to be uncorrelated. The advantage of having high transmit correlation is most prominent for the overloaded case $\alpha = K/L = 2$, where the spectral efficiency $\mathsf{C}^{\mathsf{qpsk}}$ is more than doubled for the LMMSE and decorrelator MUDs compared to the case of low transmit correlation. Intuitive explanation is that here $\mathsf{C}^{\mathsf{qpsk}}$ depends on transmit correlation only through the error covariance $\boldsymbol{\Omega}_{\Delta \boldsymbol{h}_m} = \lambda_{\Delta \boldsymbol{h}_m} \boldsymbol{I}_N$, so that $\mathsf{C}^{\mathsf{qpsk}}_m(\lambda'_{\Delta \boldsymbol{h}_m}) \geq \mathsf{C}^{\mathsf{qpsk}}_m(\lambda_{\Delta \boldsymbol{h}_m})$ when $\lambda'_{\Delta \boldsymbol{h}_m} \leq \lambda_{\Delta \boldsymbol{h}_m}$. As we saw in Example 7, correlation benefits the CE and we therefore get an improvement in the spectral efficiency as angular spread $\delta_{\mathsf{tx}}$ decreases. Note that when the transmit antennas are highly correlated, the overloaded case $\alpha = 2$ offers higher total throughput than the half loaded system $\alpha = 0.5$ for all MUDs. If the transmit correlation is low, however, higher throughput is achieved with the user load $\alpha = 0.5$ regardless of the MUD. Interestingly, for the low user load $\alpha = 0.5$, the performance of the MUDs is arranged in increasing order of complexity, but this does not hold for the overloaded case $\alpha = 2$. In the latter, the spectral efficiencies of the SUMF and the decorrelator cross at around $\delta_{\mathsf{tx}} = 8$, and for low transmit correlation decorrelator performs worse than SUMF.

In Figure 3.4, the spectral efficiencies of the LMMSE-MUD and the SUMF with different channel estimators defined in Examples 2 and 3 are plotted as a function of the angular spread at the receiver side $\delta_{\mathsf{rx}}$ (in degrees). The user load is fixed to $\alpha = 2$ and an angular spread of $d_{\mathsf{tx}} = 3$ degrees (high correlation) at the transmitter side is assumed. As expected, when the antenna correlation at the receiver side decreases, a significant gain in spectral efficiency is observed. Sim-




**(a)** Number of known training symbols $\tau_{\mathsf{tr}} = 4$.




**(b)** Number of known training symbols $\tau_{\mathsf{tr}} = 10$.

**Figure 3.2.** Spectral efficiency $\mathsf{C}^{\mathsf{qpsk}}$ vs. the user load $\alpha = K/L$ for the linear MUDs. Uncorrelated $4 \times 4$ MIMO channel, coherence time of $T_{\mathsf{coh}} = 50$ symbols and average SNR of 10 dB. Linear MMSE channel estimator.

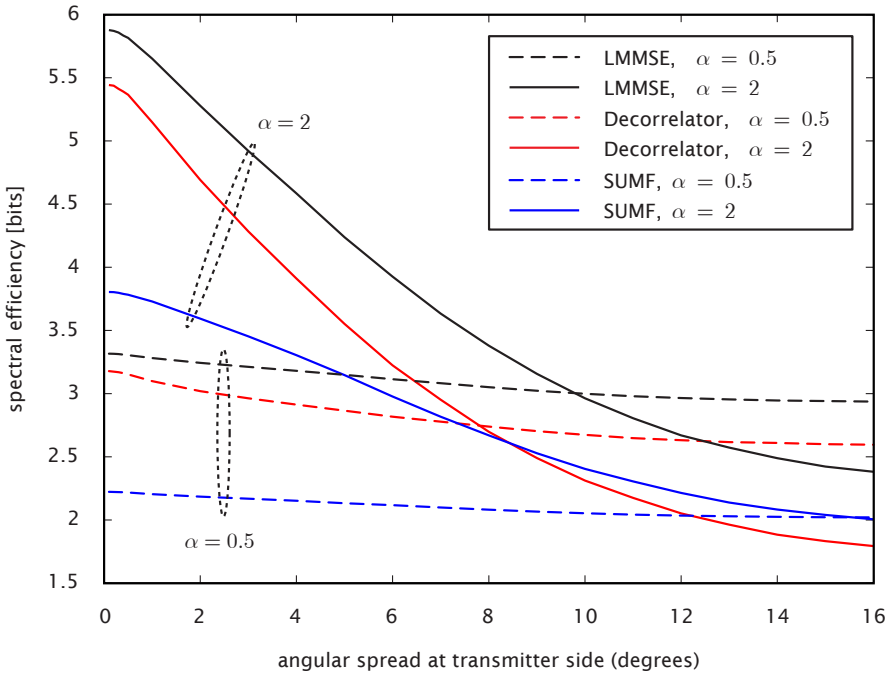



**Figure 3.3.** Spectral efficiency vs. angular spread at the transmitter $\delta_{\mathsf{tx}} \in [0.1, 16]$ for the linear MUDs. Uncorrelated receive antennas and optimum pilot assisted LMMSE channel estimator (spatial correlation known perfectly). Equal transmit power per antenna, $4 \times 4$ MIMO channel, coherence time of $T_{\mathsf{coh}} = 50$ symbols, number of pilots per fading block $\tau_{\mathsf{tr}} = 4$ and average SNR per receive antenna $\overline{\mathsf{snr}} = 10$ dB.

ilar behavior was also observed earlier in the simple single-user example given in Section 2.2. If the optimum LMMSE-CE instead of the Type-1 (neglects transmit correlation) or Type-2 (postulates uncorrelated antennas) mismatched CE is used, $\mathsf{C}^{\mathsf{qpsk}}$ roughly doubles for both the LMMSE-MUD and the SUMF for all values of $\delta_{\mathsf{rx}}$. As a consequence, in this scenario it is in fact preferable to have the optimum LMMSE-CE with SUMF instead of Type-1 CE and LMMSE-MUD. Surprisingly, the difference in $\mathsf{C}^{\mathsf{qpsk}}$ between the Type-1 and Type-2 estimators is negligible in the considered case. This implies that if the transmit correlation is not known at the channel estimator, virtually no further loss will be encountered if the receive correlation is neglected as well.

Finally, we investigate the performance of the non-linear MAP detector described in Section 3.1.2. Note that a direct implementation of this MUD has in-

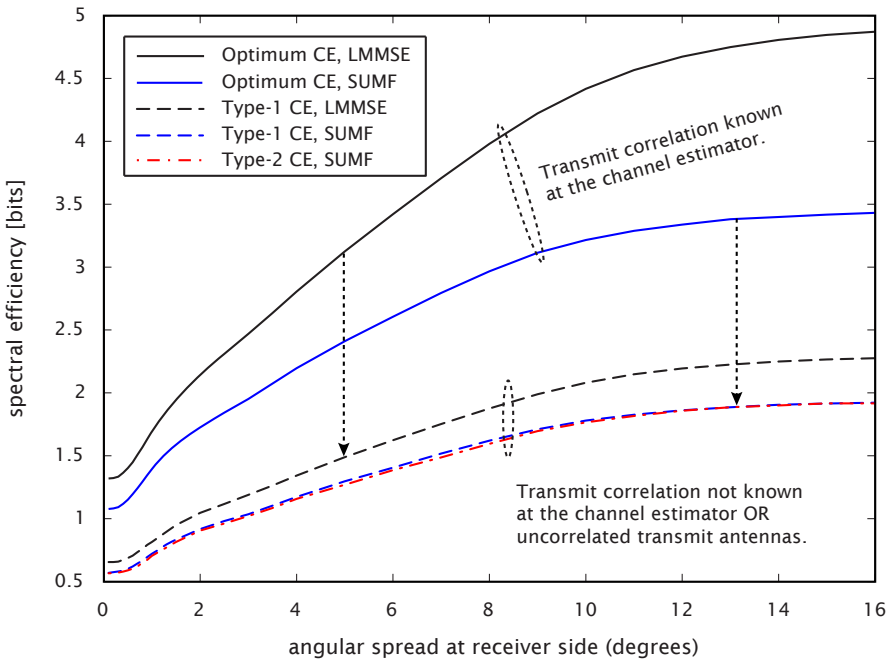



**Figure 3.4.** Spectral efficiency vs. angular spread at the receiver $\delta_{\mathsf{rx}} \in [0.1, 16]$ for the linear MMSE detector and SUMF. Correlated transmit antennas $\delta_{\mathsf{tx}} = 3$ and optimum LMMSE, Type-1 or Type-2 mismatched LMMSE channel estimator. Equal transmit power per antenna, $4 \times 4$ MIMO channel, coherence time of $T_{\mathsf{coh}} = 50$ symbols. Number of pilots per fading block $\tau_{\mathsf{tr}} = 4$, user load $\alpha = 2$ and average SNR per receive antenna $\overline{\mathsf{snr}} = 10$ dB.

feasible complexity for practical systems. It is, however, useful for benchmarking the more practical non-iterative detectors studied earlier, as well as providing an upper bound for the performance of the approximate non-linear methods such as sphere decoding and related algorithms [188–192].

The spectral efficiency of a $4 \times 4$ MIMO DS-CDMA system with MAP or LMMSE multiuser detector and $\tau_{\mathsf{tr}} = 4$ or $\tau_{\mathsf{tr}} = 10$ training symbols per fading block is plotted in Figure 3.5. The channel coherence time is set to $T_{\mathsf{coh}} = 50$ symbols and the loss in system throughput due to pilot symbols is taken into account in the plots. In all cases the LMMSE channel estimator of Example 2 is used and the antennas are assumed to be uncorrelated. The same conclusions as we made from Figures 3.2a and 3.2b can be drawn, i.e., except for low user loads using ten training symbols instead of four results to better spectral efficiency, even when the loss in effective bandwidth is taken into account due to transmission of known pilots. For

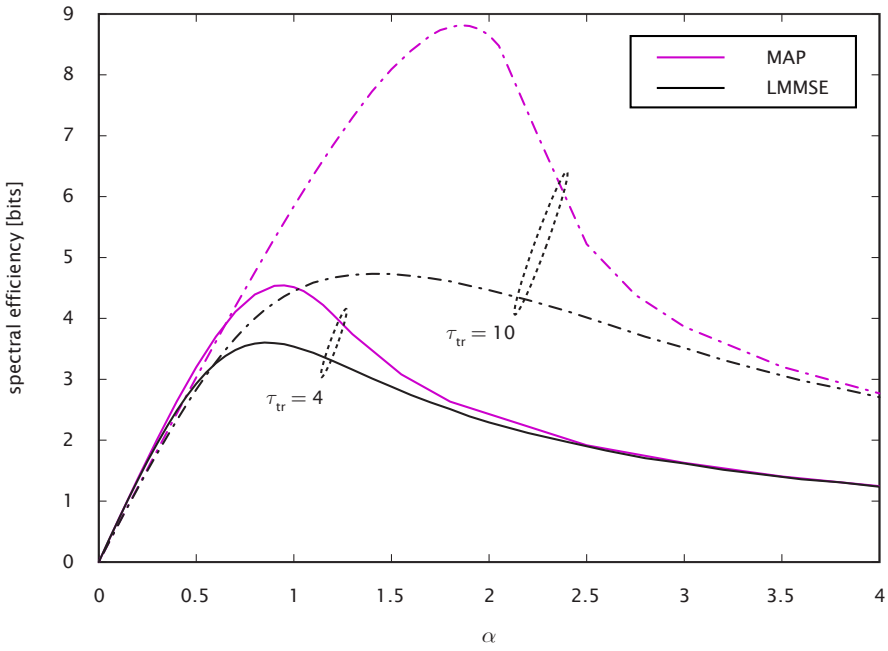



**Figure 3.5.** Spectral efficiency $\mathsf{C}^{\mathsf{qpsk}}$ vs. the user load $\alpha = K/L$ for the MAP-MUD and the LMMSE detector. Uncorrelated $4\times4$ MIMO channel, coherence time of $T_{\mathsf{coh}} = 50$ symbols and average SNR of 10 dB. Linear MMSE channel estimator.

the MAP-MUD, the loss due to severe CSI mismatch is much more pronounced than for the linear MUDs considered earlier. This makes intuitively sense since the MAP-MUD has "more to lose", so to speak, compared to the linear MUDs that will suffer more severely from the MAI in any case — even if perfect CSI is provided to them.

Figure 3.6 compares the spectral efficiencies $\mathsf{C}^{\mathsf{qpsk}}$ obtained by the non-linear MAP-MUD and the linear MMSE detector. The receive antennas are assumed to be uncorrelated and the angular spread at the transmitter side is $\delta_{\mathsf{tx}} \in [0.1, 16]$ (in degrees). The optimum pilot-aided LMMSE-CE provides the CSI for both detectors and is obtained by using $\tau_{\mathsf{tr}} = 4$ known training symbols per transmit antenna and fading block. The effect of antenna correlation to the spectral efficiency of the MAP-MUD is dramatic, going from about 13 bits to less than 2.5 bits per channel use as the transmit antennas become uncorrelated. In fact, for uncorrelated transmit antennas we lose all the benefits of the MAP-MUD compared to the LMMSE-MUD for the given configuration. Thus, the spectral efficiency obtained with the

LMMSE-MUD can be roughly doubled by using the non-iterative MAP-MUD, but only if sufficiently accurate channel estimation can be performed.

The straightforward solution to the problem of obtaining an accurate CSI is to use a greater number of pilot symbols per fading block. This, however, decreases the spectral efficiency as the training symbols consume the bandwidth from information bearing transmission so that there is an optimum trade-off between the number of training symbols and CSI accuracy. Alternatively, if the system is delay tolerant and using pilot-aided channel estimation, we could make the transmit antennas highly correlated while keeping the receive antennas as uncorrelated as possible. This of course causes a design conflict for the uplink / downlink transmission and should be carefully balanced so that a desired trade-off is reached. The final remedy is suggested by the iterative algorithms, studied in the next chapter, that may have potential to provide significant gains over non-iterative channel estimation and MUD also in MIMO CDMA systems. This topic is, however, left for future research and not considered in the present dissertation.

## 3.5 Chapter Summary and Conclusions

In this chapter, the performance of a randomly spread MIMO DS-CDMA system using non-iterative linear channel estimation and multiuser detection was studied. The considered channel estimators included the optimum pilot-aided LMMSE-CE, two covariance mismatched LMMSE-CEs and a maximum likelihood CE. The multiuser detectors included the non-linear MAP-MUD, linear MMSE and decorrelating MUDs and the single-user matched filter. Rayleigh fading single-path MIMO channel with spatial correlation was assumed between the transmitters and the receiver.

The performance analysis was carried out with the help of the replica method that provided a single-user characterization of the multiuser system in the large system limit. In contrast to some earlier results, we took into account the CSI mismatch caused by the pilot-aided channel estimation, as well as the effect of antenna correlation in the mathematical analysis. As a performance measure for the considered system, the QPSK constrained capacity with separate decoding was derived.

The results indicated that the ergodic spectral efficiency achieved with uncorrelated transmit antennas could be significantly improved if the transmit antennas were allowed to be correlated. This is in contrast to the case of perfect channel information, where correlation between the transmit antennas has no effect on the ergodic performance of the system. It is important to remark that the improve-

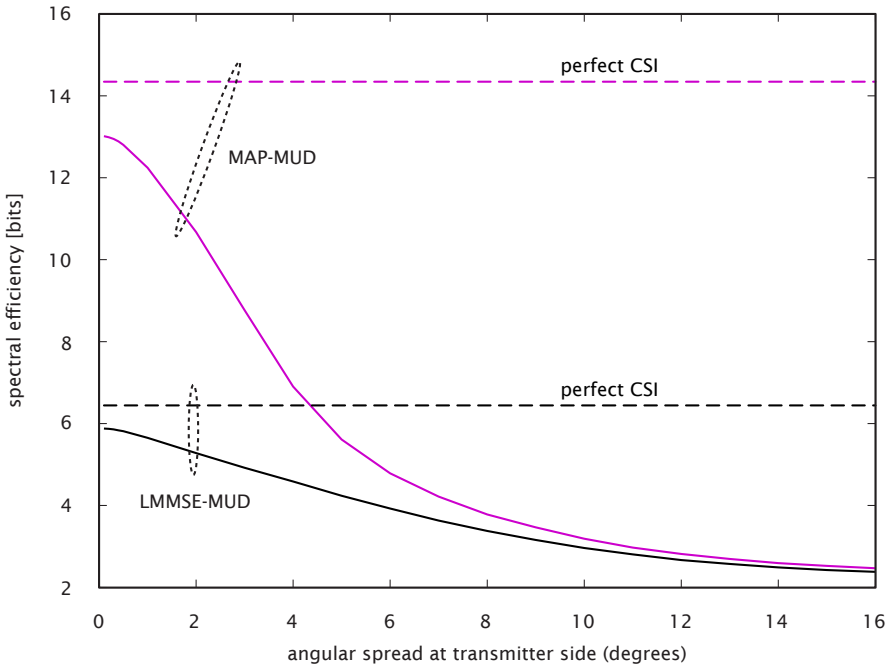



**Figure 3.6.** Spectral efficiency vs. angular spread at the transmitter $\delta_{\mathsf{tx}} \in [0.1, 16]$ for the MAP-MUD and the LMMSE detector. Uncorrelated receive antennas and optimum LMMSE channel estimator (spatial correlation known perfectly). Equal transmit power per antenna, $4 \times 4$ MIMO channel, coherence time of $T_{\mathsf{coh}} = 50$ symbols, number of pilots per fading block $\tau_{\mathsf{tr}} = 4$. User load $\alpha = 2$ and average SNR per receive antenna $\overline{\mathsf{snr}} = 10$ dB.

ment in spectral efficiency required no information at the transmitter. The channel estimator, however, needed the knowledge of the long-term spatial correlation in advance. Neglecting the transmitter side correlation at the receiver resulted to the same spectral efficiency as obtained for uncorrelated transmit antennas. The effect of neglecting the receiver side correlation turned out to have little effect on the performance.

# Chapter 4

# Iterative Receivers for DS-CDMA in Multipath Fading Channels

In this chapter, the performance of DS-CDMA systems using iterative channel estimation, multiuser detection, and single-user decoding is studied. Multipath Rayleigh fading channel model described in Section 2.2.2 is considered. For simplicity, both the transmitter and the receiver are assumed be equipped with a single antenna. Throughout the chapter, it is also presupposed that the coding and modulation is provided by the BICM introduced in Section 2.3.2, and the code word is long enough to span several independent fading blocks.

The outline of the chapter is as follows. Section 4.1 derives the iterative channel estimators and multiuser decoders that will be studied in the latter parts of the chapter Both hard and soft feedback are considered. The specific estimators studied in detail are:

- ***Channel estimators:*** iterative LMMSE channel estimator with soft feedback, approximate ML estimator using hard feedback;
- ***Data estimators:*** iterative maximum a posteriori, LMMSE and SUMF MUDDs with soft feedback, SUMF with hard feedback.

In Section 4.2, the decoupled single-user channel models related to the estimators given above are derived in the large system limit with the help of the replica method. Using the obtained single-user characterization, the performance of the iterative multiuser receivers is studied in Section 4.3. Numerical examples and discussion is provided in Section 4.4. Selected set of proofs can be found in Appendices B – F.

## 4.1 Iterative Multiuser Receivers via Bayesian Framework

The iterative multiuser channel and data estimators for the DS-CDMA system discussed in Section 2.2.2 are derived in this section. To help the reader to keep track of the derivations, the outline for the rest of the section is given below. The estimators are presented there in a decreasing order of complexity.

- Section 4.1.1 describes the general algorithm for iterative channel estimation and MUDD that will be considered in this chapter.
- Section 4.1.2 considers an iterative channel estimator that takes full advantage of the feedback information provided by the single-user decoders. This estimator turns out to be non-linear with exponential complexity.
- In Section 4.1.3, a set of low complexity iterative channel estimators based on linear filtering are introduced. Due to the simplifying assumptions, however, these estimators experience inevitable performance degradation compared to the non-linear CE given in the preceding section.
- Section 4.1.4 derives an iterative MAP detector that utilizes directly the CSI and the feedback information it receives from the channel estimator and the single-user decoders. This iterative MUDD gives the upper bound for the performance of the rest of the data decoding algorithms studied in the chapter but has exponential complexity.
- In Section 4.1.5, low complexity data estimators utilizing linear filtering and parallel interference cancellation are presented. Several suboptimal solutions resulting from different levels of model mismatch are considered.

### 4.1.1 General Framework for Iterative Channel Estimation, Detection, and Decoding

In the following, the postulated channel and data symbols for the user $k = 1, \ldots, K$, at time instant $t = \tau_{\mathsf{tr}} + 1, \ldots, T_{\mathsf{coh}}$, are written as $\tilde{\boldsymbol{h}}_{k,t}[c] \in \mathbb{C}^M$ and $\tilde{x}_{k,t}[c]$, respectively. Note that the postulated channel depends on $t$, while the true channel $\boldsymbol{h}_k[c]$ does not. The reason for this will become clear later. We also denote $\tilde{\boldsymbol{x}}_k[c] = \left[\tilde{x}_{k,\tau_{\mathsf{tr}}+1}[c] \ \cdots \ \tilde{x}_{k,T_{\mathsf{coh}}}[c]\right]^{\mathsf{T}} \in \mathbb{C}^{\tau_{\mathsf{d}}}$ for the postulated data symbols of the $k$th user in the $c$th fading block, and assign the priors (to be defined later) $\mathbb{Q}(\tilde{\boldsymbol{x}}_k[c])$ and $\mathbb{Q}(\tilde{\boldsymbol{h}}_{k,t}[c])$ to the above RVs for all $k = 1, \ldots, K$.

By assumption, the channel and the data estimator have knowledge of the received vectors $\mathcal{Y}_c$ as well as the training symbols and the spreading matrices, i.e.,

$\mathcal{I}_c = \{\mathcal{P}_c, \mathcal{S}\}$ at each fading block $c = 1, \ldots, C$. The estimators may have also received some information via feedback from the single-user decoders. Note that the content of the feedback has been obtained during the previous iteration.

Consider iteration $\ell = 1, 2, \ldots$ and let $\mathbb{P}_{\mathsf{ext}}^{(\ell-1)}(x_{k,t})$ and $\mathbb{P}_{\mathsf{app}}^{(\ell-1)}(x_{k,t})$, be the extrinsic and approximate a posteriori probabilities, respectively, of the transmitted symbol $x_{k,t} \in \mathcal{M}$. For convolutional codes, both probabilities are easy to obtain by using the BCJR algorithm [193–196]. Let the feedback from the single-user decoders be in the form of probabilities $\mathbb{Q}_{\mathsf{ext}}^{(\ell-1)}(\tilde{x}_{k,t})$ and $\mathbb{Q}_{\mathsf{app}}^{(\ell-1)}(\tilde{x}_{k,t})$, where again $\tilde{x}_{k,t} \in \mathcal{M}$. The relation between the decoder outputs and the feedback is defined by the operators $\varphi_{\mathsf{ext}}$ and $\varphi_{\mathsf{app}}$ that transform the probability measures (or distributions) $\mathbb{P}_{\mathsf{ext}}$ and $\mathbb{P}_{\mathsf{app}}$ to $\mathbb{Q}_{\mathsf{ext}}$ and $\mathbb{Q}_{\mathsf{app}}$, respectively, i.e.,

$$\varphi_{\mathsf{ext}} : \mathbb{P}_{\mathsf{ext}}^{(\ell-1)}(x) \mapsto \mathbb{Q}_{\mathsf{ext}}^{(\ell-1)}(\tilde{x} = x), \qquad x, \tilde{x} \in \mathcal{M}, \tag{4.1}$$

$$\varphi_{\mathsf{app}} : \mathbb{P}_{\mathsf{app}}^{(\ell-1)}(x) \mapsto \mathbb{Q}_{\mathsf{app}}^{(\ell-1)}(\tilde{x} = x), \qquad x, \tilde{x} \in \mathcal{M}. \tag{4.2}$$

Throughout the thesis we assume that the operators (4.1) and (4.2) do not depend on the iteration index $\ell$, and the feedback probabilities $\mathbb{Q}_{\mathsf{ext}}^{(\ell-1)}$ and $\mathbb{Q}_{\mathsf{app}}^{(\ell-1)}$ are well-defined over $\mathcal{M}$. The specific forms of $\varphi_{\mathsf{app}}$ and $\varphi_{\mathsf{ext}}$ define the type of feedback used and will be detailed in the next section.

Note that we used above the nomenclature common to iterative ISI cancellation and MUDD, where the extrinsic probabilities of the coded bits do not contain channel information, whereas the approximate APPs do (see for example, [16, 125]). Both probabilities are obtained using the knowledge of $\mathcal{C}_k$. We make two small remarks before proceeding to the iterative algorithm itself:

- The approximate APPs $\mathbb{P}_{\mathsf{app}}^{(\ell-1)}(x_{k,t}[c])$ obtained by the single-user decoders are in general different from the true APPs $\mathbb{P}\big(x_{k,t}[c] \mid \{\mathcal{Y}_c\}_{c=1}^{C}\big)$ for all $\ell = 1, 2, \ldots$;
- Also the APP-based feedback to the channel estimator has to be extrinsic to the CE in the sense defined for message passing algorithms in factor graphs [21–26, 124].

With the above in mind, a high-level algorithm for iterative channel estimation and MUDD is given in Table 4.1. The details of the steps are postponed to the later parts of the chapter, where some special cases of this framework are considered. To initiate the iterative process, we let $\mathbb{Q}_{\mathsf{app}}^{(0)}(\tilde{x}_{k,t}[c])$ and $\mathbb{Q}_{\mathsf{ext}}^{(0)}(\tilde{x}_{k,t}[c])$ be equal to (2.19). The block diagram of the receiver is depicted in Fig. 4.1, where we omitted the iteration index $\ell = 1, 2, \ldots$, for clarity.

**Table 4.1.** Iterative channel estimation and MUDD

1. Consider the problem of obtaining the CSI at time index $\vartheta \in \mathcal{D}$ during the $\ell$th iteration. Let the channel estimator postulate the priors $\{\mathbb{Q}(\tilde{\boldsymbol{h}}_{k,\vartheta}[c]) \mid \forall k \in \mathcal{K}\}$ and assume that the approximate APPs $\{\mathbb{Q}_{\mathsf{app}}^{(\ell-1)}(\tilde{x}_{k,t}[c]) \mid \forall k \in \mathcal{K}, t \in \mathcal{D} \setminus \vartheta\}$ obtained by the single-user decoders during the previous iteration have been received from the iterative MUDD. Given the information

$$\mathcal{I}_{\vartheta}^{(\ell)}[c] = \Big\{\mathcal{I}_c, \mathcal{Y}_c \setminus \boldsymbol{y}_{\vartheta}[c], \{\mathbb{Q}_{\mathsf{app}}^{(\ell-1)}(\tilde{x}_{k,t}[c]) \mid \forall k \in \mathcal{K}, t \in \mathcal{D} \setminus \vartheta\}\Big\}, \quad (4.3)$$

   and its knowledge about the system model (2.10), the iterative channel estimator calculates the posterior probabilities $\{\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{h}}_{k,t}[c] \mid \mathcal{I}_t^{(\ell)}[c]) \mid \forall k \in \mathcal{K}, t \in \mathcal{D}\}$ for all $c = 1, \ldots, C$, and sends the obtained CSI to the iterative MUDD.

2. Let the data estimator assign the postulated prior $\mathbb{Q}(\tilde{x}_{\xi,t}[c])$ to the the data symbol of user $\xi \in \mathcal{K}$ at time instant $t \in \mathcal{D}$. Given the information

$$\mathcal{I}_{\xi,t}^{(\ell)}[c] = \Big\{\mathcal{I}_c, \boldsymbol{y}_t[c], \{\mathbb{Q}_{\mathsf{ext}}^{(\ell-1)}(\tilde{x}_{j,t}[c]) \mid \forall j \in \mathcal{K} \setminus \xi\}, \{\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{h}}_{k,t}[c] \mid \mathcal{I}_t^{(\ell)}[c]) \mid \forall k \in \mathcal{K}\}\Big\}, \quad (4.4)$$

   and its knowledge about the channel (2.10), for each fading block $c = 1, \ldots, C$, the data estimator of the $\xi$th user calculates the symbol-by-symbol posterior probabilities $\{\mathbb{Q}^{(\ell)}(\tilde{x}_{\xi,t}[c] \mid \mathcal{I}_{\xi,t}^{(\ell)}[c]) \mid \forall t \in \mathcal{D}\}$ and sends them to the single-user sum-product decoder.

3. For $k = 1, \ldots, K$, the posterior probabilities of the data symbols $\{\mathbb{Q}^{(\ell)}(\tilde{x}_{k,t}[c] \mid \mathcal{I}_{k,t}^{(\ell)}[c]) \mid \forall c, t \in \mathcal{D}\}$ and the code book $\mathcal{C}_k$ are used by the sum-procuct decoder to calculate the approximate a posteriori $\mathbb{P}_{\mathsf{app}}^{(\ell)}(\boldsymbol{x}_k)$ and extrinsic $\mathbb{P}_{\mathsf{ext}}^{(\ell)}(\boldsymbol{x}_k)$ probabilities of the data symbols. For trellis codes these probabilities can be easily obtained by the BCJR algorithm [193–196].

4. The operators $\varphi_{\mathsf{ext}}$ and $\varphi_{\mathsf{app}}$ are applied to the outputs of the sum-product decoders $\mathbb{P}_{\mathsf{ext}}^{(\ell)}(x_{k,t}[c])$ and $\mathbb{P}_{\mathsf{app}}^{(\ell)}(x_{k,t}[c])$, respectively, to produce the corresponding feedback probabilities $\mathbb{Q}_{\mathsf{ext}}^{(\ell)}(\tilde{x}_{k,t}[c])$ and $\mathbb{Q}_{\mathsf{app}}^{(\ell)}(\tilde{x}_{k,t}[c])$. The former are sent to the channel estimator while the latter are stored and used by the data estimator during the next iteration.

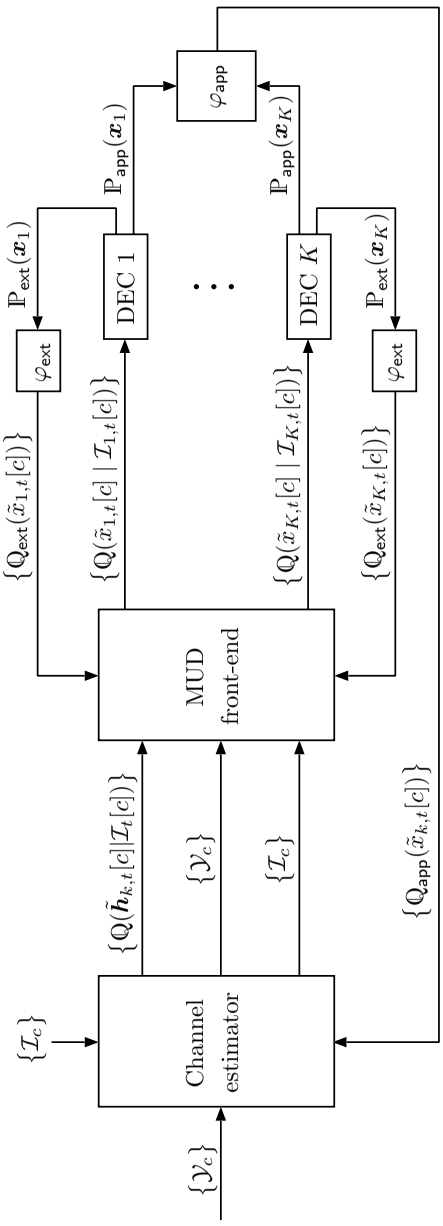



**Figure 4.1.** Simplified block diagram of a receiver employing iterative channel estimation, detection, and decoding.

### 4.1.2 Non-Linear Channel Estimation with Soft Feedback

Let us start by omitting the block index $c$ for notational simplicity and postulate a new channel model related to (2.10) as

$$\tilde{\boldsymbol{y}}_t = \begin{cases} \dfrac{1}{\sqrt{L}} \displaystyle\sum_{k=1}^{K} \boldsymbol{S}_{k,t} \tilde{\boldsymbol{h}}_{k,\vartheta} p_{k,t} + \tilde{\boldsymbol{w}}_t \in \mathbb{C}^L, & t \in \mathcal{T}, \\ \dfrac{1}{\sqrt{L}} \displaystyle\sum_{k=1}^{K} \boldsymbol{S}_{k,t} \tilde{\boldsymbol{h}}_{k,\vartheta} \tilde{x}_{k,t} + \tilde{\boldsymbol{w}}_t \in \mathbb{C}^L, & t \in \mathcal{D} \setminus \vartheta, \end{cases} \tag{4.5}$$

where the noise vectors are IID $\mathbb{Q}(\tilde{\boldsymbol{w}}_t) = \mathsf{CN}(\boldsymbol{0};\, \sigma^2 \boldsymbol{I}_L)$ and the training symbols $\mathcal{P} = \{p_{k,t} \mid \forall k, t \in \mathcal{T}\}$ are known at the receiver. We denote $\tilde{\mathcal{H}}_\vartheta = \{\tilde{\boldsymbol{h}}_{k,\vartheta}\}_{k \in \mathcal{K}}$, for the postulated channel at time instant $\vartheta$ and $\tilde{\mathcal{Y}} \setminus \tilde{\boldsymbol{y}}_\vartheta = \{\tilde{\boldsymbol{y}}_t \mid t \in \mathcal{T} \cup \mathcal{D} \setminus \vartheta\}$ for the set of all received signals in (4.5). For notational convenience, we also introduce vector representation (similarly for other variables and index sets)

$$\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} = \begin{bmatrix} \tilde{x}_{k,\tau_{\mathsf{tr}}+1} & \cdots & \tilde{x}_{k,\vartheta-1} & \tilde{x}_{k,\vartheta+1} & \cdots & \tilde{x}_{k,T_{\mathsf{coh}}} \end{bmatrix}^{\mathsf{T}} \in \mathbb{C}^{\tau_{\mathsf{d}}-1}, \tag{4.6}$$

for $\vartheta \in \mathcal{D}$ and $k \in \mathcal{K}$.

Now, postulate $\mathbb{Q}(\tilde{\mathcal{H}}_\vartheta) = \prod_{k=1}^{K} \mathbb{Q}(\tilde{\boldsymbol{h}}_{k,\vartheta})$, where $\mathbb{Q}(\tilde{\boldsymbol{h}}_{k,\vartheta}) = \mathsf{CN}(\boldsymbol{0};\, \boldsymbol{\Omega}_{\boldsymbol{h}_k})$, and let $\varphi_{\mathsf{app}}$ be the identity operator. We assign the prior probabilities

$$\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} = \boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta}) = \mathbb{P}_{\mathsf{app}}^{(\ell-1)}(\boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta}), \tag{4.7}$$

to the data symbols, where $\mathbb{P}_{\mathsf{app}}^{(\ell-1)}(\boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta})$ are the approximate APPs obtained by the sum-product decoders during the previous iteration, as discussed in the previous section. In the following we shall abbreviate the equalities of the kind (4.7) simply as $\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}) = \mathbb{P}_{\mathsf{app}}^{(\ell-1)}(\boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta})$. Note that albeit the probabilities (4.7) are APPs in the typical turbo processing jargon, they represent extrinsic information to the channel estimator.

The postulated posterior probability of the channel coefficients $\boldsymbol{h}_k$, given $\mathcal{I}_\vartheta^{(\ell)}$ and the knowledge of (4.5) reads

$$\begin{aligned} &\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{h}}_{k,\vartheta} \mid \mathcal{I}_\vartheta^{(\ell)}) \\ &= \frac{\mathbb{Q}(\tilde{\boldsymbol{h}}_{k,\vartheta})}{\mathbb{E}_{\tilde{\mathcal{H}}_\vartheta}\left\{ \displaystyle\sum_{\{\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}\}} \prod_{k \in \mathcal{K}} \mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}) \mathbb{Q}(\tilde{\mathcal{Y}} \setminus \tilde{\boldsymbol{y}}_\vartheta = \mathcal{Y} \setminus \boldsymbol{y}_\vartheta \mid \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}, \tilde{\mathcal{H}}_\vartheta, \mathcal{I}_\vartheta^{(\ell)}) \right\}} \\ &\times \mathbb{E}_{\tilde{\mathcal{H}}_\vartheta \setminus \tilde{\boldsymbol{h}}_{k,\vartheta}}\left\{ \sum_{\{\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}\}} \prod_{k \in \mathcal{K}} \mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}) \mathbb{Q}(\tilde{\mathcal{Y}} \setminus \tilde{\boldsymbol{y}}_\vartheta = \mathcal{Y} \setminus \boldsymbol{y}_\vartheta \mid \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}, \tilde{\mathcal{H}}_\vartheta, \mathcal{I}_\vartheta^{(\ell)}) \right\}, \end{aligned} \tag{4.8}$$

where the expectations are with respect to the postulated a priori distribution of the channel. The summations are over all possible transmitted signals vectors of all users during the time instants $t \in \mathcal{D} \setminus \vartheta$. The GPME is thus given by

$$\langle \tilde{\boldsymbol{h}}_{k,\vartheta} \rangle_{(\ell)} = \int \tilde{\boldsymbol{h}}_{k,\vartheta} \mathrm{d}\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{h}}_{k,\vartheta} \mid \mathcal{I}_{\vartheta}^{(\ell)}), \qquad k = 1, \ldots, K. \tag{4.9}$$

For arbitrary a priori probabilities of the data symbols[1], however, the number of computations required to calculate the the summations in (4.8) grows exponentially with $K$ and the number of elements in $\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}$, making this channel estimator highly impractical. Thus, the emphasis in this thesis is on the iterative linear channel estimators, introduced in the next section, which have only polynomial complexity.

### 4.1.3 Linear Channel Estimation with Information Feedback

Consider estimating the CSI $\{\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{h}}_{k,\vartheta}[c] \mid \mathcal{I}_{\vartheta}^{(\ell)}[c])\}$ for fixed time index $\vartheta \in \mathcal{D}$ during the $c$th fading block. As above, we let (4.3) be available at the channel estimator and drop the block index $c$ for notational convenience.

Assume that the receiver is at its $\ell$th iteration. Let the feedback based posterior mean estimates of the data symbols $\{x_{k,t} \mid \forall k, t \in \mathcal{D} \setminus \vartheta\}$ from previous iteration be given by

$$\langle \tilde{x}_{k,t} \rangle_{\mathsf{app}}^{(\ell-1)} = \sum_{\tilde{x}_{k,t} \in \mathcal{M}} \tilde{x}_{k,t} \mathbb{Q}_{\mathsf{app}}^{(\ell-1)}(\tilde{x}_{k,t}), \qquad \forall k, t \in \mathcal{D} \setminus \vartheta. \tag{4.10}$$

Following the notation introduced in (4.6), let $\langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}}^{(\ell-1)} \in \mathbb{C}^{\tau_{\mathsf{d}}-1}$ be the vector consisting of the symbols (4.10) of the user $k = 1, \ldots, K$, i.e.,

$$\begin{aligned} &\langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}}^{(\ell-1)} \\ &\quad = \Big[ \langle \tilde{x}_{k,\tau_{\mathsf{tr}}+1} \rangle_{\mathsf{app}}^{(\ell-1)} \quad \cdots \quad \langle \tilde{x}_{k,\vartheta-1} \rangle_{\mathsf{app}}^{(\ell-1)} \quad \langle \tilde{x}_{k,\vartheta+1} \rangle_{\mathsf{app}}^{(\ell-1)} \quad \cdots \quad \langle \tilde{x}_{k,T_{\mathsf{coh}}} \rangle_{\mathsf{app}}^{(\ell-1)} \Big]^{\mathsf{T}}, \end{aligned} \tag{4.11}$$

and define the error terms

$$\Delta \boldsymbol{u}_{k,m} = \Delta \boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta} h_{k,m} \in \mathbb{C}^{(\tau_{\mathsf{d}}-1)}, \tag{4.12}$$

$$\Delta \boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta} = \boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta} - \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}}^{(\ell-1)}. \tag{4.13}$$

Given $\mathcal{I}_{\vartheta}^{(\ell)}[c]$, the RVs (4.12) – (4.13) are zero-mean in the limit of large code word length, and we denote the corresponding conditional covariance matrices by

[1] By this we mean other than the case when $\prod_{k \in \mathcal{K}} \mathbb{Q}^{(\ell)}\big(\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}\big)$ is non-zero for only one combination of transmitted symbols from all users and, thus, perfectly known at the receiver.

$\boldsymbol{\Omega}^{(\ell)}_{\Delta \boldsymbol{u}_{k,m}}$ and $\boldsymbol{\Omega}^{(\ell-1)}_{\Delta \boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta}}$, respectively, where the iteration index for $\Delta \boldsymbol{u}_{k,m}$ is chosen to be $\ell$ instead of $\ell-1$ by a notational convention. For future reference, let us also write the spreading sequences related to the training and data transmission phases of (2.10) in the matrix form

$$\boldsymbol{S}_{k,\mathcal{T},m} = \operatorname{diag}\big(\boldsymbol{s}_{k,1,m},\ldots,\boldsymbol{s}_{k,\tau_{\mathsf{tr}},m}\big) \in \mathbb{C}^{\tau_{\mathsf{tr}}L\times\tau_{\mathsf{tr}}}, \tag{4.14}$$

$$\boldsymbol{S}_{k,\mathcal{D}\setminus\vartheta,m} = \operatorname{diag}\big(\boldsymbol{s}_{k,\tau_{\mathsf{tr}}+1,m},\ldots,\boldsymbol{s}_{k,\vartheta-1,m},\boldsymbol{s}_{k,\vartheta+1,m},\ldots,\boldsymbol{s}_{k,T_{\mathsf{coh}},m}\big) \in \mathbb{C}^{(\tau_{\mathsf{d}}-1)L\times(\tau_{\mathsf{d}}-1)}, \tag{4.15}$$

respectively.

The information contained in the set $\mathcal{Y}\setminus\boldsymbol{y}_\vartheta$ and the channel model (2.10) can now be written in the form

$$\boldsymbol{y}_{\mathcal{T}} = \frac{1}{\sqrt{L}}\sum_{k=1}^{K}\sum_{m=1}^{M}\boldsymbol{S}_{k,\mathcal{T},m}\boldsymbol{p}_k h_{k,m} + \boldsymbol{w}_{\mathcal{T}}, \tag{4.16}$$

$$\boldsymbol{y}_{\mathcal{D}\setminus\vartheta} = \frac{1}{\sqrt{L}}\sum_{k=1}^{K}\sum_{m=1}^{M}\boldsymbol{S}_{k,\mathcal{D}\setminus\vartheta,m}\langle\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}\rangle^{(\ell-1)}_{\mathsf{app}} h_{k,m} + \frac{1}{\sqrt{L}}\sum_{k=1}^{K}\sum_{m=1}^{M}\boldsymbol{S}_{k,\mathcal{D}\setminus\vartheta,m}\Delta\boldsymbol{u}_{k,m} + \boldsymbol{w}_{\mathcal{D}\setminus\vartheta}, \tag{4.17}$$

where the noise vectors are independent zero-mean complex Gaussian RVs with distributions $\mathbb{P}(\boldsymbol{w}_{\mathcal{T}}) = \mathsf{CN}(\boldsymbol{0};\,\sigma^2\boldsymbol{I}_{\tau_{\mathsf{tr}}L})$ and $\mathbb{P}(\boldsymbol{w}_{\mathcal{D}\setminus\vartheta}) = \mathsf{CN}(\boldsymbol{0};\,\sigma^2\boldsymbol{I}_{(\tau_{\mathsf{d}}-1)L})$. Since the fading is assumed to be an ergodic process over the code words (cf. Section 2.3.2), we can regard $\Delta\boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta}$ and $h_{k,m}$ to be independent RVs and therefore

$$\boldsymbol{\Omega}^{(\ell)}_{\Delta\boldsymbol{u}_{k,m}} = \bar{t}_{k,m}\boldsymbol{\Omega}^{(\ell-1)}_{\Delta\boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta}} \in \mathbb{C}^{(\tau_{\mathsf{d}}-1)\times(\tau_{\mathsf{d}}-1)}, \qquad k=1,\ldots,K, m=1,\ldots,M. \tag{4.18}$$

Furthermore, $\mathbb{E}\{\Delta\boldsymbol{u}_{k,m}\Delta\boldsymbol{u}_{\xi,i}\} = \boldsymbol{0}$ if $\xi\neq k$ or $m\neq i$. Note that so-far we have not changed the system model and substituting (4.10) – (4.15) to (4.16) – (4.17) gives back the received vectors $\{\boldsymbol{y}_t \mid \forall t\neq\vartheta\}$ in (2.10).

Now, create a new channel model from (4.16) – (4.17) by replacing the set of true channel coefficients $\mathcal{H} = \{h_{k,m} \mid \forall k,m\}$ by postulated ones $\tilde{\mathcal{H}}_\vartheta = \{\tilde{h}_{k,\vartheta,m} \mid \forall k,m\}$. Let also the noise vectors have a postulated variance $\tilde{\sigma}^2$, i.e., $\mathbb{Q}(\tilde{\boldsymbol{w}}_{\mathcal{T}}) = \mathsf{CN}(\boldsymbol{0};\,\tilde{\sigma}^2\boldsymbol{I}_{\tau_{\mathsf{tr}}})$ and $\mathbb{Q}(\tilde{\boldsymbol{w}}_{\mathcal{D}\setminus\vartheta}) = \mathsf{CN}(\boldsymbol{0};\,\tilde{\sigma}^2\boldsymbol{I}_{\tau_{\mathsf{d}}-1})$. If we also postulate that (4.12) are independent zero-mean Gaussian RVs[2] uncorrelated with $\langle\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}\rangle^{(\ell-1)}_{\mathsf{app}}$, and

[2]Note that given only the mean $\boldsymbol{\mu}$ and the covariance $\boldsymbol{\Omega}$ of a continuous RV defined on $\mathbb{C}^L$, the maximum entropy distribution is $\mathsf{CN}(\boldsymbol{\mu};\,\boldsymbol{\Omega})$ [11, Chapter 12].

having postulated covariance matrix $\tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta \boldsymbol{u}_{k,m}}$, the receiver's knowledge about the channel is

$$\tilde{\boldsymbol{y}}_{\mathcal{T}} = \frac{1}{\sqrt{L}} \sum_{k=1}^{K} \sum_{m=1}^{M} \boldsymbol{S}_{k,\mathcal{T},m} \boldsymbol{p}_k h_{k,\vartheta,m} + \tilde{\boldsymbol{w}}_{\mathcal{T}}, \tag{4.19}$$

$$\begin{aligned} \tilde{\boldsymbol{y}}_{\mathcal{D}\setminus\vartheta} = & \frac{1}{\sqrt{L}} \sum_{k=1}^{K} \sum_{m=1}^{M} \boldsymbol{S}_{k,\mathcal{D}\setminus\vartheta,m} \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle^{(\ell-1)}_{\mathsf{app}} \tilde{h}_{k,\vartheta,m} \\ & + \frac{1}{\sqrt{L}} \sum_{k=1}^{K} \sum_{m=1}^{M} \boldsymbol{S}_{k,\mathcal{D}\setminus\vartheta,m} \Delta \tilde{\boldsymbol{u}}_{k,m} + \tilde{\boldsymbol{w}}_{\mathcal{D}\setminus\vartheta}. \end{aligned} \tag{4.20}$$

The posterior probability of the postulated channel coefficients given information (4.3) reads

$$\mathbb{Q}^{(\ell)}(\tilde{\mathcal{H}}_\vartheta \mid \mathcal{I}^{(\ell)}_\vartheta) = \frac{\mathbb{Q}(\tilde{\mathcal{H}}_\vartheta)\mathbb{Q}(\tilde{\boldsymbol{y}}_{\mathcal{T}} = \boldsymbol{y}_{\mathcal{T}} \mid \tilde{\mathcal{H}}_\vartheta, \mathcal{I}^{(\ell)}_\vartheta)\mathbb{Q}(\tilde{\boldsymbol{y}}_{\mathcal{D}\setminus\vartheta} = \boldsymbol{y}_{\mathcal{D}\setminus\vartheta} \mid \tilde{\mathcal{H}}_\vartheta, \mathcal{I}^{(\ell)}_\vartheta)}{\mathbb{E}_{\tilde{\mathcal{H}}_\vartheta}\Big\{\mathbb{Q}(\tilde{\boldsymbol{y}}_{\mathcal{T}} = \boldsymbol{y}_{\mathcal{T}} \mid \tilde{\mathcal{H}}_\vartheta, \mathcal{I}^{(\ell)}_\vartheta)\mathbb{Q}(\tilde{\boldsymbol{y}}_{\mathcal{D}\setminus\vartheta} = \boldsymbol{y}_{\mathcal{D}\setminus\vartheta} \mid \tilde{\mathcal{H}}_\vartheta, \mathcal{I}^{(\ell)}_\vartheta)\Big\}}, \tag{4.21}$$

where

$$\begin{aligned} & \mathbb{Q}(\tilde{\boldsymbol{y}}_{\mathcal{D}\setminus\vartheta} = \boldsymbol{y}_{\mathcal{D}\setminus\vartheta} \mid \tilde{\mathcal{H}}_\vartheta, \mathcal{I}^{(\ell)}_\vartheta) \\ & \quad = \mathbb{E}_{\{\Delta\tilde{\boldsymbol{u}}_{k,m} \mid \forall k,m\}}\{\mathbb{Q}(\tilde{\boldsymbol{y}}_{\mathcal{D}\setminus\vartheta} = \boldsymbol{y}_{\mathcal{D}\setminus\vartheta} \mid \{\Delta\tilde{\boldsymbol{u}}_{k,m}\}, \tilde{\mathcal{H}}_\vartheta, \mathcal{I}^{(\ell)}_\vartheta)\}. \end{aligned} \tag{4.22}$$

By definition

$$\mathbb{Q}(\tilde{\boldsymbol{y}}_{\mathcal{T}} = \boldsymbol{y}_{\mathcal{T}} \mid \tilde{\mathcal{H}}_\vartheta, \mathcal{I}^{(\ell)}_\vartheta) = \mathsf{CN}\left(\frac{1}{\sqrt{L}} \sum_{k=1}^{K} \sum_{m=1}^{M} \boldsymbol{S}_{k,\mathcal{T},m} \boldsymbol{p}_k \tilde{h}_{k,m,\vartheta};\, \tilde{\sigma}^2 \boldsymbol{I}_{\tau_{\mathsf{tr}}}\right), \tag{4.23}$$

and solving the Gaussian integrals with respect to $\{\Delta\tilde{\boldsymbol{u}}_{k,m} \mid \forall k,m\}$ gives

$$\begin{aligned} & \mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{y}}_{\mathcal{D}\setminus\vartheta} = \boldsymbol{y}_{\mathcal{D}\setminus\vartheta} \mid \tilde{\mathcal{H}}_\vartheta, \mathcal{I}^{(\ell)}_\vartheta) \\ & \quad = \mathsf{CN}\left(\frac{1}{\sqrt{L}} \sum_{k=1}^{K} \sum_{m=1}^{M} \boldsymbol{S}_{k,\mathcal{D}\setminus\vartheta,m} \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle^{(\ell-1)}_{\mathsf{app}} \tilde{h}_{k,m,\vartheta};\, \tilde{\boldsymbol{\Omega}}^{(\ell)}_{\mathsf{err}}\right), \end{aligned} \tag{4.24}$$

where

$$\tilde{\boldsymbol{\Omega}}^{(\ell)}_{\mathsf{err}} = \tilde{\sigma}^2 \boldsymbol{I}_{(\tau_{\mathsf{d}}-1)L} + \frac{1}{L} \sum_{k=1}^{K} \sum_{m=1}^{M} \boldsymbol{S}_{k,\mathcal{D}\setminus\vartheta,m} \tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta \boldsymbol{u}_{k,m}} \boldsymbol{S}^{\mathsf{H}}_{k,\mathcal{D}\setminus\vartheta,m}. \tag{4.25}$$

The marginal probability of the $k$th user's postulated channel is therefore given by

$$\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{h}}_{k,\vartheta} \mid \mathcal{I}_{\vartheta}^{(\ell)}) = \frac{\mathbb{Q}(\tilde{\boldsymbol{h}}_{k,\vartheta})\mathbb{E}_{\tilde{\mathcal{H}}_{\vartheta}\setminus\tilde{\boldsymbol{h}}_{k,\vartheta}}\{\mathbb{Q}(\tilde{\boldsymbol{y}}_{\mathcal{T}} = \boldsymbol{y}_{\mathcal{T}} \mid \tilde{\mathcal{H}}_{\vartheta}, \mathcal{I}_{\vartheta}^{(\ell)})\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{y}}_{\mathcal{D}\setminus\vartheta} = \boldsymbol{y}_{\mathcal{D}\setminus\vartheta} \mid \tilde{\mathcal{H}}_{\vartheta}, \mathcal{I}_{\vartheta}^{(\ell)})\}}{\mathbb{E}_{\tilde{\mathcal{H}}_{\vartheta}}\{\mathbb{Q}(\tilde{\boldsymbol{y}}_{\mathcal{T}} = \boldsymbol{y}_{\mathcal{T}} \mid \tilde{\mathcal{H}}_{\vartheta}, \mathcal{I}_{\vartheta}^{(\ell)})\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{y}}_{\mathcal{D}\setminus\vartheta} = \boldsymbol{y}_{\mathcal{D}\setminus\vartheta} \mid \tilde{\mathcal{H}}_{\vartheta}, \mathcal{I}_{\vartheta}^{(\ell)})\}}, \tag{4.26}$$

and the corresponding GPME (4.9) is parametrized by:

1. the operator $\varphi_{\mathsf{app}}$, defines the type of feedback;
2. the postulated covariance matrix $\tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta\boldsymbol{u}_{k,m}}$, determines the estimators knowledge about the feedback error statistics;
3. the postulated prior $\mathbb{Q}(\tilde{\boldsymbol{h}}_{k,\vartheta})$ and noise variance $\tilde{\sigma}^2$, give the type of linear filtering used.

By choosing these parameters appropriately, we can derive all the usual iterative channel estimators with linear filtering. In the following cases, we assume the channel estimator knows the correct statistics of the channel, i.e., the postulated prior is $\mathbb{Q}(\tilde{\boldsymbol{h}}_{k,t}) = \mathsf{CN}(\mathbf{0};\, \boldsymbol{\Omega}_{\boldsymbol{h}_k})$, for all $t = \tau_{\mathsf{tr}} + 1, \ldots, T_{\mathsf{coh}}$.

**Example 8.** Let $\varphi_{\mathsf{app}}$ be the identity operator, so that $\mathbb{Q}^{(\ell-1)}_{\mathsf{app}}(\tilde{x}_{k,t}) = \mathbb{P}^{(\ell-1)}_{\mathsf{app}}(x_{k,t})$, for all $k \in \mathcal{K}$ and $t \in \mathcal{D} \setminus \vartheta$. Furthermore, let $\tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta\boldsymbol{u}_{k,m}} = \boldsymbol{\Omega}^{(\ell)}_{\Delta\boldsymbol{u}_{k,m}}$ and $\tilde{\sigma}^2 = \sigma^2$. The GPME (4.9) with (4.26) is then the LMMSE channel estimator for (2.10), given $\{\mathcal{P}, \mathcal{S}, \mathcal{Y} \setminus \boldsymbol{y}_{\vartheta}, \{\langle\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}\rangle^{(\ell-1)}_{\mathsf{app}}\}_{k=1}^{K}, \{\boldsymbol{\Omega}^{(\ell)}_{\Delta\boldsymbol{u}_{k,m}}\}_{k=1}^{K}\}$. ♢

**Example 9.** Postulate $\tilde{\sigma}^2 = 0$ and $\tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta\boldsymbol{u}_{k,m}} = \mathbf{0}$. Let

$$\hat{x}^{(\ell-1)}_{k,t} = \underset{x_{k,t}\in\mathcal{M}}{\arg\max}\ \mathbb{P}^{(\ell-1)}_{\mathsf{app}}(x_{k,t}), \qquad \forall t \in \mathcal{D} \setminus \vartheta, \tag{4.27}$$

and define

$$\varphi_{\mathsf{app}} : \mathbb{P}^{(\ell-1)}_{\mathsf{app}}(x_{k,t}) \mapsto \delta_{\tilde{x}_{k,t}}(\hat{x}^{(\ell-1)}_{k,t}), \qquad \tilde{x}_{k,t} = x_{k,t}. \tag{4.28}$$

The symbols $\langle\tilde{x}_{k,t}\rangle^{(\ell-1)}_{\mathsf{app}}$ defined in (4.10) represent now hard feedback symbols and the estimator treats them as error free pilots since the error covariance is neglected. Thus, (4.9) with (4.26) yields the hard feedback based "maximum likelihood" channel estimator studied approximately in [138]. ♢

### 4.1.4 Iterative MAP Detector

In this section, we consider a non-linear data estimator that uses the extrinsic probabilities from the single-user decoders to iteratively approximate the prior probabilities of the transmitted symbols of the interfering users (see [126, Eqs. (14) – (15)]).

Consider the data transmission part $t \in \mathcal{D}$ in (2.10) and denote the user of interest by $\xi \in \mathcal{K}$. Let (4.4) be available to the data estimator and postulate a new channel model for the $\ell$th iteration

$$\tilde{\boldsymbol{y}}_t[c] = \frac{1}{\sqrt{L}} \boldsymbol{S}_{\xi,t} \tilde{\boldsymbol{h}}_{\xi,t}[c] \tilde{x}_{\xi,t}[c] + \frac{1}{\sqrt{L}} \sum_{j \in \mathcal{K} \setminus \xi} \boldsymbol{S}_{j,t} \tilde{\boldsymbol{h}}_{j,t}[c] \tilde{x}_{j,t}[c] + \tilde{\boldsymbol{w}}_t[c] \in \mathbb{C}^L, \quad (4.29)$$

where the postulated prior of the user of interest is $\mathbb{Q}(\tilde{x}_{\xi,t}[c]) = \mathbb{P}(x_{\xi,t}[c])$, and the data symbols of the interfering users have postulated a priori probabilities

$$\mathbb{Q}^{(\ell)}(\tilde{x}_{j,t}[c]) = \mathbb{Q}^{(\ell-1)}_{\mathsf{ext}}(\tilde{x}_{j,t}[c]), \qquad \forall j \in \mathcal{K} \setminus \xi. \quad (4.30)$$

As before, we let $\tilde{\sigma}^2$ to be the postulated noise variance and $\tilde{\boldsymbol{w}}_t[c] \sim \mathsf{CN}(\boldsymbol{0};\, \tilde{\sigma}^2 \boldsymbol{I}_L)$ are IID Gaussian RVs. After omitting the block index $c$ for notational simplicity, the iterative non-linear data estimator for the $\xi$th user calculates the probabilities

$$\begin{aligned} &\mathbb{Q}^{(\ell)}(\tilde{x}_{\xi,t} \mid \mathcal{I}^{(\ell)}_{\xi,t}) \\ &\quad = \frac{\mathbb{Q}(\tilde{x}_{\xi,t}) \sum\limits_{\{\tilde{x}_{j,t}\}} \prod\limits_{j \in \mathcal{K} \setminus \xi} \mathbb{Q}^{(\ell)}(\tilde{x}_{j,t}) \mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t \mid \{\tilde{x}_{k,t}\}_{k=1}^K, \mathcal{I}^{(\ell)}_{\xi,t})}{\sum\limits_{\tilde{x}_{\xi,t} \in \mathcal{M}} \mathbb{Q}(\tilde{x}_{\xi,t}) \sum\limits_{\{\tilde{x}_{j,t}\}} \prod\limits_{j \in \mathcal{K} \setminus \xi} \mathbb{Q}^{(\ell)}(\tilde{x}_{j,t}) \mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t \mid \{\tilde{x}_{k,t}\}_{k=1}^K, \mathcal{I}^{(\ell)}_{\xi,t})}, \end{aligned} \quad (4.31)$$

where the summations are over the symbols whose a priori probabilities were given in (4.30). The channel estimates are introduced via

$$\begin{aligned} &\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t \mid \{\tilde{x}_{k,t}\}_{k=1}^K, \mathcal{I}^{(\ell)}_{\xi,t}) \\ &\quad = \int \mathbb{Q}(\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t \mid \{\tilde{\boldsymbol{h}}_{k,t}\}_{k=1}^K, \{\tilde{x}_{k,t}\}_{k=1}^K, \mathcal{I}^{(\ell)}_{\xi,t}) \prod_{k=1}^K \mathrm{d}\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{h}}_{k,t} \mid \mathcal{I}^{(\ell)}_t), \end{aligned} \quad (4.32)$$

where the CSI $\{\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{h}}_{k,t} \mid \mathcal{I}^{(\ell)}_t)\}_{k=1}^K$ is provided by the channel estimator. The obtained probabilities $\{\mathbb{Q}^{(\ell)}(\tilde{x}_{k,t} \mid \mathcal{I}^{(\ell)}_{k,t}) \mid \forall k \in \mathcal{K}, t \in \mathcal{D}\}$ for all fading blocks are then forwarded to the respective sum-product decoders of the users.

**Example 10.** Let $\tilde{\sigma}^2 = \sigma^2$ and $\varphi_{\mathsf{ext}}$ be the identity operator, i.e., $\mathbb{Q}_{\mathsf{ext}}^{(\ell-1)}(\tilde{x}_{k,t}) = \mathbb{P}_{\mathsf{ext}}^{(\ell-1)}(x_{k,t})\,\forall k$. We call the iterative MUDD consisting of this non-linear data estimator and the bank of single-user sum-product decoders, the iterative MAP-MUDD. $\diamondsuit$

Note that the iterative MAP-MUDD uses directly all the information provided by the channel estimator and the extrinsic probabilities obtained during the previous iteration. This gives an upper bound for the performance of the class of iterative MUDDs with the same feedback and channel information, but requires $O(|\mathcal{M}|^K)$ summations for each estimated symbol. This is too high for large practical systems and, therefore, lower complexity parallel interference cancellation approach is considered in the next section.

### 4.1.5 Iterative Multiuser Detection and Decoding with Parallel Interference Cancellation

Having obtained the optimum, but computationally complex, MAP detector in previous section, we next consider how to estimate the data symbol $x_{\xi,t}$ of the $\xi$th user at the $\ell$th iteration by using linear filtering and parallel interference cancellation. The derived class of estimators avoids the exponential complexity in the number of users present in the system and are therefore more suitable for practical applications. For notational convenience we drop again the block index $c = 1, \ldots, C$.

Consider the $\ell$th iteration and let $\xi \in \mathcal{K}$ be the user of interest. Recall that $\mathcal{I}_{\xi,t}^{(\ell)}$ defined in (4.4) is available to the data estimator and assume that the posterior mean estimates of the channel coefficients of all users and the data symbols of the interfering users, i.e.,

$$\langle \tilde{\boldsymbol{h}}_{k,t} \rangle_{(\ell)} = \int \tilde{\boldsymbol{h}}_{k,t} \mathrm{d}\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{h}}_{k,t} \mid \mathcal{I}_t^{(\ell)}), \qquad \forall k \in \mathcal{K}, \tag{4.33}$$

$$\langle \tilde{x}_{j,t} \rangle_{\mathsf{ext}}^{(\ell-1)} = \sum_{\tilde{x}_{j,t} \in \mathcal{M}} \tilde{x}_{j,t} \mathbb{Q}_{\mathsf{ext}}^{(\ell-1)}(\tilde{x}_{j,t}), \qquad \forall j \in \mathcal{K} \setminus \xi, \tag{4.34}$$

respectively, have been calculated with the help of $\mathcal{I}_{\xi,t}^{(\ell)}$. Define also the RVs

$$\Delta x_{j,t} = x_{j,t} - \langle \tilde{x}_{j,t} \rangle_{\mathsf{ext}}^{(\ell-1)}, \qquad \forall j \in \mathcal{K} \setminus \xi, \tag{4.35}$$

$$\Delta \boldsymbol{h}_{k,t} = \boldsymbol{h}_k - \langle \tilde{\boldsymbol{h}}_{k,t} \rangle_{(\ell)}, \qquad \forall k \in \mathcal{K}, \tag{4.36}$$

$$\Delta \boldsymbol{v}_{k,t} = \Delta \boldsymbol{h}_{k,t} x_{k,t}, \qquad \forall k \in \mathcal{K}, \tag{4.37}$$

which are all zero-mean in the limit of large code word length and given $\mathcal{I}_{\xi,t}^{(\ell)}$.

Let $\tilde{\Omega}^{(\ell-1)}_{\Delta x_{j,t}}$ and $\tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta \boldsymbol{v}_{k,t}}$ be the receiver's knowledge about the true covariances $\Omega^{(\ell-1)}_{\Delta x_{j,t}}$ and $\boldsymbol{\Omega}^{(\ell)}_{\Delta \boldsymbol{v}_{k,t}}$ of (4.35) and (4.37), respectively. Denote the postulated noise variance in (2.10) by $\tilde{\sigma}^2$. In Appendix B, we derive the marginalized posterior probabilities of the data symbol $x_{\xi,t}$, that will be shown in Remark 4 to be analogous the outputs of linear filtering and parallel interference cancellation at the receiver. The posterior probabilities are given by (see Appendix B for details)

$$\mathbb{Q}^{(\ell)}(\tilde{x}_{\xi,t} \mid \mathcal{I}^{(\ell)}_{\xi,t}) = \frac{\mathbb{Q}(\tilde{x}_{\xi,t})\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t \mid \tilde{x}_{\xi,t}, \mathcal{I}^{(\ell)}_{\xi,t})}{\mathbb{E}_{\tilde{x}_{\xi,t}}\{\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t \mid \tilde{x}_{\xi,t}, \mathcal{I}^{(\ell)}_{\xi,t})\}}, \tag{4.38}$$

where $\mathbb{Q}(\tilde{x}_{\xi,t})$ is the postulated a priori distribution for the desired user's data symbols and the postulated channel model reads

$$\tilde{\boldsymbol{y}}_t = \frac{1}{\sqrt{L}}\boldsymbol{S}_{\xi,t}\langle\tilde{\boldsymbol{h}}_{\xi,t}\rangle_{(\ell)}\tilde{x}_{\xi,t} + \frac{1}{\sqrt{L}}\sum_{j\in\mathcal{K}\setminus\xi}\boldsymbol{S}_{j,t}\langle\tilde{\boldsymbol{h}}_{j,t}\rangle_{(\ell)}\langle\tilde{x}_{j,t}\rangle_{(\ell)} + \tilde{\boldsymbol{w}}^{\mathsf{pic},(\ell)}_{\xi}, \tag{4.39}$$

where $\tilde{\boldsymbol{w}}^{\mathsf{pic},(\ell)}_{\xi} \sim \mathsf{CN}(\boldsymbol{0};\, \tilde{\boldsymbol{\Omega}}^{\mathsf{pic},(\ell)}_{\xi})$ and the modified noise covariance is given by

$$\begin{aligned}\tilde{\boldsymbol{\Omega}}^{\mathsf{pic},(\ell)}_{\xi} = \tilde{\sigma}^2\boldsymbol{I}_L &+ \frac{1}{L}\boldsymbol{S}_{\xi,t}\tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta\boldsymbol{v}_{\xi,t}}\boldsymbol{S}^{\mathsf{H}}_{\xi,t} \\ &+ \frac{1}{L}\sum_{j\in\mathcal{K}\setminus\xi}\boldsymbol{S}_{j,t}\left(\tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta\boldsymbol{v}_{j,t}} + \langle\tilde{\boldsymbol{h}}_{j,t}\rangle_{(\ell)}\tilde{\Omega}^{(\ell-1)}_{\Delta x_{j,t}}\langle\tilde{\boldsymbol{h}}^{\mathsf{H}}_{j,t}\rangle_{(\ell)}\right)\boldsymbol{S}^{\mathsf{H}}_{j,t}.\end{aligned} \tag{4.40}$$

Note that when the second term on the RHS of (4.39) is moved to the left hand side (LHS) of the equation, we get the parallel interference cancellation (PIC) operation. The probabilities (4.38) are parametrized by:

1. $\varphi_{\mathsf{ext}} : \mathbb{P} \mapsto \mathbb{Q}$, defines the type of interference cancellation (soft / hard);
2. $\tilde{\Omega}^{(\ell-1)}_{\Delta x_{j,t}}$ and $\tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta\boldsymbol{v}_{k,t}}$, quantify the estimator's knowledge about the error statistics;
3. $\tilde{\sigma}^2$, defines the type of linear filtering used by the data estimator.

By choosing these parameters appropriately, all the usual iterative data estimators using PIC can be obtained, as will be shown below.

**Remark 4.** The GPME based on (4.31) or (4.38) is given by

$$\langle\tilde{x}_{\xi,t}\rangle_{(\ell)} = \int \tilde{x}_{\xi,t}\,\mathrm{d}\mathbb{Q}^{(\ell)}(\tilde{x}_{\xi,t} \mid \mathcal{I}^{(\ell)}_{\xi,t}), \qquad \xi \in \mathcal{K}. \tag{4.41}$$

When iterative decoding is considered, however, instead of the posterior mean (4.41), the BCJR algorithm needs in fact the probabilities (4.31) or (4.38) with the correct

prior $\mathbb{Q}(\tilde{x}_{\xi,t}) = \mathbb{P}(x_{\xi,t})$ for the user of interest. Under QPSK and Gray mapping, it is trivial to obtain the desired probabilities from the posterior mean estimate (4.41) if non-linear estimator (4.31) is used. A problem arises, however, when the GPME is represented by a linear filter as we shall see next.

Consider the case where Gaussian prior $\mathbb{Q}(\tilde{x}_{\xi,t}) = \mathsf{CN}(0;\, 1)$ is postulated for the user of interest $\xi$. Then, the GPME (4.41) with the probabilities (4.38) simplifies the to the familiar form

$$\langle \tilde{x}_{\xi,t} \rangle_{(\ell)} = \boldsymbol{m}_{\xi,t}^{\mathsf{H}} \tilde{\boldsymbol{y}}_{\xi,t} = \frac{\frac{1}{\sqrt{L}} \langle \tilde{\boldsymbol{h}}_{\xi,t} \rangle_{(\ell)}^{\mathsf{H}} \boldsymbol{S}_{\xi,t}^{\mathsf{H}} (\tilde{\boldsymbol{\Omega}}_{\xi}^{\mathsf{pic},(\ell)})^{-1}}{1 + \frac{1}{L} \langle \tilde{\boldsymbol{h}}_{\xi,t} \rangle_{(\ell)}^{\mathsf{H}} \boldsymbol{S}_{\xi,t}^{\mathsf{H}} (\tilde{\boldsymbol{\Omega}}_{\xi}^{\mathsf{pic},(\ell)})^{-1} \boldsymbol{S}_{\xi,t} \langle \tilde{\boldsymbol{h}}_{\xi,t} \rangle_{(\ell)}} \tilde{\boldsymbol{y}}_{\xi,t} \in \mathbb{C}, \tag{4.42}$$

where we denoted

$$\tilde{\boldsymbol{y}}_{\xi,t} = \tilde{\boldsymbol{y}}_t - \frac{1}{\sqrt{L}} \sum_{j \in \mathcal{K} \setminus \xi} \boldsymbol{S}_{j,t} \langle \tilde{\boldsymbol{h}}_{j,t} \rangle_{(\ell)} \langle \tilde{x}_{j,t} \rangle_{(\ell)}. \tag{4.43}$$

By selecting the parameters appropriately, $\boldsymbol{m}_{\xi,t}^{\mathsf{H}} \in \mathbb{C}^{1 \times L}$ becomes, e.g., the the LMMSE estimator or SUMF with PIC. This form does not, however, produce the information desired by the decoders as remarked earlier. An easy solution to this problem exists if the CDMA system under consideration is sufficiently large. Then the output of a linear data estimator is in general accurately approximated by the Gaussian distribution [55, 56, 197]. The approximate symbol probabilities can thus be obtained by considering $\langle \tilde{x}_{\xi,t} \rangle_{(\ell)}$ to be the output of a channel

$$\langle \tilde{x}_{\xi,t} \rangle_{(\ell)} = \boldsymbol{m}_{\xi,t}^{\mathsf{H}} \boldsymbol{S}_{\xi,t} \langle \tilde{\boldsymbol{h}}_{\xi,t} \rangle_{(\ell)} x_{\xi,t} + \tilde{w}_{\xi,t}, \qquad \tilde{w}_{\xi,t} \sim \mathsf{CN}(0;\, \boldsymbol{m}_{\xi,t}^{\mathsf{H}} \tilde{\boldsymbol{\Omega}}_{\xi}^{\mathsf{pic},(\ell)} \boldsymbol{m}_{\xi,t}), \tag{4.44}$$

where $\mathbb{P}(x_{k,t})$ is given in (2.19). See, e.g., [117] for an example of this approach. Since we study large systems, in the following the outputs (4.38) and (4.42) are considered to be equivalent. ♢

**Example 11.** Consider estimating the $\xi$th user and let the postulated noise variance be correct $\tilde{\sigma}^2 = \sigma^2$. Define $\varphi_{\mathsf{ext}}$ to be the identity operator, so that, $\mathbb{Q}_{\mathsf{ext}}^{(\ell-1)}(\tilde{x}_{\xi,t}) = \mathbb{P}_{\mathsf{ext}}^{(\ell-1)}(x_{\xi,t})\, \forall j \in \mathcal{K} \setminus \xi$. Given $\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_{k,t}}^{(\ell)} = \boldsymbol{\Omega}_{\Delta \boldsymbol{v}_{k,t}}^{(\ell)}$, for all $k = 1, \ldots, K$, the GPME (4.41) with posterior probabilities (4.38) is an extension of the LMMSE data estimator studied in [59] to include soft PIC. We call this data estimator the LMMSE-PIC MUDD for the rest of the chapter. ♢

**Example 12.** Let $\tilde{\sigma}^2 \to \infty$ and define $\varphi_{\mathsf{ext}}$ to be the identity operator or

$$\varphi_{\mathsf{ext}} : \mathbb{P}_{\mathsf{ext}}^{(\ell)}(x_{k,t}) \mapsto \delta_{\tilde{x}_{k,t}}(\hat{x}_{k,t}), \qquad \tilde{x}_{k,t} = x_{k,t}, \tag{4.45}$$

where

$$\hat{x}_{k,t} = \underset{x_{k,t}\in\mathcal{M}}{\arg\max}\ \mathbb{P}^{(\ell)}_{\mathsf{ext}}(x_{k,t}). \tag{4.46}$$

Then, the GPME (4.41) with marginal probabilities (4.38) reduces to the "SUMF-Based Soft IC" and the "Hard-IC" receivers, respectively, studied under the assumption of perfect CSI in [125, Proposition 2]. ♢

## 4.2 Decoupling Results

In this section, the single-user characterization for the multiuser system using the iterative multiuser estimators derived in Section 4.1 is presented. The decoupling of the multiuser channel is obtained via an application of the replica method by using the same methodology as in [85, 89, 92]. In the analysis, the standard assumptions in the replica trick are considered to be valid and the replica symmetry is assumed to hold — see Assumptions 6 and 7 in Appendix C, respectively. Before proceeding we make an assumption for the rest of the chapter.

**Assumption 5.** Let $\{\mathcal{K}_u\}_{u=1}^{U}$ be a finite partition of $\mathcal{K}$ into $U$ user groups. Furthermore, let all users in the same group have equal power delay profiles. ♢

We also note that the replica method relies on the large system limit where $K = \alpha L \to \infty$ with fixed system load $0 < \alpha < \infty$ and number of user groups $U$.

### 4.2.1 Linear Channel Estimation with Information Feedback

Let us consider user $k \in \mathcal{K}$ in the multiuser system defined in Sections 2.2.2 and 4.1. Fix the time instant $\vartheta \in \mathcal{D}$ and let the channel estimator be at its $\ell$th iteration. Define a set of single-user channels, indexed by $m = 1, \ldots, M$ where $M$ is the number of multipaths, during the training and the data transmission phases

$$\boldsymbol{z}_{k,\mathcal{T},m} = \boldsymbol{p}_k h_{k,m} + \boldsymbol{w}_{k,\mathcal{T},m} \in \mathbb{C}^{\tau_{\mathsf{tr}}}, \boldsymbol{w}_{k,\mathcal{T},m} \sim \mathsf{CN}(\mathbf{0};\, \boldsymbol{C}^{(\ell)}_{\mathcal{T}}), \tag{4.47}$$

$$\boldsymbol{z}_{k,\mathcal{D}\setminus\vartheta,m} = \boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta} h_{k,m} + \boldsymbol{w}_{k,\mathcal{D}\setminus\vartheta,m} \in \mathbb{C}^{\tau_{\mathsf{d}}-1}, \tag{4.48}$$

respectively. The additive noise vectors are zero-mean complex Gaussian $\boldsymbol{w}_{k,\mathcal{T},m} \sim \mathsf{CN}(\mathbf{0};\, \boldsymbol{C}^{(\ell)}_{\mathcal{T}})$ and $\boldsymbol{w}_{k,\mathcal{D}\setminus\vartheta,m} \sim \mathsf{CN}(\mathbf{0};\, \boldsymbol{C}^{(\ell)}_{\mathcal{D}\setminus\vartheta})$ and IID for $m = 1, \ldots, M$. Following the notation of (4.6), $\boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta}$ are the transmitted data symbols for time indices $t \in \mathcal{D} \setminus \vartheta$. Let the postulated channel related to (4.47) – (4.48) be

$$\tilde{\boldsymbol{z}}_{k,\mathcal{T},m} = \boldsymbol{p}_k \tilde{h}_{k,\vartheta,m} + \tilde{\boldsymbol{w}}_{k,\mathcal{T},m} \in \mathbb{C}^{\tau_{\mathsf{tr}}}, \tag{4.49}$$

$$\tilde{\boldsymbol{z}}_{k,\mathcal{D}\setminus\vartheta,m} = \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle^{(\ell-1)}_{\mathsf{app}} \tilde{h}_{k,\vartheta,m} + \Delta\tilde{\boldsymbol{u}}_{k,m} + \tilde{\boldsymbol{w}}_{k,\mathcal{D}\setminus\vartheta,m} \in \mathbb{C}^{(\tau_{\mathsf{d}}-1)}, \tag{4.50}$$

where $\tilde{\boldsymbol{w}}_{k,\mathcal{T},m} \sim \mathsf{CN}(\boldsymbol{0};\, \tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)})$ and $\tilde{\boldsymbol{w}}_{k,\mathcal{D}\setminus\vartheta,m} \sim \mathsf{CN}(\boldsymbol{0};\, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)})$ are IID complex Gaussian RVs for $m = 1, \ldots, M$ with postulated covariances $\tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)}$ and $\tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)}$. The feedback vector $\langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}}^{(\ell-1)}$ is defined in (4.11) and $\Delta\tilde{\boldsymbol{u}}_{k,m} \sim \mathsf{CN}(\boldsymbol{0};\, \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{u}_{k,m}}^{(\ell)})$ corresponds to the receiver's knowledge about $\Delta \boldsymbol{u}_{k,m}$ defined in (4.12). We also define a set

$$\mathcal{J}_{k,\vartheta,m}^{(\ell)} = \Big\{ \boldsymbol{z}_{k,\mathcal{T},m}, \boldsymbol{z}_{k,\mathcal{D}\setminus\vartheta,m}, \boldsymbol{p}_k, \{\mathbb{Q}_{\mathsf{app}}^{(\ell-1)}(\tilde{x}_{k,t})\}_{t\in\mathcal{D}\setminus\vartheta} \Big\}, \tag{4.51}$$

and conditional expectation

$$\mathbb{E}_k^{\mathsf{c}}\{\cdots\} = \mathbb{E}\Big\{ \cdots \mid \boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta}, \{\mathcal{J}_{k,\vartheta,m}^{(\ell)}\}_{m=1}^{M} \Big\}, \tag{4.52}$$

for notational convenience. The noise covariances in (4.47) – (4.50) are then given by

$$\boldsymbol{C}_{\mathcal{T}}^{(\ell)} = \sigma^2 \boldsymbol{I}_{\tau_{\mathsf{tr}}} + \alpha \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \boldsymbol{\Sigma}_{k,\mathcal{T},m}(\boldsymbol{C}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)}), \tag{4.53}$$

$$\boldsymbol{C}_{\mathcal{D}\setminus\vartheta}^{(\ell)} = \sigma^2 \boldsymbol{I}_{\tau_{\mathsf{d}}-1} + \alpha \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \boldsymbol{\Sigma}_{k,\mathcal{D}\setminus\vartheta,m}(\boldsymbol{C}_{\mathcal{D}\setminus\vartheta}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)}), \tag{4.54}$$

$$\tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)} = \tilde{\sigma}^2 \boldsymbol{I}_{\tau_{\mathsf{tr}}} + \alpha \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \tilde{\boldsymbol{\Sigma}}_{k,\mathcal{T},m}(\tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)}), \tag{4.55}$$

$$\tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)} = \tilde{\sigma}^2 \boldsymbol{I}_{\tau_{\mathsf{d}}-1} + \alpha \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \tilde{\boldsymbol{\Sigma}}_{k,\mathcal{D}\setminus\vartheta,m}(\tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)}), \tag{4.56}$$

respectively, where

$$\begin{aligned}
&\boldsymbol{\Sigma}_{k,\mathcal{T},m}(\boldsymbol{C}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)}) \\
&\quad = \mathbb{E}_k^{\mathsf{c}} \Big\{ \Big( \boldsymbol{u}_{k,\mathcal{T},m} - \langle \tilde{\boldsymbol{u}}_{k,\mathcal{T},m} \rangle_{k,m}^{(\ell)} \Big) \Big( \boldsymbol{u}_{k,\mathcal{T},m} - \langle \tilde{\boldsymbol{u}}_{k,\mathcal{T},m} \rangle_{k,m}^{(\ell)} \Big)^{\mathsf{H}} \Big\},
\end{aligned} \tag{4.57}$$

$$\begin{aligned}
&\boldsymbol{\Sigma}_{k,\mathcal{D}\setminus\vartheta,m}(\boldsymbol{C}_{\mathcal{D}\setminus\vartheta}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)}) \\
&\quad = \mathbb{E}_k^{\mathsf{c}} \Big\{ \Big( \boldsymbol{u}_{k,\mathcal{D}\setminus\vartheta,m} - \langle \tilde{\boldsymbol{u}}_{k,\mathcal{D}\setminus\vartheta,m} \rangle_{k,m}^{(\ell)} \Big) \Big( \boldsymbol{u}_{k,\mathcal{D}\setminus\vartheta,m} - \langle \tilde{\boldsymbol{u}}_{k,\mathcal{D}\setminus\vartheta,m} \rangle_{k,m}^{(\ell)} \Big)^{\mathsf{H}} \Big\},
\end{aligned} \tag{4.58}$$

$$\begin{aligned}
&\tilde{\boldsymbol{\Sigma}}_{k,\mathcal{T},m}(\tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)}) \\
&\quad = \mathbb{E}_k^{\mathsf{c}} \Big\{ \Big( \tilde{\boldsymbol{u}}_{k,\mathcal{T},m} - \langle \tilde{\boldsymbol{u}}_{k,\mathcal{T},m} \rangle_{k,m}^{(\ell)} \Big) \Big( \tilde{\boldsymbol{u}}_{k,\mathcal{T},m} - \langle \tilde{\boldsymbol{u}}_{k,\mathcal{T},m} \rangle_{k,m}^{(\ell)} \Big)^{\mathsf{H}} \Big\},
\end{aligned} \tag{4.59}$$

$$\begin{aligned}
&\tilde{\boldsymbol{\Sigma}}_{k,\mathcal{D}\setminus\vartheta,m}(\tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)}) \\
&\quad = \mathbb{E}_k^{\mathsf{c}} \Big\{ \Big( \tilde{\boldsymbol{u}}_{k,\mathcal{D}\setminus\vartheta,m} - \langle \tilde{\boldsymbol{u}}_{k,\mathcal{D}\setminus\vartheta,m} \rangle_{k,m}^{(\ell)} \Big) \Big( \tilde{\boldsymbol{u}}_{k,\mathcal{D}\setminus\vartheta,m} - \langle \tilde{\boldsymbol{u}}_{k,\mathcal{D}\setminus\vartheta,m} \rangle_{k,m}^{(\ell)} \Big)^{\mathsf{H}} \Big\}.
\end{aligned} \tag{4.60}$$

We used the shorthand notation

$$\boldsymbol{u}_{k,\mathcal{T},m} = \boldsymbol{p}_k h_{k,m}, \qquad \tilde{\boldsymbol{u}}_{k,\mathcal{T},m} = \boldsymbol{p}_k \tilde{h}_{k,\vartheta,m}, \tag{4.61}$$

$$\boldsymbol{u}_{k,\mathcal{D}\setminus\vartheta,m} = \boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta} h_{k,m}, \qquad \tilde{\boldsymbol{u}}_{k,\mathcal{D}\setminus\vartheta,m} = \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}}^{(\ell-1)} \tilde{h}_{k,\vartheta,m} + \Delta\tilde{\boldsymbol{u}}_{k,m}, \tag{4.62}$$

above and defined the single-user GPME

$$\langle \cdots \rangle_{k,m}^{(\ell)} = \frac{\mathbb{E}_{\tilde{h}_{k,\vartheta,m},\Delta\tilde{\boldsymbol{u}}_{k,m}} \{ \cdots \, \mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{z}}_{k,\mathcal{T},m}, \tilde{\boldsymbol{z}}_{k,\mathcal{D}\setminus\vartheta,m} \mid \tilde{h}_{k,\vartheta,m}, \Delta\tilde{\boldsymbol{u}}_{k,m}, \mathcal{J}_{k,\vartheta,m}^{(\ell)}) \}}{\mathbb{E}_{\tilde{h}_{k,\vartheta,m},\Delta\tilde{\boldsymbol{u}}_{k,m}} \{ \mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{z}}_{k,\mathcal{T},m}, \tilde{\boldsymbol{z}}_{k,\mathcal{D}\setminus\vartheta,m} \mid \tilde{h}_{k,\vartheta,m}, \Delta\tilde{\boldsymbol{u}}_{k,m}, \mathcal{J}_{k,\vartheta,m}^{(\ell)}) \}}, \tag{4.63}$$

where

$$\begin{aligned} &\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{z}}_{k,\mathcal{T},m}, \tilde{\boldsymbol{z}}_{k,\mathcal{D}\setminus\vartheta,m} \mid \tilde{h}_{k,\vartheta,m}, \Delta\tilde{\boldsymbol{u}}_{k,m}, \mathcal{J}_{k,\vartheta,m}^{(\ell)}) \\ &\quad = \mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{z}}_{k,\mathcal{T},m} = \boldsymbol{z}_{k,\mathcal{T},m} \mid \tilde{h}_{k,\vartheta,m}, \mathcal{J}_{k,\vartheta,m}^{(\ell)}) \\ &\qquad \times \mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{z}}_{k,\mathcal{D}\setminus\vartheta,m} = \boldsymbol{z}_{k,\mathcal{D}\setminus\vartheta,m} \mid \tilde{h}_{k,\vartheta,m}, \Delta\tilde{\boldsymbol{u}}_{k,m}, \mathcal{J}_{k,\vartheta,m}^{(\ell)}). \end{aligned} \tag{4.64}$$

**Claim 4.** *Let $T = \tau_{\mathsf{d}} C \to \infty$ and $K = \alpha L \to \infty$ with $\alpha$ and $\tau_{\mathsf{d}}$ finite and fixed. Also, let $\vartheta \in \mathcal{D}$ be an arbitrary time index during the data transmission phase, and $\ell = 1, 2, \ldots$ the iteration index. Conditioned on the set $\{\{\boldsymbol{p}_k\}_{k=1}^K, \{\boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta}\}_{k=1}^K, \{\mathbb{Q}_{\mathsf{app}}^{(\ell-1)}(\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta})\}_{k=1}^K, p_{\mathsf{pdp}}\}$, the joint distribution of the true and postulated channel coefficients and the estimates $\{\langle \tilde{h}_{k,\vartheta,M} \rangle_{(\ell)}\}_{m=1}^M$ of the multiuser system in Section 4.1.3 converges in probability to the joint distribution of the true and postulated channel coefficients and the estimates $\{\langle \tilde{h}_{k,\vartheta,m} \rangle_{k,m}^{(\ell)}\}_{m=1}^M$ of the above single-user system.*

**Proof:** See Appendix C. □

### 4.2.2 Iterative MAP Detector

Let us consider the $\xi$th user in the multiuser system discussed in the previous section. Fix the time instant $t \in \mathcal{D}$ and let the iterative MUDD be at its $\ell$th iteration. Define a set of $M$ single-user channels

$$z_{\xi,t,m} = h_{\xi,m} x_{\xi,t} + w_{\xi,t,m}, \qquad m = 1, \ldots, M, \tag{4.65}$$

where $w_{\xi,t,m} \sim \mathsf{CN}(0;\, D_t^{(\ell)})$ are IID complex Gaussian RVs. Let

$$\tilde{z}_{\xi,t,m} = \tilde{h}_{\xi,t,m} \tilde{x}_{\xi,t} + \tilde{w}_{\xi,t,m}, \qquad m = 1, \ldots, M, \tag{4.66}$$

be the receiver's knowledge about (4.65), where $\tilde{w}_{\xi,t,m} \sim \mathsf{CN}(0;\, \tilde{D}_t^{(\ell)})\ \forall m$ are IID complex Gaussian RVs with postulated noise covariance $\tilde{D}_t^{(\ell)}$. We also define for notational convenience the conditional expectation

$$\mathbb{E}_k^{\mathsf{d}}\{\cdots\} = \mathbb{E}\left\{\cdots \mid \{h_{k,m}\}_{m=1}^M, \mathcal{J}_{k,t}^{(\ell)}\right\}, \qquad k = 1, \ldots, K, \tag{4.67}$$

where

$$\mathcal{J}_{\xi,t}^{(\ell)} = \left\{\{z_{\xi,t,m}\}_{m=1}^M, \{\mathbb{Q}^{(\ell)}(\tilde{h}_{\xi,t,m} \mid \mathcal{I}_t^{(\ell)})\}_{m=1}^M\right\}, \tag{4.68}$$

and

$$\mathcal{J}_{j,t}^{(\ell)} = \left\{\{z_{j,t,m}\}_{m=1}^M, \{\mathbb{Q}^{(\ell)}(\tilde{h}_{j,t,m} \mid \mathcal{I}_t^{(\ell)})\}_{m=1}^M, \mathbb{Q}_{\mathsf{ext}}^{(\ell-1)}(\tilde{x}_{j,t})\right\}, \tag{4.69}$$

for $j \in \mathcal{K} \setminus \xi$. The noise variances in (4.65) and (4.66) are given by the fixed point equations

$$D_t^{(\ell)} = \sigma^2 + \alpha \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^K \sum_{m=1}^M \Sigma_{k,t,m}(D_t^{(\ell)}, \tilde{D}_t^{(\ell)}), \tag{4.70}$$

$$\tilde{D}_t^{(\ell)} = \tilde{\sigma}^2 + \alpha \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^K \sum_{m=1}^M \tilde{\Sigma}_{k,t,m}(D_t^{(\ell)}, \tilde{D}_t^{(\ell)}), \tag{4.71}$$

where

$$\Sigma_{k,t,m}(D_t^{(\ell)}, \tilde{D}_t^{(\ell)}) = \mathbb{E}_k^{\mathsf{d}}\{|h_{k,m}x_{k,t} - \langle \tilde{h}_{k,t,m}\tilde{x}_{k,t}\rangle_k^{(\ell)}|^2\}, \tag{4.72}$$

$$\tilde{\Sigma}_{k,t,m}(D_t^{(\ell)}, \tilde{D}_t^{(\ell)}) = \mathbb{E}_k^{\mathsf{d}}\{|\tilde{h}_{k,t,m}\tilde{x}_{k,t} - \langle \tilde{h}_{k,t,m}\tilde{x}_{k,t}\rangle_k^{(\ell)}|^2\}, \tag{4.73}$$

and the notation $\langle\cdots\rangle_k^{(\ell)}$ in (4.72) – (4.73) denotes for the single-user GPME

$$\langle\cdots\rangle_k^{(\ell)} = \sum_{\tilde{x}_{k,t}\in\mathcal{M}} \mathbb{Q}(\tilde{x}_{k,t}) \times \frac{\int\cdots\prod_{m=1}^M \mathbb{Q}(\tilde{z}_{k,t,m} = z_{k,t,m} \mid \tilde{h}_{k,t,m}, \tilde{x}_{k,t}, \mathcal{J}_{k,t}^{(\ell)}) \mathrm{d}\mathbb{Q}^{(\ell)}(\tilde{h}_{k,t,m} \mid \mathcal{I}_t^{(\ell)})}{\sum_{\tilde{x}_{k,t}\in\mathcal{M}} \mathbb{Q}(\tilde{x}_{k,t}) \int \mathbb{Q}(\tilde{z}_{k,t,m} = z_{k,t,m} \mid \tilde{h}_{k,t,m}, \tilde{x}_{k,t}, \mathcal{J}_{k,t}^{(\ell)}) \mathrm{d}\mathbb{Q}^{(\ell)}(\tilde{h}_{k,t,m} \mid \mathcal{I}_t^{(\ell)})} \tag{4.74}$$

of the user $k = 1, \ldots, K$. Furthermore, the a priori probabilities of the data symbols in (4.72) – (4.73) are given for the desired user by $\mathbb{Q}(\tilde{x}_{\xi,t}) = \mathbb{P}(x_{\xi,t} = \tilde{x}_{\xi,t})$ (see (2.19) – (2.20)) and for the interfering users by

$$\mathbb{Q}^{(\ell)}(\tilde{x}_{j,t}) = \mathbb{Q}_{\mathsf{ext}}^{(\ell-1)}(\tilde{x}_{j,t}), \qquad j \in \mathcal{K} \setminus \xi. \tag{4.75}$$

The channel information $\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{h}}_{k,t} \mid \mathcal{I}_t^{(\ell)}) = \prod_{m=1}^{M} \mathbb{Q}^{(\ell)}(\tilde{h}_{k,t,m} \mid \mathcal{I}_t^{(\ell)}), k = 1, \ldots, K$, is provided by the channel estimator.

**Claim 5.** *Let $T = \tau_{\mathsf{d}} C \to \infty$ and $K = \alpha L \to \infty$ with $\alpha$ and $\tau_{\mathsf{d}}$ finite and fixed. Conditioned on the set $\{\mathcal{P}, \mathcal{H}, \{\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{h}}_{k,t} \mid \mathcal{I}_t^{(\ell)})\}_{k=1}^{K}, \{\mathbb{Q}_{\mathsf{ext}}^{(\ell-1)}(\tilde{x}_{j,t})\}_{j \in \mathcal{K} \setminus \xi}, p_{\mathsf{pdp}}\}$, the joint distribution of the true and postulated inputs and the estimate $\langle \tilde{x}_{\xi,t} \rangle_{(\ell)}$ of the multiuser system in Section 4.1.4 at the $\ell$th iteration converges in probability to the joint distribution of the true and postulated inputs and the estimate $\langle \tilde{x}_{\xi,t} \rangle_{\xi}^{(\ell)}$ of the single-user system described above.*

**Proof:** Omitted. □

### 4.2.3 Iterative Multiuser Detection and Decoding with Parallel Interference Cancellation

Consider the same setup as in Section 4.2.2. Let the true single-user channel be given by (4.65) and postulate for $m = 1, \ldots, M$, the set of channels

$$\tilde{z}_{\xi,t,m} = \langle \tilde{h}_{\xi,t,m} \rangle_{(\ell)} \tilde{x}_{\xi,t} + \Delta\tilde{v}_{\xi,t,m} + \tilde{w}_{\xi,t,m}, \qquad \tilde{w}_{\xi,t,m} \sim \mathsf{CN}(0;\, \tilde{D}_t^{(\ell)}), \quad (4.76)$$

where $\{\langle \tilde{h}_{\xi,t,m} \rangle_{(\ell)}\}_{m=1}^{M}$ are the posterior mean estimates of the channel (4.33) and given by the channel estimator. As in Section 4.2.2, the noise variances are given by (4.70) – (4.71), with (4.72) – (4.73) replaced by

$$\Sigma_{k,t,m}(D_t^{(\ell)}, \tilde{D}_t^{(\ell)}) = \mathbb{E}_k^{\mathsf{d}} \left\{ |h_{k,m} x_{k,t} - \langle \tilde{v}_{k,t,m} \rangle_k^{(\ell)}|^2 \right\}, \quad (4.77)$$

$$\tilde{\Sigma}_{k,t,m}(D_t^{(\ell)}, \tilde{D}_t^{(\ell)}) = \mathbb{E}_k^{\mathsf{d}} \left\{ |\tilde{v}_{k,t,m} - \langle \tilde{v}_{k,t,m} \rangle_k^{(\ell)}|^2 \right\}, \quad (4.78)$$

where $k = 1, \ldots, K$ and

$$\tilde{v}_{k,t,m} = \langle \tilde{h}_{k,t,m} \rangle_{(\ell)} \tilde{x}_{k,t} + \Delta\tilde{v}_{k,t,m}. \quad (4.79)$$

The RV

$$\Delta\tilde{\boldsymbol{v}}_{k,t} = \left[ \Delta\tilde{v}_{k,t,1} \; \cdots \; \Delta\tilde{v}_{k,t,M} \right]^{\mathsf{T}} \in \mathbb{C}^M, \quad (4.80)$$

is the receiver's knowledge about (4.37), as discussed in Section 4.1.5. The single-user GPME of the $k$th user is given by

$$\langle \cdots \rangle_k^{(\ell)} = \frac{\mathbb{E}_{\tilde{x}_k, \Delta\tilde{\boldsymbol{v}}_{k,t}} \{ \cdots \prod_{m=1}^{M} \mathbb{Q}(\tilde{z}_{k,t,m} = z_{k,t,m} \mid \tilde{x}_{k,t}, \Delta\tilde{v}_{k,t,m}, \mathcal{J}_{k,t}^{(\ell)}) \}}{\mathbb{E}_{\tilde{x}_k, \Delta\tilde{\boldsymbol{v}}_{k,t}} \{ \prod_{m=1}^{M} \mathbb{Q}(\tilde{z}_{k,t,m} = z_{k,t,m} \mid \tilde{x}_{k,t}, \Delta\tilde{v}_{k,t,m}, \mathcal{J}_{k,t}^{(\ell)}) \}}. \quad (4.81)$$

The expectations over the data symbols are taken for the desired user $\xi$ with respect to the postulated a priori probability $\mathbb{Q}(\tilde{x}_{\xi,t})$, and for the other users over the Gaussian distribution

$$\mathbb{Q}^{(\ell)}(\tilde{x}_{j,t}) = \mathsf{CN}(\langle \tilde{x}_{j,t} \rangle_{\mathsf{ext}}^{(\ell-1)}, \tilde{\Omega}_{\Delta x_{j,t}}^{(\ell-1)}), \qquad j \in \mathcal{K} \setminus \xi, \tag{4.82}$$

where the mean is given by (4.34), and $\tilde{\Omega}_{\Delta x_{j,t}}^{(\ell-1)}$ is the postulated variance of the estimation error in the feedback symbols.

**Claim 6.** *Let $T = \tau_{\mathsf{d}} C \to \infty$ and $K = \alpha L \to \infty$ with $\alpha$ and $\tau_{\mathsf{d}}$ finite and fixed. Conditioned on the set $\{\mathcal{P}, \mathcal{H}, \{\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{h}}_{k,t} \mid \mathcal{I}_t^{(\ell)})\}_{k=1}^K, \{\mathbb{Q}_{\mathsf{ext}}^{(\ell-1)}(\tilde{x}_{j,t})\}_{j \in \mathcal{K} \setminus \xi}, p_{\mathsf{pdp}}\}$, the joint distribution of the true and postulated inputs and the estimate $\langle \tilde{x}_{k,t} \rangle_{(\ell)}$ of the multiuser system in Section 4.1.5 at the $\ell$th iteration converges in probability to the joint distribution of the true and postulated inputs and the estimate $\langle \tilde{x}_{k,t} \rangle_{(\ell)}$ of the single-user system described above.*

**Proof:** Omitted. □

The consequence of the Claims 4 – 6 is that the performance of an iterative multiuser DS-CDMA system described in Sections 2.2.2 and 4.1, can be analyzed by concentrating on the equivalent single-user system defined by the appropriate equations in (4.47) – (4.81). The next section reports the performance analysis of the iterative DS-CDMA system based on the equivalent single-user description given above.

## 4.3 Performance of Large DS-CDMA Systems Using Iterative Channel Estimation, Detection, and Decoding

We now turn to the analysis of the multiuser DS-CDMA system described in Section 2.2.2 that uses the iterative estimators derived in Section 4.1. As in Chapter 3, the large system performance is obtained with the help of the decoupling results reported earlier. Thus, for the rest of this chapter the replica symmetric solutions of Claims 4 – 6 are assumed to be valid and all results are obtained by studying the equivalent single-user system defined by (4.47) – (4.81).

Due to the relatively high amount of different results that will follow, the outline of the section is provided below.

- Section 4.3.1 briefly recaps the assumptions made in the density evolution for the rest of the analysis.
- Section 4.3.2 considers the iterative channel estimators described in Section 4.1.3. The output statistics of the following CEs in increasing order of model mismatch are obtained:
    1. Iterative LMMSE channel estimator with soft information feedback and full knowledge of second order error statistics;
    2. Iterative "maximum likelihood" channel estimator using hard feedback and neglecting all error statistics.
- Section 4.3.3 concentrates on the performance analysis of the iterative data decoders introduced in Sections 4.1.4 and 4.1.5. The following iterative estimators, arranged in decreasing order of complexity, are considered:
    1. Iterative MAP-MUDD with soft feedback;
    2. Iterative LMMSE-PIC MUDD with soft feedback;
    3. Iterative SUMF with soft and hard feedback.
- Section 4.3.4 briefly recaps the notion of multiuser efficiency and presents a related performance measure suitable for mismatched channel information.

### 4.3.1 Density Evolution with Gaussian Approximation

Consider the BICM encoded channel inputs $\{x_{k,t}\}$ in (2.10). By (2.19) – (2.20), the extrinsic probabilities of the data symbol $x_{k,t}$, obtained by the BCJR algorithm during the $\ell$th iteration, factor as

$$\mathbb{P}_{\mathsf{ext}}^{(\ell)}\left(x_{k,t}=\frac{1}{\sqrt{2}}(a_{k,t,1}+\mathrm{j}a_{k,t,2})\right)=\mathbb{P}_{\mathsf{ext}}^{(\ell)}(a_{k,t,1})\mathbb{P}_{\mathsf{ext}}^{(\ell)}(a_{k,t,2}), \tag{4.83}$$

in the limit of large code word length $T\to\infty$. For later use, let

$$\hat{a}_{k,t,q}^{\mathsf{ext},(\ell)}=\underset{a_{k,t,q}\in\{\pm 1\}}{\arg\max}\ \mathbb{P}_{\mathsf{ext}}^{(\ell)}(a_{k,t,q}),\qquad q=1,2, \tag{4.84}$$

be the extrinsic information based hard estimate of $a_{k,t,q}$, and define the error probability

$$\varepsilon_{k}^{\mathsf{ext},(\ell)}=\frac{1}{2T}\sum_{j=1}^{2}\sum_{t=1}^{T}\Pr\left(\hat{a}_{k,t,q}^{\mathsf{ext},(\ell)}\neq a_{k,t,q}\right). \tag{4.85}$$

Naturally, equations completely analogous to (4.83) – (4.85) for the approximate a posteriori based feedback can be defined.

To simplify the density evolution, we make the Gaussian approximation [125, 126, 133, 173, 174] for the log-likelihood ratios obtained by the sum-product decoders, i.e., given $a_{k,t,q}, q = 1, 2,$

$$\lambda_{a_{k,t,q}}^{\mathsf{ext},(\ell)} = \log\left(\frac{\mathbb{P}_{\mathsf{ext}}^{(\ell)}(a_{k,t,q}=+1)}{\mathbb{P}_{\mathsf{ext}}^{(\ell)}(a_{k,t,q}=-1)}\right) \sim \mathsf{N}(2a_{k,t,q}\mu_k^{\mathsf{ext},(\ell)}, 4\mu_k^{\mathsf{ext},(\ell)}), \tag{4.86}$$

where

$$\mu_k^{\mathsf{ext},(\ell)} = \left[Q^{-1}(\varepsilon_k^{\mathsf{ext},(\ell)})\right]^2, \tag{4.87}$$

and $Q^{-1}$ is the functional inverse of the $Q$-function [5]. The approximate APPs $\mathbb{P}_{\mathsf{app}}^{(\ell)}(x_{k,t})$ are handled in a completely analogous manner. One should also remember from Section 2.2.2, that the power delay profiles $\bar{\boldsymbol{t}}_k \in \mathbb{R}^M$ have a distribution $p_{\mathsf{pdp}}$ over the users. Therefore, one needs to take the expectations over the joint distribution of $\bar{\boldsymbol{t}}_k$ and $\varepsilon_k^{\mathsf{ext},(\ell)}$ (or $\varepsilon_k^{\mathsf{app},(\ell)}$ for the case of channel estimation).

The probabilities $\mathbb{P}_{\mathsf{ext}}^{(\ell)}(x_{k,t})$ obtained through (4.86) – (4.87) are transformed via $\varphi_{\mathsf{ext}}$ to $\mathbb{Q}_{\mathsf{ext}}^{(\ell)}(\tilde{x}_{k,t})$ as discussed in Section 4.1.1. The posterior mean estimate reads

$$\langle \tilde{x}_{k,t} \rangle_{\mathsf{ext}}^{(\ell)} = \sum_{\tilde{x}_{k,t} \in \mathcal{M}} \tilde{x}_{k,t} \mathbb{Q}_{\mathsf{ext}}^{(\ell)}(\tilde{x}_{k,t}) \tag{4.88}$$

and the MSE of the extrinsic information based symbols conditioned on the feedback is denoted

$$\Omega_{\Delta x_{k,t}}^{\mathsf{ext},(\ell)} = \mathbb{E}\left\{ |x_{k,t} - \langle \tilde{x}_{k,t} \rangle_{\mathsf{ext}}^{(\ell)}|^2 \mid \langle \tilde{x}_{k,t} \rangle_{\mathsf{ext}}^{(\ell)} \right\}. \tag{4.89}$$

Note that the explicit form of (4.89) depends on the type of feedback used. Completely analogous notation is used for the feedback based on approximate APPs $\mathbb{P}_{\mathsf{app}}^{(\ell)}(x_{k,t})$.

In the following, we omit the user and time indices $k$ and $t$ when they are deemed unnecessary for the presentation.

### 4.3.2 Linear Channel Estimation with Information Feedback

**Proposition 8.** *For the LMMSE channel estimator described in Example 8, we have*

$$\tilde{\boldsymbol{\Sigma}}_{k,\mathcal{T},m}(\tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)}) = \boldsymbol{\Sigma}_{k,\mathcal{T},m}(\boldsymbol{C}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)}), \tag{4.90}$$

$$\tilde{\boldsymbol{\Sigma}}_{k,\mathcal{D}\setminus\vartheta,m}(\tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)}) = \boldsymbol{\Sigma}_{k,\mathcal{D}\setminus\vartheta,m}(\boldsymbol{C}_{\mathcal{D}\setminus\vartheta}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)}), \tag{4.91}$$

*in* (4.53) – (4.56)*. As a result,* (4.57) – (4.60) *simplifies to*

$$\tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)} = \boldsymbol{C}_{\mathcal{T}}^{(\ell)} = C_{\mathsf{tr}}^{(\ell)} \boldsymbol{I}_{\tau_{\mathsf{tr}}}, \qquad C_{\mathsf{tr}}^{(\ell)} \in \mathbb{R}, \tag{4.92}$$

$$\tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)} = \boldsymbol{C}_{\mathcal{D}\setminus\vartheta}^{(\ell)} = C_{\mathsf{d}}^{(\ell)} \boldsymbol{I}_{\tau_{\mathsf{d}}-1}, \qquad C_{\mathsf{d}}^{(\ell)} \in \mathbb{R}. \tag{4.93}$$

*The MSE of the $m$th path for the user $k$ at time index $\vartheta \in \mathcal{D}$ reads*

$$\begin{aligned}&\mathsf{mse}_{k,\vartheta,m} \\ &= \mathbb{E}_{\langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}}^{(\ell-1)}} \left\{ \bar{t}_{k,m} \left( 1 + \bar{t}_{k,m} \left[ \frac{\tau_{\mathsf{tr}}}{C_{\mathsf{tr}}^{(\ell)}} + \sum_{t \in \mathcal{D}\setminus\vartheta} \frac{|\langle \tilde{x}_{k,t} \rangle_{\mathsf{app}}^{(\ell-1)}|^2}{C_{\mathsf{d}}^{(\ell)} + \bar{t}_{k,m} \Omega_{\Delta x_k}^{\mathsf{app},(\ell-1)}} \right] \right)^{-1} \right\},\end{aligned} \tag{4.94}$$

*where the noise variances $C_{\mathsf{tr}}^{(\ell)}$ and $C_{\mathsf{d}}^{(\ell)}$ are given by the solutions to the coupled fixed point equations*

$$C_{\mathsf{tr}}^{(\ell)} = \sigma^2 + \alpha \sum_{m=1}^{M} \mathbb{E} \left\{ \bar{t}_m \left( 1 + \bar{t}_m \left[ \frac{\tau_{\mathsf{tr}}}{C_{\mathsf{tr}}^{(\ell)}} + \sum_{t \in \mathcal{D}\setminus\vartheta} \frac{|\langle \tilde{x}_t \rangle_{\mathsf{app}}^{(\ell-1)}|^2}{C_{\mathsf{d}}^{(\ell)} + \bar{t}_m \Omega_{\Delta x}^{\mathsf{app},(\ell-1)}} \right] \right)^{-1} \right\}, \tag{4.95}$$

$$\begin{aligned}C_{\mathsf{d}}^{(\ell)} = \sigma^2 + \alpha \sum_{m=1}^{M} \mathbb{E} \Bigg\{ &\frac{\bar{t}_m C_{\mathsf{d}}^{(\ell)}}{C_{\mathsf{d}}^{(\ell)} + \bar{t}_m \Omega_{\Delta x}^{\mathsf{app},(\ell-1)}} \Big[ \Omega_{\Delta x}^{\mathsf{app},(\ell-1)} \\ &+ \frac{C_{\mathsf{d}}^{(\ell)} |\langle \tilde{x}_t \rangle_{\mathsf{app}}^{(\ell-1)}|^2}{C_{\mathsf{d}}^{(\ell)} + \bar{t}_m \Omega_{\Delta x}^{\mathsf{app},(\ell-1)}} \left( 1 + \bar{t}_m \left[ \frac{\tau_{\mathsf{tr}}}{C_{\mathsf{tr}}^{(\ell)}} + \sum_{t \in \mathcal{D}\setminus\vartheta} \frac{|\langle \tilde{x}_t \rangle_{\mathsf{app}}^{(\ell-1)}|^2}{C_{\mathsf{d}}^{(\ell)} + \bar{t}_m \Omega_{\Delta x}^{\mathsf{app},(\ell-1)}} \right] \right)^{-1} \Big] \Bigg\},\end{aligned} \tag{4.96}$$

*respectively. Due to soft feedback,*

$$\Omega_{\Delta x_t}^{\mathsf{app},(\ell-1)} = 1 - |\langle \tilde{x}_t \rangle_{\mathsf{app}}^{(\ell-1)}|^2. \tag{4.97}$$

*The power delay profile $\bar{\boldsymbol{t}} = \begin{bmatrix} \bar{t}_1 & \cdots & \bar{t}_M \end{bmatrix}^{\mathsf{T}}$ has distribution $p_{\mathsf{pdp}}$, and the expectations in* (4.95) – (4.96) *should be taken with respect to the joint distribution of $\bar{\boldsymbol{t}}$, the feedback symbols $\langle \tilde{\boldsymbol{x}}_{\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}}^{(\ell-1)}$, and conditional variance* (4.97)*.*

**Proof:** See Appendix D. □

**Remark 5.** If we let $\tau_{\mathsf{d}} = 0$ or $\varepsilon_k^{\mathsf{app},(\ell-1)} = 0 \ \forall k \implies |\langle \tilde{x}_t \rangle_{\mathsf{app}}^{(\ell-1)}|^2 = 1$ and $\Omega_{\Delta x_t}^{\mathsf{app},(\ell-1)} = 0$, Proposition 8 reduces to the previous result [59, Thm. 2], as expected. Furthermore, from (4.95) – (4.96) we find that the use of soft feedback can never increase the per-path MSE of this channel estimator. ♢

**Corollary 3.** *Let us consider the special case where all the users have the same total received average power $\bar{t} > 0$. Furthermore, let the PDP have $M$ equal power multipaths so that*

$$p_{\mathsf{pdp}} = \delta_{(\bar{t}/M)\boldsymbol{e}_M}(\bar{\boldsymbol{t}}), \qquad \bar{\boldsymbol{t}} \in \mathbb{R}^M. \tag{4.98}$$

*If we define the ratios*

$$\Delta_{\mathsf{tr}} = \tau_{\mathsf{tr}}/T_{\mathsf{coh}} \quad \textit{and} \quad \Upsilon = T_{\mathsf{coh}}/M, \tag{4.99}$$

*taking the limit $\tau_{\mathsf{tr}}, \tau_{\mathsf{d}}, M \to \infty$ while keeping $\Delta_{\mathsf{tr}}, \Upsilon$ finite and fixed gives the asymptotic normalized per-path MSE*

$$\xi^{(\ell)} = \frac{\mathsf{mse}^{(\ell)}}{\bar{t}/M} = \left[1 + \Upsilon\bar{t}\left(\frac{\Delta_{\mathsf{tr}}}{C_{\mathsf{tr}}^{(\ell)}} + \frac{(1-\Delta_{\mathsf{tr}})}{C_{\mathsf{d}}^{(\ell)}}\mathbb{E}\{|\langle\tilde{x}_t\rangle_{\mathsf{app}}^{(\ell-1)}|^2\}\right)\right]^{-1}, \tag{4.100}$$

*where $t \in \mathcal{D}$ is a dummy variable. The noise variances* (4.95) *and* (4.96) *are given by the simplified fixed point equations*

$$C_{\mathsf{tr}}^{(\ell)} = \sigma^2 + \alpha\bar{t}\xi^{(\ell)}, \tag{4.101}$$

$$C_{\mathsf{d}}^{(\ell)} = \sigma^2 + \alpha\left[1 - (1-\xi^{(\ell)})\mathbb{E}\{|\langle\tilde{x}_t\rangle_{\mathsf{app}}^{(\ell-1)}|^2\}\right], \tag{4.102}$$

*respectively.*

**Proof:** The result follows from simple algebra and is therefore omitted. □

**Remark 6.** The scenario considered in Corollary 3 is highly ideal, but allows for simplified numerical evalution of the fixed point equations given in Proposition 8. One can make a physical interpretation for the case as follows:

1. The system has a very broad bandwidth and the environment rich scattering so that there are many solvable multipath components with relatively equal received powers;
2. The transmission rate is high compared to the user mobility so that one fading block contains a long sequence of transmitted symbols;
3. Very long code words are used so that they span several fading blocks.

Furthermore, if we use the notation of [6, 158] and denote the delay and Doppler spread of the channel by $T_m$ and $B_d$, respectively, $\Upsilon$ turns out to be the inverse of the channel spread factor, i.e., $\Upsilon^{-1} = T_m B_d$. Accurate channel estimation is known to be feasible when $T_m B_d \ll 1$, which in our notation translates to $\Upsilon \gg 1$. ♢

**Proposition 9.** *If $1+\alpha M < T_{\mathsf{coh}}$, the per-path MSE of the approximate ML estimator given in Example 9 converges in the large system limit to*

$$\mathsf{mse}_{k,m}^{\mathsf{ML},(\ell)} = \frac{\sigma^2}{T_{\mathsf{coh}}-1-\alpha M} + \frac{4\varepsilon_k^{\mathsf{app},(\ell-1)}\bar{t}_{k,m}(\tau_{\mathsf{d}}-1)\left[1+\varepsilon_k^{\mathsf{app},(\ell-1)}(\tau_{\mathsf{d}}-2)\right]}{(T_{\mathsf{coh}}-1)^2} + \alpha\mathbb{E}\left\{\frac{4\varepsilon_{\mathsf{app}}^{(\ell-1)}\bar{t}(\tau_{\mathsf{d}}-1)\left[T_{\mathsf{coh}}-2-\varepsilon_{\mathsf{app}}^{(\ell-1)}(\tau_{\mathsf{d}}-2)\right]}{(T_{\mathsf{coh}}-1)^2(T_{\mathsf{coh}}-1-\alpha M)}\right\}, \quad (4.103)$$

*where $\bar{t}=\sum_m \bar{t}_m$, and the expectation is with respect to the joint distribution of the PDP and the BEP of the feedback.*

**Proof:** See Appendix D. □

**Corollary 4.** *Consider again the special case of Corollary 3. Then, for the approximate ML channel estimator*

$$\begin{aligned}\xi^{\mathsf{ML},(\ell)} &= \frac{\mathsf{mse}^{\mathsf{ML},(\ell)}}{\bar{t}/M} \\ &\xrightarrow{\tau_{\mathsf{tr}},\tau_{\mathsf{d}},M\to\infty} \frac{\sigma^2}{\bar{t}(\Upsilon-\alpha)} + 4(\varepsilon_{\mathsf{app}}^{(\ell-1)})^2(1-\Delta_{\mathsf{tr}})^2 \\ &\qquad +\alpha\frac{4\varepsilon_{\mathsf{app}}^{(\ell-1)}(1-\Delta_{\mathsf{tr}})[1-\varepsilon_{\mathsf{app}}^{(\ell-1)}(1-\Delta_{\mathsf{tr}})]}{\Upsilon-\alpha}, \end{aligned} \quad (4.104)$$

*is the asymptotic normalized MSE when $\Delta_{\mathsf{tr}}=\tau_{\mathsf{tr}}/T_{\mathsf{coh}}$ and $\Upsilon=T_{\mathsf{coh}}/M$ are finite and fixed.*

**Proof:** The result follows from simple algebra and is therefore omitted. □

**Remark 7.** The results in [138] were obtained by making several approximations in the analysis. In order to compare our exact replica symmetric solution for the approximate ML channel estimator to the main result of [138, Sec. III], consider Corollary 4. Following the assumptions in [138], set $\bar{t}=1$ along with the approximations: $4\left[\varepsilon_{\mathsf{app}}^{(\ell-1)}(1-\Delta_{\mathsf{tr}})\right]^2/(\Upsilon-\alpha)\approx 0$, and $T_{\mathsf{coh}} \gg M \implies T_{\mathsf{coh}}-\alpha M \approx T_{\mathsf{coh}}$ so that $\Upsilon-\alpha \approx \Upsilon$. The first two terms of (4.104) now coincide with the variance $\Delta_a$ in [138, Sec. III], and (4.114) with (4.104) corresponds to the variance of the interference term [138, Eq. (13)][3]. ♢

[3]We remark that [138, Eq. (13)] has an error in it. There is also no separation in the bit error probabilities (in the notation of [138]) $P_e$ related to channel estimation and interference cancellation. This implies that the authors use the same type of feedback (extrinsic information or APP based) for both tasks. The type of feedback is not defined in the paper.

### 4.3.3 Iterative Data Detection and Decoding with Feedback and Mismatched CSI

Let us assume that for iteration $\ell = 1, 2, \ldots$ and any $t \in \mathcal{D}$, the non-random limiting distribution

$$F^{(\ell)}\big(P_1, \ldots, P_M, \tilde{P}_1^{(\ell)}, \ldots, \tilde{P}_M^{(\ell)}, \mathsf{mse}_1^{(\ell)}, \ldots, \mathsf{mse}_M^{(\ell)}\big) = \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \prod_{m=1}^{M} \mathbb{1}_{\mathcal{A}_{k,m}^{P}}(P_m) \mathbb{1}_{\tilde{\mathcal{A}}_{k,t,m}^{P}}(\tilde{P}_m^{(\ell)}) \mathbb{1}_{\mathcal{A}_{k,t,m}^{\mathsf{mse}}}(\mathsf{mse}_m^{(\ell)}), \quad (4.105)$$

exists almost surely, and is independent of $t$. The auxiliary sets for the indicator functions are defined as

$$\mathcal{A}_{k,m}^{P} = \{P_m \geq 0 \mid |h_{k,m}|^2 \leq P_m\}, \quad (4.106)$$

for the received powers of the true channel, and

$$\tilde{\mathcal{A}}_{k,t,m}^{P} = \{\tilde{P}_m^{(\ell)} \geq 0 \mid |\langle \tilde{h}_{k,t,m} \rangle_{(\ell)}|^2 \leq \tilde{P}_m^{(\ell)}\}, \quad (4.107)$$

$$\mathcal{A}_{k,t,m}^{\mathsf{mse}} = \{\mathsf{mse}_m^{(\ell)} \geq 0 \mid \mathsf{mse}_{k,t,m}^{(\ell)} \leq \mathsf{mse}_m^{(\ell)}\}, \quad (4.108)$$

for the received powers and MSEs of the channel estimates, respectively. With some abuse of notation, we refer to (4.105) also when the MSEs are given by the approximate ML channel estimator, denoted by $\{\mathsf{mse}_m^{\mathsf{ML},(\ell)}\}_{m=1}^{M}$.

**Proposition 10.** *Consider the SUMF-based iterative MUDD in Example 12. For hard or soft feedback and any channel estimator, the noise variance in* (4.70) *is given by*

$$D^{(\ell)} = \sigma^2 + \alpha \sum_{m=1}^{M} \mathbb{E}\{|h_m x - \langle \tilde{h}_m \rangle_{(\ell)} \langle \tilde{x} \rangle_{\mathsf{ext}}^{(\ell)}|^2\}. \quad (4.109)$$

*The expectation should be taken with respect to the limiting empirical distribution of the true and estimated channel and data symbols, calculated over the user population whose power delay profile is drawn according to* $p_{\mathsf{pdp}}$.

**Proof:** See Appendix D. □

**Proposition 11.** *Let the channel estimation be performed by the LMMSE estimator of Example 8, or the approximate ML estimator of Example 9. The noise variance for the SUMF with the LMMSE based channel estimator reads*

$$D_{\mathsf{sumf}}^{(\ell)} = \sigma^2 + \alpha \sum_{m=1}^{M} \mathbb{E}\left\{\mathsf{mse}_m^{(\ell)} + \Omega_{\Delta x}^{(\ell-1)} \tilde{P}_m^{(\ell)}\right\}, \quad (4.110)$$

*where the expectations are with respect to* (4.105) *and the error variance of the extrinsic information based feedback* $\Omega_{\Delta x}^{(\ell-1)}$*. When the CSI is provided by the approximate ML channel estimator, the noise variance for the SUMF utilizing hard and soft PIC is given by*

$$D_{\mathsf{sumf-hard}}^{\mathsf{ML},(\ell)} = \sigma^2 + \alpha \sum_{m=1}^{M} \mathbb{E}\left\{ \mathsf{mse}_m^{\mathsf{ML},(\ell)} + \Omega_{\Delta x}^{(\ell-1)} P_m \left( 1 - 2\varepsilon_{\mathsf{app}}^{(\ell-1)} \frac{\tau_{\mathsf{d}} - 1}{T_{\mathsf{coh}} - 1} \right) \right\}, \tag{4.111}$$

$$D_{\mathsf{sumf-soft}}^{\mathsf{ML},(\ell)} = \sigma^2 + \alpha \sum_{m=1}^{M} \mathbb{E}\left\{ \mathsf{mse}_m^{\mathsf{ML},(\ell)} + \Omega_{\Delta x}^{(\ell-1)} \big( P_m - \mathsf{mse}_m^{\mathsf{ML},(\ell)} \big) \right\}, \tag{4.112}$$

*respectively. The instantaneous SINR for the SUMF with LMMSE based channel estimators reads*

$$\mathsf{sinr}_{k,t}^{(\ell)} = \left( \sum_{m=1}^{M} |\langle \tilde{h}_{k,t,m} \rangle_{(\ell)}|^2 \right)^2 \Bigg/ \left( \sum_{m=1}^{M} |\langle \tilde{h}_{k,t,m} \rangle_{(\ell)}|^2 \big( D_{\mathsf{sumf}}^{(\ell)} + \mathsf{mse}_{k,t,m}^{(\ell)} \big) \right). \tag{4.113}$$

**Proof:** Equations (4.110) and (4.113) follow from the fact that for the LMMSE CE of Example 8, the channel estimate and the error are uncorrelated. For the case with approximate ML channel estimator, using Lemma 1 in Appendix D gives (4.111).□

**Remark 8.** It is easy to verify that if we let $\langle \tilde{x} \rangle_{\mathsf{ext}}^{(\ell)} = 0 \implies \Omega_{\Delta x}^{(\ell-1)} = 1$, Proposition 11 gives [59, Proposition 2], and setting $\langle \tilde{h}_m \rangle_{(\ell)} = h_m \implies \mathsf{mse}_m^{(\ell)} = 0$, reduces it to [125, Proposition 2]. ♢

The case of ML channel estimator is in general slightly cumbersome to deal with numerically. We therefore consider again the special case of Corollary 3, that gives the next simplified result.

**Corollary 5.** *Let the PDP be drawn according to* (4.98)*. Then,*

$$D_{\mathsf{sumf-soft}}^{\mathsf{ML},(\ell)} = \sigma^2 + \alpha \bar{t} \left[ \xi^{\mathsf{ML},(\ell)} + \big( 1 - \mathbb{E}\{ |\langle \tilde{x} \rangle_{\mathsf{ext}}^{(\ell)}|^2 \} \big) \big( 1 - \xi^{\mathsf{ML},(\ell)} \big) \right], \tag{4.114}$$

$$D_{\mathsf{sumf-hard}}^{\mathsf{ML},(\ell)} = \sigma^2 + \alpha \bar{t} \left( \xi^{\mathsf{ML},(\ell)} + 4\varepsilon_{\mathsf{ext}}^{(\ell-1)} \big[ 1 - 2\varepsilon_{\mathsf{app}}^{(\ell-1)} (1 - \Delta_{\mathsf{tr}}) \big] \right), \tag{4.115}$$

$$\mathsf{sinr}^{\mathsf{ML},(\ell)} = \frac{\bar{t} \big[ 1 - 2\varepsilon_{\mathsf{app}}^{(\ell-1)} (1 - \Delta_{\mathsf{tr}}) \big]^2}{D_{\mathsf{sumf}}^{\mathsf{ML},(\ell)} \big[ 1 + \xi^{\mathsf{ML},(\ell)} - 4\varepsilon_{\mathsf{app}}^{(\ell-1)} (1 - \Delta_{\mathsf{tr}}) \big]}, \tag{4.116}$$

*in the limit* $\tau_{\mathsf{tr}}, \tau_{\mathsf{d}}, M \to \infty$ *when* $\Delta_{\mathsf{tr}} = \tau_{\mathsf{tr}}/T_{\mathsf{coh}}$ *and* $\Upsilon = T_{\mathsf{coh}}/M$ *are finite and fixed. The normalized per-path MSE* $\xi^{\mathsf{ML},(\ell)}$ *is given by* (4.104) *and the noise variance* $D^{\mathsf{ML},(\ell)}_{\mathsf{sumf}}$ *in* (4.116) *is either* $D^{\mathsf{ML},(\ell)}_{\mathsf{sumf-soft}}$ *or* $D^{\mathsf{ML},(\ell)}_{\mathsf{sumf-hard}}$*, depending on the type of feedback used for interference cancellation.*

**Remark 9.** If we make the approximation $1 - \Delta_{\mathsf{tr}} \approx 1$ in Corollary 5, and $\xi^{\mathsf{ML},(\ell)}$ is modified as discussed in Remark 7, (4.115) – (4.116) reduce to the main result of [138, Sec. IV]. ♢

**Proposition 12.** *Let the multiuser decoding be performed by the non-linear MAP-MUDD of Example 10, or the LMMSE-PIC MUDD described in Example 11. Assume the channel estimation is performed by the LMMSE estimator of Example 8. The instantaneous post-detection SINR at* $\ell$*th iteration for the user* $k$ *and time index* $t \in \mathcal{D}$ *is given by*

$$\mathsf{sinr}^{(\ell)}_{k,t} = \sum_{m=1}^{M} \frac{|\langle \tilde{h}_{k,t,m} \rangle_{(\ell)}|^2}{D^{(\ell)} + \mathsf{mse}^{(\ell)}_{k,t,m}}. \tag{4.117}$$

*The noise variance* $D^{(\ell)}$ *is given for the MAP-MUDD or the LMMSE-PIC MUDD by the solution to the fixed point equation*

$$D^{(\ell)} = \sigma^2 + \alpha \sum_{m=1}^{M} \mathbb{E}\left\{ \left( \frac{D^{(\ell)}}{D^{(\ell)} + \mathsf{mse}^{(\ell)}_m} \right)^2 \times \left[ \mathsf{mse}^{(\ell)}_m \left( \frac{D^{(\ell)} + \mathsf{mse}^{(\ell)}_m}{D^{(\ell)}} \right) + \tilde{P}^{(\ell)}_m \, V\Big( D^{(\ell)}, \{\tilde{P}^{(\ell)}_m\}_{m=1}^M, \{\mathsf{mse}^{(\ell)}_m\}_{m=1}^M \Big) \right] \right\}, \tag{4.118}$$

*where for the MAP-MUDD*

$$V\Big( D^{(\ell)}, \{\tilde{P}^{(\ell)}_m\}_{m=1}^M, \{\mathsf{mse}^{(\ell)}_m\}_{m=1}^M \Big) = 1 - \mathbb{E}\Bigg\{ \sum_{a_1 \in \{\pm 1\}} \frac{1 + a_1 \langle \tilde{a}_1 \rangle^{(\ell-1)}_{\mathsf{ext}}}{2} \times \int \tanh\left( \sum_{m=1}^{M} \frac{\tilde{P}^{(\ell)}_m}{D^{(\ell)} + \mathsf{mse}^{(\ell)}_m} + \nu \sqrt{\sum_{m=1}^{M} \frac{\tilde{P}^{(\ell)}_m}{D^{(\ell)} + \mathsf{mse}^{(\ell)}_m}} + a_1 \frac{\lambda^{(\ell-1)}_{a_1}}{2} \right) \mathrm{D}\nu \Bigg| \{\tilde{P}^{(\ell)}_m\}_{m=1}^M, \{\mathsf{mse}^{(\ell)}_m\}_{m=1}^M \Bigg\}, \tag{4.119}$$

*and for the LMMSE-PIC MUDD*

$$V\Big(D^{(\ell)}, \{\tilde{P}_m^{(\ell)}\}_{m=1}^M, \{\mathsf{mse}_m^{(\ell)}\}_{m=1}^M\Big) = \mathbb{E}\left\{\Omega_{\Delta x}^{(\ell-1)}\left(1+\Omega_{\Delta x}^{(\ell-1)}\sum_{m=1}^{M}\frac{\tilde{P}_m^{(\ell)}}{D^{(\ell)}+\mathsf{mse}_m^{(\ell)}}\right)^{-1}\;\middle|\;\{\tilde{P}_m^{(\ell)}\}_{m=1}^M, \{\mathsf{mse}_m^{(\ell)}\}_{m=1}^M\right\}. \tag{4.120}$$

*In the following the noise variances for the MAP and LMMSE MUDDs that are solutions to* (4.118) – (4.120) *are denoted by* $D_{\mathsf{map}}^{(\ell)}$ *and* $D_{\mathsf{lmmse}}^{(\ell)}$*, respectively.*

**Proof:** See Appendix E. □

**Remark 10.** Let $\mathsf{mse}_m = 0 \;\forall m$ in (4.118) – (4.120). We immediately retrieve the results in [85, 125, 126]. On the other hand, if we set $\lambda_{a_1}^{(\ell-1)} = \lambda_{a_2}^{(\ell-1)} = 0$ and consider the distribution (4.98) for the user PDPs, we get [59, Eq (12)] after some algebra, as expected. ♢

Interestingly, there is a common part in (4.118) for both data estimators that does not depend on the extrisic information based feedback at all. Note that these terms vanish if and only if $\mathsf{mse} \to 0$. Furthermore, there is a connection with the estimator specific terms $V(\cdots)$ to the related terms in the case of perfect CSI.

**Remark 11.** Consider the terms (4.119) and (4.120), specific to the MAP-MUDD and the LMMSE-PIC MUDD, respectively. One can verify that for fixed $D^{(\ell)}$ and a single path $M = 1$,

$$\mathbb{E}\Big\{V\Big(D^{(\ell)}, \{\tilde{P}_m^{(\ell)}\}_{m=1}^M, \{\mathsf{mse}_m^{(\ell)}\}_{m=1}^M\Big)\Big\} = \mathbb{E}\left\{\left|x - \langle\tilde{x}\rangle_{(\ell)}\right|^2\right\}, \tag{4.121}$$

where $\langle\tilde{x}\rangle_{(\ell)}$ is the estimate of the desired user's data symbols and given for the MAP and LMMSE MUDDs by (4.74) and (4.81), respectively. Furthermore, these MSEs of the data symbols are equal to the corresponding terms for the case of perfect CSI [85, 125, 126], with the noise variance increased by the MSE of the channel estimates and the channel power reduced accordingly. ♢

**Corollary 6.** *Let us assume the same conditions for the channel parameters as in Corollary 5 and define*

$$\mathsf{sinr}^{(\ell)}\big(D^{(\ell)}\big) = \frac{\bar{t}(1-\xi^{(\ell)})}{D^{(\ell)}}. \tag{4.122}$$

*Then, in the limit* $\tau_{\mathsf{tr}}, \tau_{\mathsf{d}}, M \to \infty$ *with* $\Delta_{\mathsf{tr}} = \tau_{\mathsf{tr}}/T_{\mathsf{coh}}$ *and* $\Upsilon = T_{\mathsf{coh}}/M$ *finite and fixed,*

$$D^{(\ell)}_{\mathsf{map}} = \sigma^2 + \alpha \bar{t}\, \mathbb{E}\Bigg\{ 1 - (1-\xi^{(\ell)}) \sum_{a_1 \in \{\pm 1\}} \frac{1 + a_1 \langle \tilde{a}_1 \rangle^{(\ell-1)}_{\mathsf{ext}}}{2} \times \int \tanh\left( \mathsf{sinr}^{(\ell)}(D^{(\ell)}_{\mathsf{map}}) + \nu \sqrt{\mathsf{sinr}^{(\ell)}(D^{(\ell)}_{\mathsf{map}})} + a_1 \frac{\lambda^{(\ell-1)}_{a_1}}{2} \right) \mathrm{D}\nu \Bigg\}, \tag{4.123}$$

*for the MAP-MUDD and*

$$D^{(\ell)}_{\mathsf{lmmse}} = \sigma^2 + \alpha \bar{t}\, \mathbb{E}\left\{ \xi^{(\ell)} + \frac{\Omega^{(\ell-1)}_{\Delta x}(1-\xi^{(\ell)})}{1 + \Omega^{(\ell-1)}_{\Delta x} \mathsf{sinr}^{(\ell)}(D^{(\ell)}_{\mathsf{lmmse}})} \right\}, \tag{4.124}$$

*for the LMMSE-PIC MUDD.*

**Proof:** The result follows from simple algebra and is therefore omitted. □

Note that when the solution to the fixed point equation (4.123) or (4.124) is obtained, the post-detection SINRs for the MAP and LMMSE-PIC MUDDs are given by (4.122) with $\mathsf{sinr}^{(\ell)}(D^{(\ell)}_{\mathsf{lmmse}})$ and $\mathsf{sinr}^{(\ell)}(D^{(\ell)}_{\mathsf{map}})$, respectively.

### 4.3.4 Multiuser Efficiency and Related Performance Measures

Consider the case of perfect CSI and let

$$\mathsf{snr}_k[c] = \frac{\|\boldsymbol{h}_k[c]\|^2}{\sigma^2}, \tag{4.125}$$

be the instantaneous received SNR of the $k$th user during $c$th fading block in (2.10). Furthermore, let $\mathsf{sinr}^{(\ell)}_k[c]$ be the corresponding SINR of the same user at the output of the MUDD during iteration $\ell = 1, 2, \ldots$, and given in Section 4.3. If we define

$$\eta^{(\ell)} = \frac{\sigma^2}{D^{(\ell)}}, \qquad 0 \leq \eta^{(\ell)} \leq 1, \tag{4.126}$$

where $D^{(\ell)}$ is the noise variance of the single-user system given in Sections 4.2.2 and 4.2.3, the output SINR of the iterative MUDD reads

$$\mathsf{sinr}^{(\ell)}_k[c] = \eta^{(\ell)}\, \mathsf{snr}_k[c]. \tag{4.127}$$

Note that for ergodic Rayleigh fading channel and BICM this is consistent with the definition of the asymptotic multiuser efficiency $\eta_{\sigma^2\to 0}$ in (1.1) if we take the high SNR limit $\sigma^2 \to 0$ and let the functions $\varepsilon_{\mathsf{su}}(\overline{\mathsf{snr}})$ and $\varepsilon_{\mathsf{mu}}(\overline{\mathsf{snr}})$ be coded BERs. Following the notation of [125, 126], we let

$$\Psi : [0,1] \to [0,1] : \eta^{(\ell-1)} \mapsto \eta^{(\ell)}, \tag{4.128}$$

be the mapping function that describes the DE for a specific iterative MUDD.

One might be interested in tracking (4.128) for the DE-GA when channel estimation is added to the system. In this case, however, some caveats in how to define the corresponding mapping function exists. Below we follow one approach that considers the loss in effective SNR, arising from both the MAI and the imperfect CSI.

Let us consider the simplest case of equal power users with uniform power delay profiles and LMMSE channel estimation. We omit the user and block indices and define a new parameter related to (4.126) – (4.127) as

$$\eta_{\mathsf{ce}}^{(\ell)} = \frac{\mathsf{sinr}^{(\ell)}}{\mathsf{snr}} = \frac{\sigma^2}{\frac{(\bar{t}/M)(D^{(\ell)}+\mathsf{mse}^{(\ell)})}{\bar{t}/M-\mathsf{mse}^{(\ell)}}} \tag{4.129}$$

$$\xrightarrow{\text{Corollary 3}} \frac{\sigma^2}{D^{(\ell)}}(1-\xi^{(\ell)}), \qquad 0 \le \eta_{\mathsf{ce}}^{(\ell)} \le 1, \tag{4.130}$$

where (4.130) corresponds to the simplified case of large number of multipaths, considered in Corollary 3. Naturally $\eta_{\mathsf{ce}}^{(\ell)} \to \eta^{(\ell)}$ when $\mathsf{mse}^{(\ell)} \to 0$ or $\xi^{(\ell)} \to 0$. One should note, however, that by using this definition:

- We are comparing a multiuser system with channel mismatch to a single-user system having perfect CSI;
- In addition to the choice of iterative MUDD, $\eta_{\mathsf{ce}}^{(\ell)}$ depends on the choice of channel estimator and the system parameters related to it (number of training symbols $\tau_{\mathsf{tr}}$ per block, coherence time $T_{\mathsf{coh}}$, number of multipaths $M$) via $\mathsf{mse}^{(\ell)}$ or $\xi^{(\ell)}$ as well;
- Even with error free feedback $\varepsilon_{\mathsf{ext}}^{(\ell)}, \varepsilon_{(\ell)}^{\mathsf{app}} \to 0$, for all finite coherence times $T_{\mathsf{coh}}$ or ratios $\Upsilon$, we have $\eta_{\mathsf{ce}}^{(\ell)} < 1$ since $\mathsf{mse}^{(\ell)} > 0$ and $\xi^{(\ell)} > 0$, respectively.

Therefore, instead of describing just the MAI suppression capacity of the iterative MUDD like (4.126) does, $\eta_{\mathsf{ce}}^{(\ell)}$ provides information about the efficiency of the entire iterative channel estimation and MUDD scheme. For the following, we let

$$\Psi_{\mathsf{ce}} : [0,1] \to [0,1] : \eta_{\mathsf{ce}}^{(\ell-1)} \mapsto \eta_{\mathsf{ce}}^{(\ell)}, \tag{4.131}$$

be a mapping function related to (4.128), that describes the evolution of (4.129) – (4.130) with Gaussian approximation.

Finally, let us consider the situation when the feedback symbols tend to correct decisions, that is,

$$D_{\mathsf{lmmse}}^{(\ell)} \xrightarrow{\Omega_{\Delta x}^{(\ell-1)}\to 0} D_{\mathsf{map}}^{(\ell)} = \sigma^2 + \alpha \sum_{m=1}^{M} \mathbb{E}\left\{ \frac{D_{\mathsf{map}}^{(\ell)} \mathsf{mse}_m^{(\ell)}}{D_{\mathsf{map}}^{(\ell)} + \mathsf{mse}_m^{(\ell)}} \right\}, \tag{4.132}$$

$$D_{\mathsf{sumf}}^{(\ell)} \xrightarrow{\Omega_{\Delta x}^{(\ell-1)}\to 0} \sigma^2 + \alpha \sum_{m=1}^{M} \mathbb{E}\left\{ \mathsf{mse}_m^{(\ell)} \right\}. \tag{4.133}$$

Plugging the solution of (4.132) or (4.133), depending on the iterative MUDD, to (4.129) gives the upper bound for maximum achievable $\eta_{\mathsf{ce}}^{(\ell)}$ for a given MSE of channel estimates. For iterative channel estimator, on the other hand, $\mathsf{mse}_m^{(\ell)}$ can be lower bounded by considering the corresponding non-iterative channel estimator with $T_{\mathsf{coh}} - 1$ known training symbols.

Note that in contrast to the case of perfect CSI, where

$$D_{\mathsf{sumf}}^{(\ell)} = D_{\mathsf{lmmse}}^{(\ell)} = D_{\mathsf{map}}^{(\ell)} \xrightarrow{\Omega_{\Delta x}^{(\ell-1)}\to 0} \sigma^2, \tag{4.134}$$

with CSI mismatch the performance of the SUMF and the LMMSE-PIC / MAP MUDDs can be different in this this limit. For the special case $\bar{t}_{k,m} = \bar{t}/M$,

$$D_{\mathsf{map}} = \frac{1}{2}\left( \sigma^2 - (1-\alpha M)\mathsf{mse}^{(\ell)} + \sqrt{4\mathsf{mse}^{(\ell)}\sigma^2 + \left[\sigma^2 - (1-\alpha M)\mathsf{mse}^{(\ell)}\right]^2} \right). \tag{4.135}$$

The maximum difference in the average post-detection SINR between the LMMSE-PIC / MAP-MUDD and SUMF occurs for this scenario at load

$$\alpha = \frac{\mathsf{mse} + \sigma^2}{M\mathsf{mse}}. \tag{4.136}$$

From (4.133), (4.135) and (4.136) we get

$$\left.\frac{\mathsf{sinr}_{\mathsf{map}}}{\mathsf{sinr}_{\mathsf{sumf}}}\right|_{\alpha=\frac{\mathsf{mse}+\sigma^2}{M\mathsf{mse}}} \xrightarrow{\Omega_{\Delta x}^{(\ell-1)}\to 0} 2\frac{\mathsf{mse} + \sigma^2}{\mathsf{mse} + \sigma^2 + \sqrt{\sigma^2(\mathsf{mse} + \sigma^2)}} \leq 2 \quad (\approx 3 \text{ dB}), \tag{4.137}$$

where the MSE of the channel estimates is assumed to be non-zero. Note that $M \to \infty \implies \mathsf{mse} \to 0$ and, thus, for large numbers of multipaths the maximum loss for SUMF approaces zero (in dBs). Thus, for wideband channels there are, in general,

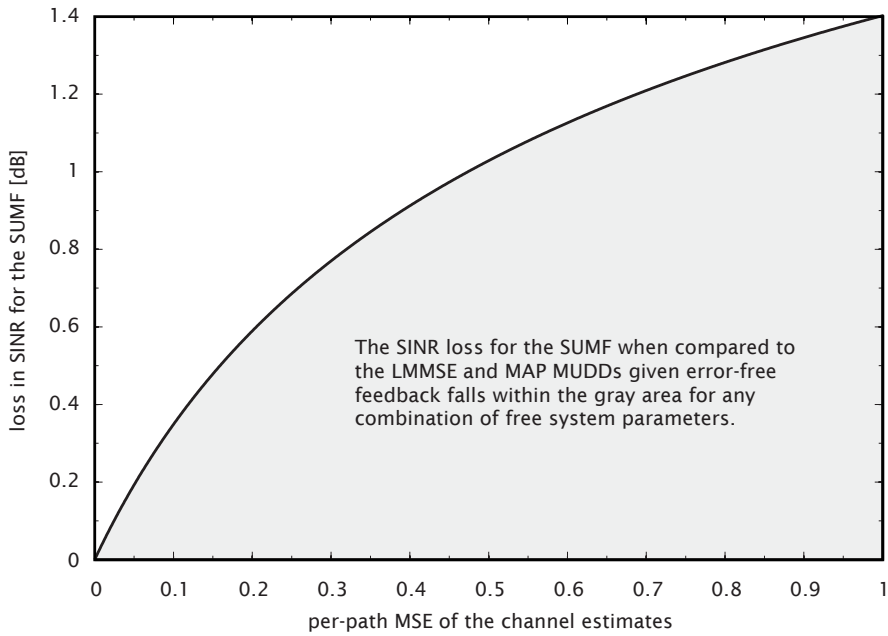



**Figure 4.2.** Loss in output SINR for the iterative SUMF when compared to LMMSE / MAP MUDD given error-free feedback. Equal power paths, $\overline{t} = 1$ and $\overline{\mathsf{snr}} = 10$ dB.

little differences between the performances of the different iterative MUDDs, if they converge to their maximum values of $\eta_{\mathsf{ce}}$. It is important to remember, however, that their convergence properties may differ strongly.

Finally, Figure 4.2 depicts the asymptotic loss for the SUMF in the case of a finite number of equal power paths with $\overline{t} = 1$, $\overline{\mathsf{snr}} = 10$ dB and genie-aided feedback. The black line gives an upper bound for the SINR loss in dBs as obtained in (4.137). Any combination of values for the user load, the number of multipaths and the MSE of the channel estimates corresponds to a point within the gray area. One can thus infer from the above figure that even for small number of multipaths, the performance loss is much smaller than the upper bound of 3 dBs for typical system parameters. Therefore, even for small numbers of solvable multipaths we get the same conclusion as in the wideband limit that the performance of the iterative MUDDs can be expected to be roughly the same given they converge to their maximum values of $\eta_{\mathsf{ce}}$.

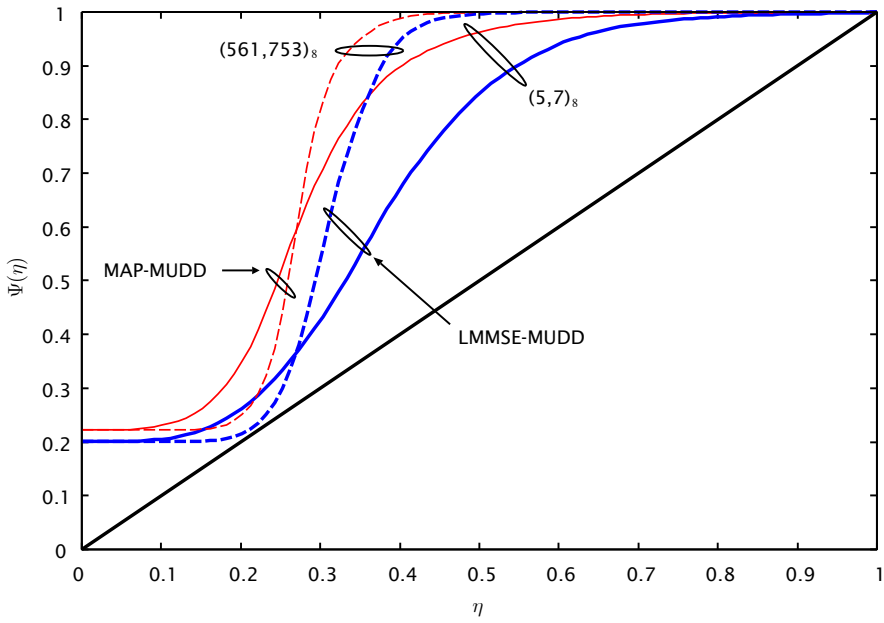



**Figure 4.3.** DE-GA with equal power users and AWGN channel. User load $\alpha = 1.8$, and average SNR of $\overline{\mathsf{snr}} = 6$ dB.

## 4.4 Numerical Examples and Discussion

In this section we present a set of numerical examples derived from the analytical results obtained in Sections 4.2 and 4.3. We again remind the reader that the numerical examples given here are based on the asymptotic large system analysis and for finite systems are approximations.

For all considered cases the binary ECC for the BICM is a half-rate convolutional code. Two maximum free distance codes defined by the polynomials in octal notation $(5, 7)_8$ and $(561, 753)_8$, with respective constraint lengths of three and nine, are used [198]. The codes were selected to represent two extremes — the first one is a very simple "textbook code" whereas the latter is a much stronger code adopted in the current state-of-the-art cellular UMTS network. Modulation mapping is Gray encoded QPSK and, thus, the BICM has code rate $R = 1$ and the average SNR per information bit is $\overline{\mathsf{snr}}_k = \bar{t}_k/\sigma^2$.

We start the numerical examples by considering the density evolution of two iterative MUDDs under the assumption of perfect CSI at the receiver. Figure 4.3 depicts $\Psi$ given in (4.128) for the DE-GA of the MAP-MUDD (see Section 4.1.4) and the LMMSE-PIC MUDD (see Example 11). Equal power users and AWGN

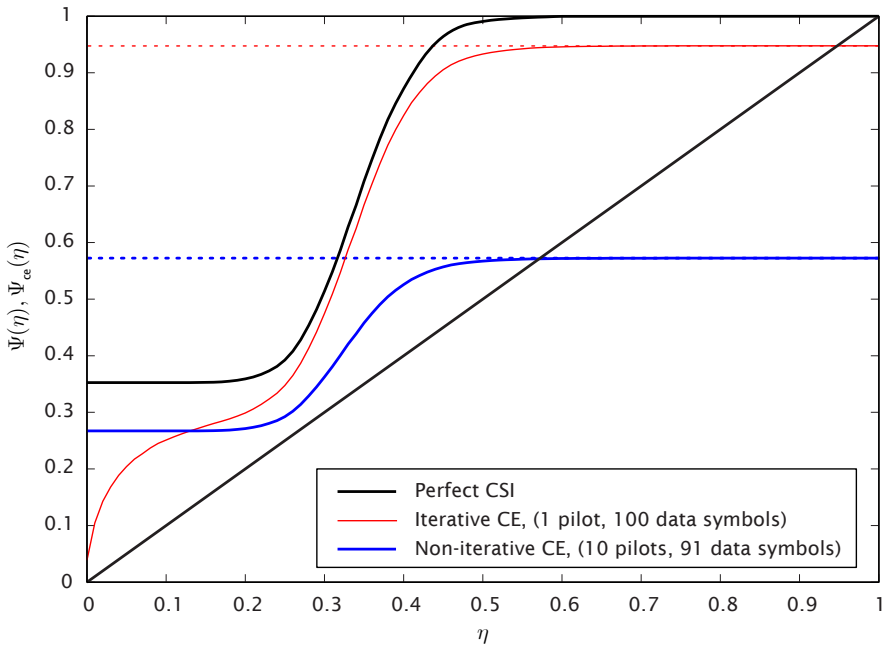



**Figure 4.4.** Mapping function $\Psi$ for the case of perfect CSI and $\Psi_{\mathsf{ce}}$ when channel estimation is employed. Three equal power paths, channel coherence time of $T_{\mathsf{coh}} = 101$ symbols, $\overline{\mathsf{snr}} = 6$ dB, user load $\alpha = 1.2$, LMMSE-PIC MUDD and $(561, 753)_8$ convolutional code for all cases. Channel estimation by non-iterative or iterative LMMSE estimator. The dotted lines show the upper bounds obtained by using (4.135).

channel was assumed. Note that the curve for the LMMSE-PIC MUDD and $(5, 7)_8$ code can be found also in [125, Fig. 4]. As expected, the MAP-MUDD obtains higher post-detection SINR for the same level of feedback reliability due to its more efficient MAI suppression. This allows the system to be more heavily loaded while still guaranteeing a single-user performance. Another observation to be made is that while the combination of LMMSE-PIC MUDD and $(561, 753)_8$ code is close to its maximum load at $\alpha = 1.8$, the shorter memory $(5, 7)_8$ code converges for much higher loads (see the curves for higher loads in [125, Fig. 4]).

In Figure 4.4, we have plotted $\Psi_{\mathsf{ce}}$ given in (4.131) for the case of LMMSE based channel estimation and LMMSE-PIC MUDD. For comparison, the upper bounds discussed in Section 4.3.4 (dashed lines), and the corresponding curve $\Psi$ for the case of perfect CSI are also included. Ergodic Rayleigh fading channel with three equal power multipaths is assumed. Only the $(561, 753)_8$ code is considered

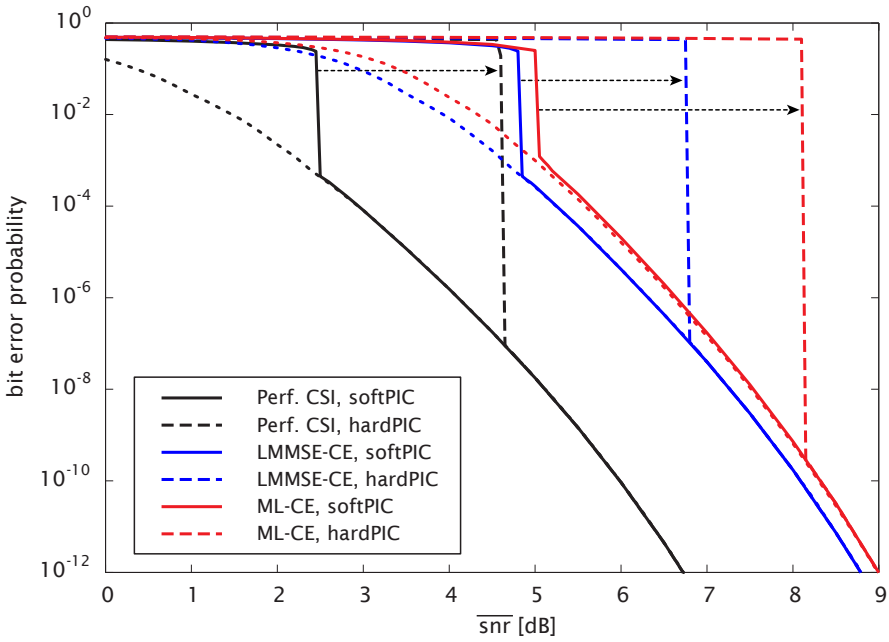



**Figure 4.5.** Bit error probability vs. SNR of the SUMF based MUDs with soft and hard feedback and LMMSE or ML channel estimation. Training overhead of 10 %, user load $\alpha = 0.7$ and inverse of channel spread factor $\Upsilon = 20$. Rate-$1/2$ convolutional code $(561, 753)_8$ and Gray encoded QPSK. Dotted lines show the minimum BEP bounds obtained using (4.135).

and all users are restricted to have the same average received power. It is clear that for both cases with channel estimation, the iterative MUDD converges to its maximum value of $\eta_{\mathsf{ce}}$. Therefore, in this case where the user load is $\alpha = 1.2$, the limiting factor in the performance is not the MAI but the imperfect CSI, and no better performance can be obtained by using MAP-MUDD instead of the LMMSE-PIC MUDD (cf. Section 4.3.4). We also remark that the upper bound for the case of iterative channel estimation was obtained by using (4.135), where the MSE of the channel estimates was lower bounded by assuming a non-iterative LMMSE channel estimator using $\tau_{\mathsf{tr}} = T_{\mathsf{coh}} - 1$ known training symbols. Therefore, the combination of iterative LMMSE-CE and LMMSE-PIC MUDD in fact achieves the optimum performance for the given channel conditions. It is quite remarkable that this is can be accomplished by using only one pilot symbol. With non-iterative LMMSE channel estimation and ten training symbols, on the other hand, severe loss in output SINR is observed. The interesting shape of the curve $\Psi_{\mathsf{ce}}$ for the case of iterative

channel estimation is due to the use of approximate APPs in the feedback to the CE. The intuitive explanation goes as follows. For low input SNR, the channel decoder output is dominated by the symbol-by-symbol a posteriori probabilities obtained by using the channel information rather than the extrinsic information that arises from the code constraints. Therefore, for small output SINR after MUD front-end, we can have a situation where $\varepsilon^{\mathsf{ext}} = 0.5$ while $\varepsilon^{\mathsf{app}} < 0.5$. For the extrinsic information based feedback symbols then $|\langle \tilde{x} \rangle_{\mathsf{ext}}|^2 = 0$ and no MAI suppression is attainable. This corresponds to the flat region at the beginning of the DE curves in Figure 4.3. At the same time, however, the feedback to the channel estimator can be reliable enough to lower the MSE of the channel estimates since $|\langle \tilde{x} \rangle_{\mathsf{app}}|^2 > 0$. This in turn affects (4.118) and (4.129), so that $\Psi_{\mathsf{ce}}(\eta_{\mathsf{ce}}) > \eta_{\mathsf{ce}}$.

Let us next look at the performance of the multiuser DS-CDMA system when the receiver has converged to its maximum value of $\eta_{\mathsf{ce}}$. Figure 4.5 depicts the BEP vs. SNR for the SUMF with linear channel estimators under the simplifying assumptions of Corollary 3. To guarantee convergence for all considered cases within the given SNR range, the user load was set to $\alpha = 0.7$ and pilot overhead of $10$ % was used. Inverse channel spread factor $\Upsilon = T_{\mathsf{coh}}/M = 20$ was assumed. From the previous density evolution analysis it is clear that a notable performance loss should occur with linear channel estimation when compared to the case of perfect CSI. The asymptotic performance of the MUDs was again obtained by using the techniques discussed in Section 4.3.4 and plotted with dotted lines. First observation from the figure is that for both the soft and the hard PIC there is a phase transition in BEP from one half to the minimum attainable, for the given receiver and system parameters. Furthermore, there is very little difference between the non-iterative ML channel estimator (ML-CE) given in Example 9 and the non-iterative LMMSE channel estimator (LMMSE-CE) in the latter region. There is, however, a significant difference in the threshold SNR when the phase transition occurs for the soft and hard feedback, as illustrated by the three arrows in the figure. For the case of perfect CSI this is well known and with the channel estimation the effect is roughly the same (LMMSE-CE) or worse (ML-CE). At high SNR, the loss caused by imperfect CSI is roughly 2 dBs, but the difference in the convergence threshold is up to $5.5$ dBs. Due to the poor performance of the hard feedback based PIC, we drop it from further discussion in this section and concentrate on presenting results for the iterative MUDDs that use soft interference cancellation.

As a final comparison between the two feedback strategies, we shall look at the iterative approximate ML and LMMSE channel estimators described in Examples 8 and 9. Figure 4.6 depicts the normalized MSE for the approximate ML-CE

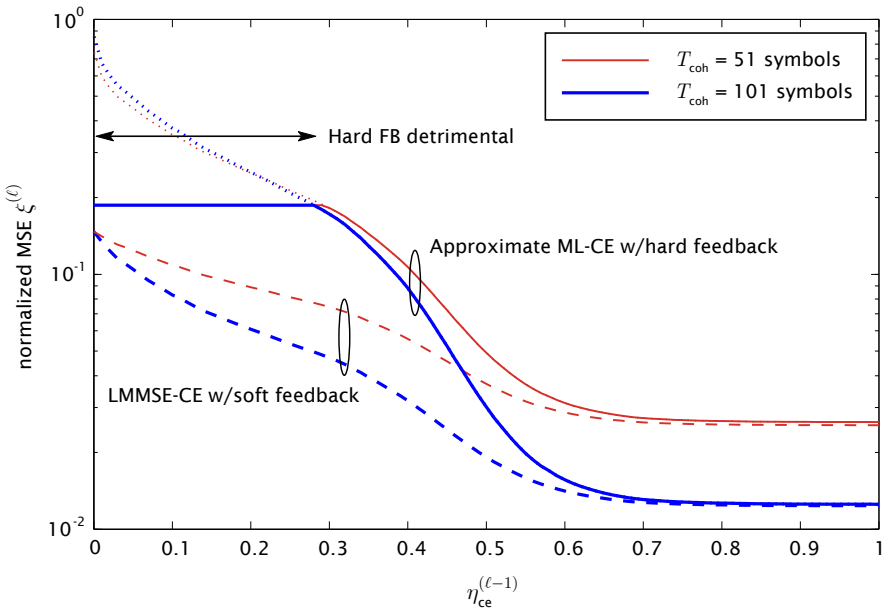



**Figure 4.6.** Normalized MSE $\xi^{(\ell)} = \mathsf{mse}^{(\ell)}/(\bar{t}/M)$ vs. $\eta_{\mathsf{ce}}^{(\ell-1)}$ for the linear channel estimators with soft and hard feedback given in Examples 8 and 9, respectively. Three equal power paths, $\tau_{\mathsf{tr}} = 10$ pilot symbols, user load $\alpha = 1.2$ and $\overline{\mathsf{snr}} = 4$ dB.

and the LMMSE-CE. The solid lines represent the normalized MSE of an iterative (approximate) ML-CE that uses feedback only when it helps to improve the performance of the pilots-only case. The flat part of the solid curves roughly at $\eta_{\mathsf{ce}} \in [0, 0.3]$ corresponds to the MSE obtained by the estimator when using only known pilot symbols. The dotted lines represent the MSE of the channel estimates when the feedback is enabled. Thus, in this region the MSE is higher with feedback than without and performing iterative channel estimation is detrimental. When the reliability of the feedback improves, the MSE can be lowered by using the feedback symbols as additional pilots. The dashed lines give the normalized MSE of the LMMSE-CE with soft feedback. As noted in Remark 5, the feedback is never harmful for this channel estimator, so there is no need to check whether to use it or not. The MSE is lowered for all values of $\eta_{\mathsf{ce}}$. As expected, the LMMSE-CE provides lower MSEs than the approximate ML-CE, although when the feedback symbols get realiable enough, the performance of these channel estimators is virtually the same. For the rest of the section, we shall concentrate on the LMMSE-based channel estimator only.

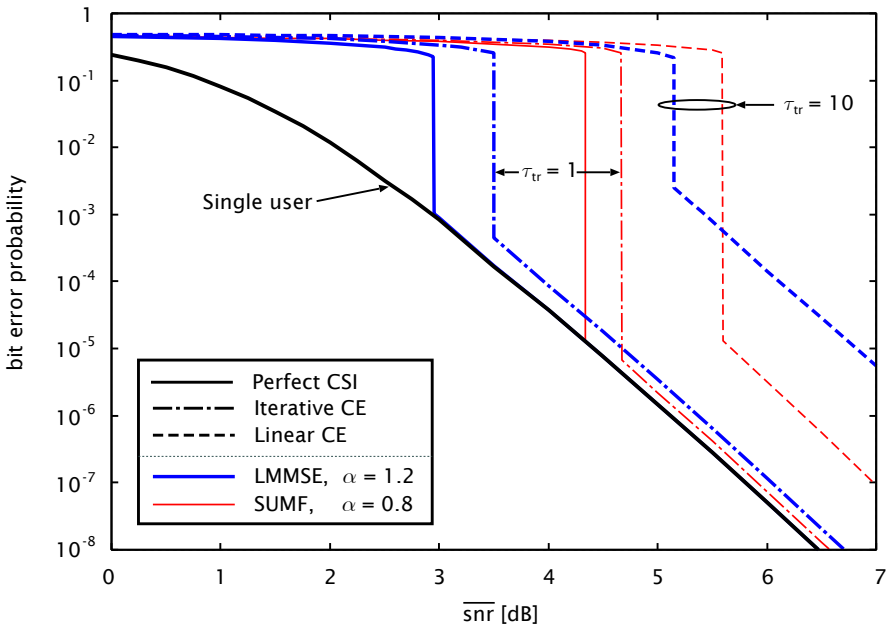



**Figure 4.7.** Bit error probability vs. SNR for the LMMSE-PIC MUDD and the SUMF with soft feedback. Iterative or non-iterative LMMSE-CE. Three equal power paths and coherence time of $T_{\text{coh}} = 101$ symbols. Rate-$1/2$ convolutional code with generator polynomials $(561, 753)_8$ and Gray encoded QPSK.

Bit error probability vs. SNR for an iterative system consisting of the LMMSE or SUMF-based MUDDs with soft PIC and the iterative LMMSE-CE is shown in Figure 4.7. The user load is set to $\alpha = 1.2$, channel coherence time of $T_{\text{coh}} = 101$ symbols and three $M = 3$ equal power paths are assumed. This time we have not plotted the BEP lower bounds for clarity. As the DE-GA analysis in Figure 4.4 already implied, using iterative channel estimation and MUDD only one pilot is needed to converge to the BEP lower bound (not shown) and close to single-user performance with perfect CSI. Increasing the training overhead to ten pilots but using non-iterative LMMSE channel estimation, however, causes an additional 2 dB loss in performance. Due to the relatively poor MAI suppression capability of the SUMF even soft feedback is used, the load was reduced $\alpha = 0.8$ in order to converge within the given range. This serves as an example of the different convergence properties of the different iterative MUDDs although we found previously that their genie-aided performances are essentially equal. We conclude that mismatch in channel infor-

mation can easily destroy the benefits of having a data detector capable of efficient MAI suppression, especially if the system needs to operate at relatively low SNR. This can, however, be effectively circumvented by using iterative channel estimator utilizing soft feedback.

The previous result showed that for fixed channel load $\alpha = 1.2$, iterative CE was able to provide the same BEP as linear CE, with 2 dBs lower SNR and only one pilot symbol when LMMSE-PIC MUDD was employed at the receiver. We next consider the case when the system is operating at moderate to high SNR and we are allowed to vary the user load in order to achieve maximum throughput. The loss in spectral efficiency due to transmission of known training symbols is taken into account in the results. In addition to the LMMSE-based data decoder, the non-linear MAP-MUDD is considered as well.

Spectral efficiency vs. training overhead for two mobility scenarios, namely $\Upsilon = T_{\mathsf{coh}}/M = 30$ and $\Upsilon = 80$, under the simplifying assumptions of Corollary 3 are plotted in Figures 4.8a and 4.8b, respectively. The system load is adjusted to meet the minimum bit error rate requirement BER $\leq 10^{-5}$ and only selected combinations of system parameters are plotted for clarity. We know that the LMMSE based channel estimator discussed in Section 4.1.3 is suboptimal when there is uncertainty in the transmitted symbols. Obtaining an upper bound for its performance by studying the optimum estimator discussed in Section 4.1.2 is, however, difficult. We therefore plot instead an upper bound for the considered channel estimator by assuming a genie-aided feedback, much like we did previously with the iterative MUDDs. We remark the following:

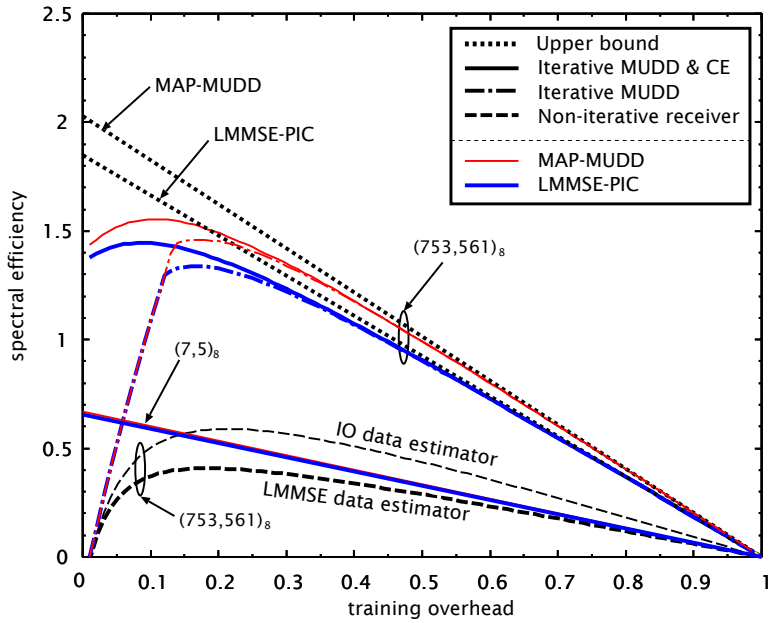
- The spectral efficiency with the $(5, 7)_8$ code and non-iterative channel estimation was found to be close zero in all cases. The corresponding curves were therefore omitted from the figures.
- Significant improvement over the non-iterative data estimators studied in [59, 74, 85] can be achieved by using iterative MUDD, even with non-iterative LMMSE channel estimator. As expected, the receivers using non-linear MAP detector show notable gains in spectral efficiency over the LMMSE based receivers. The difference is, however, smaller in the iterative cases.
- For $\Upsilon = 30$, the upper bounds (omitted) and the curves for the fully iterative receiver overlap almost perfectly in the case of $(5, 7)_8$ code. The maximum spectral efficience is, however, around $0.68$ bits per chip. Note that throughput of over $> 1.3$ bits per chip is achievable with non-iterative channel estimator and $(561, 753)_8$ code. This opposite to the case of perfect channel knowledge shown in Figure 4.3. There the system with iterative MUDD

achieves higher spectral efficiency with the $(5, 7)_8$ code for the same SNR.

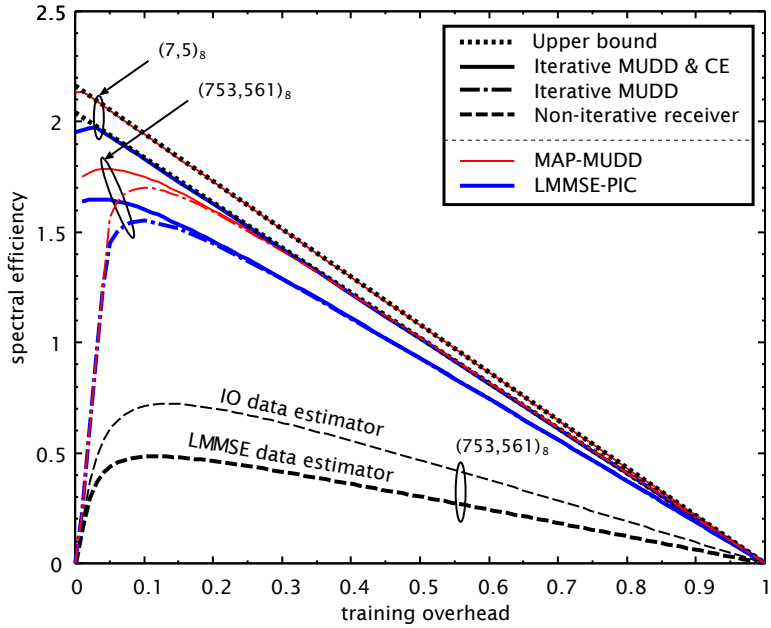
- For $\Upsilon = 80$, the highest loads with iterative MUDD and CE are obtained by using the $(5, 7)_8$ code. In this case, the performance of the iterative LMMSE channel estimator also follows closely the upper bound, showing that everything else being equal, little can be gained by using a more complex channel estimator.
- Iterations over the channel estimator provide only minor improvements in the spectral efficiency if the $(561, 753)_8$ code is used, and the performance is quite far from the upper bounds in this case. This hints that matching the channel code to the provisional channel conditions might be very important for the iterative MUDD and CE.

Figures 4.9a and 4.9b show the minimum training overhead that achieves the target BEP $\leq 10^{-5}$ as a function of the inverse channel spread factor $\Upsilon$, for the channel loads $\alpha = 1.0$ and $\alpha = 1.8$, respectively. The simplifying assumptions of Corollary 3 are considered and the receiver is equipped with the LMMSE-CE and LMMSE-PIC MUDD or MAP-MUDD. The following is observed:

- For channel load $\alpha = 1.0$, the iterative receiver allows for successful communication with vanishing pilot overhead if $\Upsilon = 12$ for the $(561, 753)_8$ code, and $\Upsilon = 42$ for the $(5, 7)_8$ code. In the case of $\alpha = 1.8$, the situation reverses and vanishing pilot overhead within the plotted region for both LMMSE-PIC MUDD and MAP-MUDD is achieved when using the $(5, 7)_8$ code, whereas MAP-MUDD is required for the $(561, 753)_8$ code.
- With the shorter constraint length code, performing iterations over the channel estimator allows for transmission with negligible pilot overhead whenever coherent communication is possible with the given system set-up. Thus, there is a phase transition in the amount of required training overhead as a function of the channel spread factor.
- With the LMMSE-based channel estimator and $(5, 7)_8$ code, the LMMSE-PIC MUDD achieves in practice optimum performance under the given system parameters.
- Using the constraint length nine code requires uniformly less training than the constraint length three code for the same error rate performance when the load is $\alpha = 1.0$. This is true also for the higher load $\alpha = 1.8$ if linear channel estimation is used. For fully iterative receiver, however, using the $(5, 7)_8$ code is beneficial if $\Upsilon$ is sufficiently large.

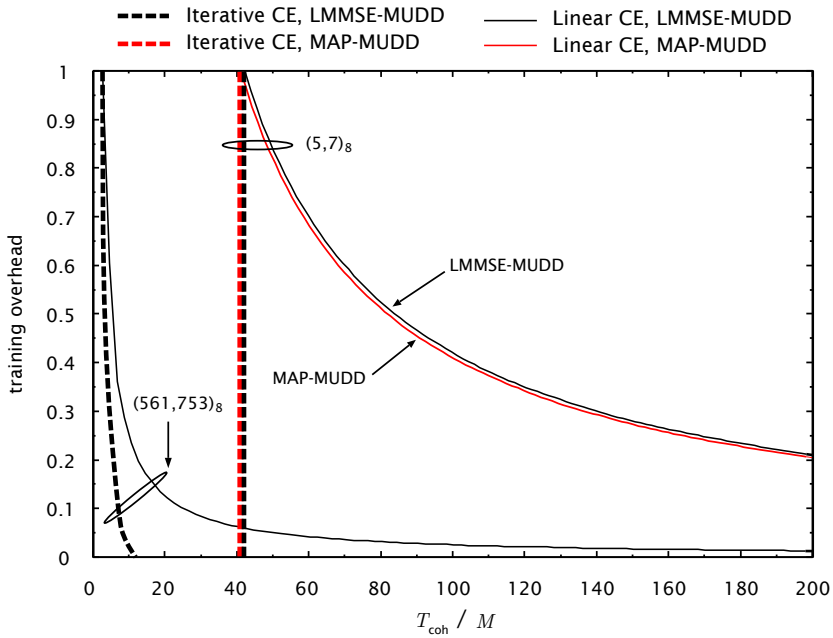



**(a)** Inverse channel spread factor $\Upsilon = 30$. For $(5,7)_8$ code, the upper bounds are indistinguishable from the iterative receiver and therefore omitted.

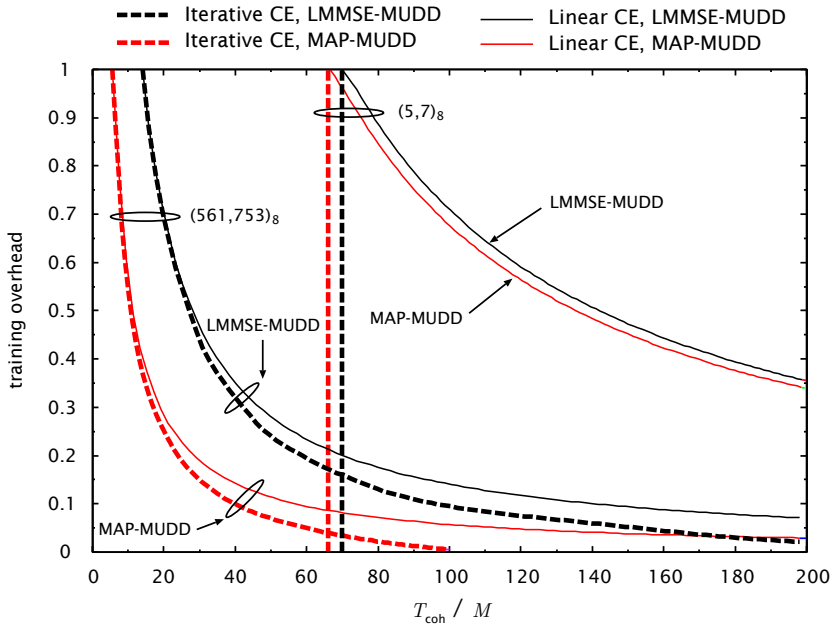



**(b)** Inverse channel spread factor $\Upsilon = 80$. For $(561,753)_8$ code, the upper bounds are not drawn for clarity.

**Figure 4.8.** Spectral efficiency $\alpha R(1-\Delta_{\mathsf{tr}})$ vs. the training overhead $\Delta_{\mathsf{tr}}$. Average SNR of 6 dB, target BER $\leq 10^{-5}$, and convolutional code $(5,7)_8$ or $(561,753)_8$.




**(a)** User load $\alpha = K/L = 1.0$. For the $(561, 753)_8$ code, the MAP-MUDD and LMMSE-PIC MUDD curves are overlapping.




**(b)** User load $\alpha = K/L = 1.8$.

**Figure 4.9.** Minimum training overhead vs. $\Upsilon = T_{\text{coh}}/M$ for target BER $\leq 10^{-5}$. Average SNR of 6 dB, LMMSE channel estimator and convolutional code $(5, 7)_8$ or $(561, 753)_8$.

**Table 4.2.** Minimum $\Upsilon$ for $\Delta_{\mathsf{tr}} < 0.001$ and BER $< 10^{-5}$ at $\overline{\mathsf{snr}} = 6$ dB

| Load | Code | Pilot | Bias |
|---|---|---|---|
| $\alpha = 1.0$ | $(7,5)_8$ | 43 | 42 |
| | $(753,561)_8$ | 12 | 12 |
| $\alpha = 1.8$ | $(7,5)_8$ | 71 | 71 |
| | $(753,561)_8$ | — | 200 |

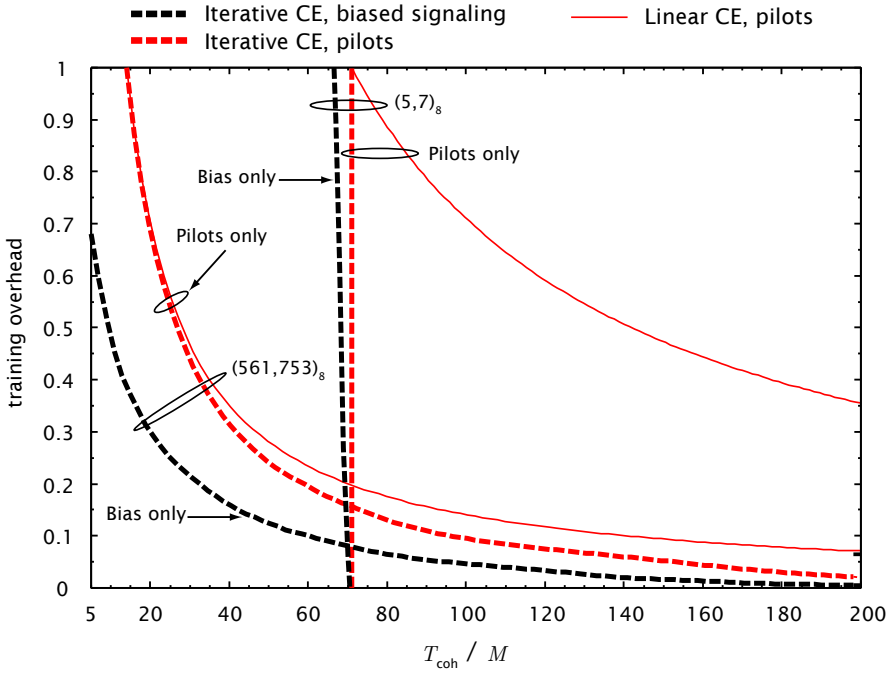



**Figure 4.10.** Minimum training overhead vs. $\Upsilon = T_{\mathsf{coh}}/M$ for target BER $\leq 10^{-5}$. Pilot symbol and biased signaling based channel estimation. Average SNR of 6 dB, LMMSE channel estimator, LMMSE-PIC MUDD and convolutional code $(5,7)_8$ or $(561,753)_8$. User load $\alpha = K/L = 1.8$.

Total training overhead $\Delta_{\mathsf{tot}}$ (see Section 2.4) vs. the inverse channel spread factor $\Upsilon = T_{\mathsf{coh}}/M$ for the iterative receiver with LMMSE-PIC MUDD and LMMSE-CE is shown in Figure 4.10. The average SNR is $\overline{\mathsf{snr}} = 6$ dB, user load is fixed at $\alpha = 1.8$ and target bit error rate BER is set to $10^{-5}$. Conventional pilot-aided channel estimation and the probability biased signaling introduced in Section 2.4 are considered. Table 4.2 summarizes the approximate minimum $\Upsilon \in [0, 200]$ that

achieves BER $< 10^{-5}$ and training overhead $\Delta_{\mathsf{tot}} < 0.001$ with iterative channel estimation and MUDD for the same system parameters. The following is observed:

1. With biased signaling the phase transition for the $(7, 5)_8$ code disappears, although the required amount of training still drops from maximum to negligible within a very narrow region.
2. With iterative CE and MUDD, the numerical experiments suggest that the bias based ($\Delta_{\mathsf{tr}} = 0$) channel estimation is uniformly better than the traditional approach of transmitting known pilots. This is in agreement with the theoretical findings in [172]. We would like to point out, however, that a more careful study of the combination of bias transformation and coding has to be carried out before final conclusions.

## 4.5 Chapter Summary and Conclusions

In this chapter, the large system analysis of randomly spread code division multiple access over a frequency-selective Rayleigh fading channel was considered. Iterative channel estimation and multiuser detection based on extrinsic feedback from the single-user decoders was studied. By means of the replica method, both estimators were shown to have an equivalent decoupled single-user characterization in the large system limit, that could be analyzed separately. In contrast to some earlier results, we took into account the CSI mismatch in the iterative multiuser decoding and studied an iterative channel estimator that utilized information feedback to refine the initial training symbol based decisions.

The specific channel estimators considered included an LMMSE-CE with soft feedback and an approximate ML-CE that used hard feedback. The iterative data estimators included non-linear MAP-MUDD and LMMSE-PIC MUDD, both of which used soft feedback. Single-user matched filter with soft or hard feedback was also considered. The performance of the system was investigated by means of DE-GA analysis. Analytical evaluation of the bit error rate and spectral efficiency was carried out. In addition to new results regarding soft feedback based channel estimation and data detection and decoding, we also considered the hard feedback based scheme studied previously in [138]. Our result is exact in the large system limit, whereas the previous result was obtained by making several approximations in the analysis.

The theoretical results indicated that the soft feedback has never detrimental effect on the estimators that take into account the error statistics. On the other hand, the hard feedback can increase both the MSE of the channel estimates and the BER

of the bit decisions. In contrast to the case of perfect CSI, where all considered estimators converge to single-user performance when the feedback tends to correct decisions, with channel mismatch the performance of the iterative SUMF can suffer a loss of up to 3 dBs in the output SINR compared to LMMSE or MAP MUDDs. For reasonable system configurations, however, the difference was found to be very small.

Via numerical examples, it was demonstrated that for loads of up to $\alpha = 1.8$, iterative channel estimation and MUDD was able to meet the desired quality of service (in terms of target BER) with vanishing training overhead for all practical channel conditions. The iterative CE was also found to provide superior performance compared to simple pilot-aided CE when the goal was to maximize the system spectral efficiency. With non-iterative channel estimation the system was found to be very vulnerable to underestimated training overhead, whereas iterative CE provided robust performance also in scenarios where the amount of pilots was well below the optimum. Interestingly, when iterative channel estimation was used in the system, the effect of ECC on the performance of the entire system became highly non-trivial and dependent on the channel parameters. This suggests that matching the code to both the CE and MUDD, as well as, the provisional channel conditions is an important part of optimizing the system performance.

Finally, the novel training method based on probability biased signaling and introduced in Section 2.4 was examined via numerical examples. It was found that the proposed scheme can provide performance gain over the traditional pilot symbol based channel estimation. Further research is, however, needed to efficiently implement the signal biasing scheme in a way that does not hamper the error rate performance of the error correction code or cause severe degradation in the spectral efficiency of the system.

# Chapter 5

# Conclusions

## 5.1 Summary and Discussion

The large system analysis of randomly spread CDMA channels with mismatched CSI at the receiver was carried out. Both channel estimation and multiuser detection and decoding algorithms were investigated with and without information feedback. Flat fading multi-antenna channels and multipath fading single-antenna channels were considered.

The signal model for the single and multiple antenna systems studied in the latter parts of the thesis were introduced in Chapter 2. Necessary background information on channel coding schemes and the mathematical methods used in the analysis were discussed. A novel training method based on probability biased signaling was proposed.

Multi-antenna CDMA with per-antenna random spreading was the topic of the Chapter 3. Spatially correlated block Rayleigh fading MIMO channels were considered. A set of linear channel estimators and non-iterative multiuser detectors were derived as special instances of a general Bayesian inference problem. In addition to the optimum pilot-aided LMMSE-CE, several suboptimal channel estimators were considered. The multiuser detectors included the non-linear MAP-MUD as well as the linear MMSE, decorrelating and conventional detectors. By an application of the replica method, the multiuser system with the derived estimators was shown to admit an equivalent single-user characterization in the large system limit. Using the decoupled channel model, the QPSK constrained ergodic spectral efficiency with single-user decoding was obtained. The analytical results showed that when pilot-aided channel estimation is employed at the receiver, the ergodic capacity of the system increases with the correlation between the transmit antennas. This obser-

vation was in contrast to the previously reported studies considering perfect CSI [92, 154], where the transmit correlation was found to have no effect on the ergodic spectral efficiency of the system. Notably, no information at the transmitter was required to attain the improvement in performance, but the channel estimator was required to have a priori knowledge about the long term transmit correlation.

In Chapter 4, iterative DS-CDMA receivers for multipath fading channels were considered. An iterative algorithm based on extrinsic information exchange between the channel estimator and the MUDD block was proposed. Both hard and soft feedback schemes, as well as linear and non-linear estimators were included in the framework. The multiuser system with the given iterative receiver was shown to decouple to a set of independent single-user channels via the replica method. The bit error rate performance and spectral efficiency were studied. The analytical results showed that while the use of soft feedback could never impede the performance of the iterative process, the use of hard feedback was potentially detrimental. The numerical experiments suggested that near single-user BER performance with perfect CSI was attainable in overloaded multiuser systems using channel estimation. Furthermore, this could be achieved with a vanishing training overhead when the system bandwidth and data transmission rate was allowed to grow in proportion with the amount of resolvable multipath components in the channel. Optimizing the user load for maximum system spectral efficiency as a function of training overhead revealed that iterative MUDD provided significant gains over non-iterative MUD even if linear pilot-aided channel estimation was used. As expected, further improvements were obtained by using iterative channel estimation. As a side effect, with iterative CE the system became also very robust against underestimated training overhead — something the non-iterative system was found to be vulnerable to. Taking the channel estimation into account in the analysis was also revealed to affect the optimal choice of error correction code in a non-trivial manner. Matching the ECC to the projected channel conditions was observed to have a great impact on the maximum achievable spectral efficiency.

## 5.2 Contributions of the Thesis

The main contributions of the thesis are summarized below.

***Large system analysis of MIMO DS-CDMA systems in spatially correlated channels:***

- The results extend the previous results reported in [92, 154] to the case of mismatched channel information at the MUD;
- Related work concerning uncorrelated MIMO channels can be found in [183] and [199]. Here the analysis was extended to antenna correlation and to channel estimators that were misinformed about the channel statistics;
- The observation that transmit correlation can be very beneficial for multi-antenna communication, even when not known at the transmitter, may provide new design approaches for practical systems.

***Large system analysis of iterative DS-CDMA systems in multipath fading channels:***

- The results of [59] were generalized to iterative channel estimation and multiuser detection and decoding. Equivalently, the work presented in this thesis can be seen as a follow-up to [125, 126] where iterative MUDD with perfect CSI was considered. The previously reported results in [137] were extended to iterative channel estimation and the analysis of [138] were performed without resorting to approximations.
- To the best of our knowledge, the analysis of general iterative receiver with soft / hard feedback and linear / non-linear estimators was performed the first time. Related work that derives capacity bounds for optimal receivers with channel estimation can be found in [172, 200].
- The finding that iterative channel estimation and MUDD provides spectrally very efficient method for coherent communication may increase the interest to fully iterative systems also in practical applications. Furthermore, the observation that the choice of channel code depends heavily on the channel conditions might provide new approaches to how to optimize the overall performance of a system with channel estimation.

## 5.3 Future Research Directions

In the following we provide some future research topics related to the work carried our in the present dissertation that might be of interest.

***Multi-Antenna Transmission Schemes:***

- *Iterative receivers for MIMO DS-CDMA.* Here the study of multi-antenna systems was limited to non-iterative receiver. As the results in Chapter 4 showed, iterative channel and data estimation have potential to provide significant improvements in performance over non-iterative receivers. Natural continuation of the research in Chapter 3 is to extend it to the iterative case of Chapter 4.
- *Comparison of the MIMO DS-CDMA considered in this thesis to multiuser MIMO without per-antenna spreading.* As shown in [92, 93], the per-antenna spreading is simpler but sub-optimal approach for uncorrelated MIMO channels with perfect CSI at the receiver. How the combination of channel estimation and antenna correlation would affect the situation is an interesting topic for investigation.
- *Channel information at the transmitter.* Utilizing some form of a priori channel knowledge at the transmitter is known to improve the performance of both single-user and multiuser communication systems (see for example [8, 9] and references therein). Investigating adaptive coding and modulation or linear pre-coding schemes in systems with channel estimation and CSI mismatch at the receiver would be an important topic, especially if the system is delay constrained.

***Single-Antenna Iterative Receivers:***

- *Code optimization.* In the present thesis, only two convolutional codes with constraint lengths three and nine were considered. An important follow-up would be a careful investigation of code optimization for systems employing iterative channel estimation and MUDD, for example, in the spirit of [134] where LDPC codes in MIMO systems were considered.
- *User power profile optimization.* The asymptotic analysis of [126] revealed that optimizing the received power profile of the users via linear programming allowed for greatly improved channel loads and spectral efficiency of iterative MUDD. Interesting future research topic would be to extend this approach to the iterative receivers considered in the present dissertation.

# Appendix A

# Diagonalization of the Noise Covariance Matrices $\tilde{C}$ and $\tilde{D}$ for MIMO DS-CDMA

Here we show that the postulated noise covariance matrices $\tilde{\boldsymbol{C}}$ and $\tilde{\boldsymbol{D}}$ of the decoupled single-user channels presented in Section 3.2, and the postulated correlation matrix at the receiver side $\tilde{\boldsymbol{R}}$, are simultaneously diagonalized by a unitary matrix. The same result follows immediately for the covariance matrices $\boldsymbol{C}$ and $\boldsymbol{D}$, only with the postulated covariance matrix $\tilde{\boldsymbol{R}}$ replaced by the correct one $\boldsymbol{R}$.

Let us first consider the postulated noise covariance of the channel estimator

$$\tilde{\boldsymbol{C}} = \tilde{\sigma}^2 \boldsymbol{I}_N + \alpha \sum_{m=1}^{M} \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_m}. \tag{A.1}$$

Since $\tilde{\boldsymbol{C}}$ is Hermitian, we can write $\tilde{\boldsymbol{C}} = \boldsymbol{V} \tilde{\boldsymbol{\Lambda}}_{\boldsymbol{C}} \boldsymbol{V}^{\mathsf{H}}$, where $\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{C}}$ is a diagonal matrix containing the eigenvalues of $\tilde{\boldsymbol{C}}$ while the columns of $\boldsymbol{V}$ are the corresponding eigenvectors. We can thus write the RHS of (A.1) as

$$\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{C}} = \tilde{\sigma}^2 \boldsymbol{I}_N + \alpha \boldsymbol{V}^{\mathsf{H}} \Big( \sum_{m=1}^{M} \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_m} \Big) \boldsymbol{V} = \tilde{\sigma}^2 \boldsymbol{I}_N + \alpha \tilde{\boldsymbol{\Lambda}}_{\Delta \boldsymbol{h}_1 + \cdots + \Delta \boldsymbol{h}_M}, \tag{A.2}$$

where the diagonal matrix $\tilde{\boldsymbol{\Lambda}}_{\Delta \boldsymbol{h}_1 + \cdots + \Delta \boldsymbol{h}_M}$ contains the eigenvalues of $\sum_{m=1}^{M} \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_m}$ [201]. If $\boldsymbol{V}$ simultaneously diagonalizes $\{\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_m}\}_{m=1}^{M}$, i.e., they all share the same eigenvectors but possibly different eigenvalues, then

$$\tilde{\boldsymbol{\Lambda}}_{\Delta \boldsymbol{h}_1 + \cdots + \Delta \boldsymbol{h}_M} = \tilde{\boldsymbol{\Lambda}}_{\Delta \boldsymbol{h}_1} + \cdots + \tilde{\boldsymbol{\Lambda}}_{\Delta \boldsymbol{h}_M}, \tag{A.3}$$

where $\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_m} = \boldsymbol{V} \tilde{\boldsymbol{\Lambda}}_{\Delta \boldsymbol{h}_m} \boldsymbol{V}^{\mathsf{H}}$, for all $m = 1, \ldots, M$. The other, rather pathological possibility, is that for some subset of $\{1, \ldots, M\}$ we have $\boldsymbol{V}^{\mathsf{H}} \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_m} \boldsymbol{V} \neq \tilde{\boldsymbol{\Lambda}}_{\Delta \boldsymbol{h}_m}$

while the sum of these matrices is diagonal. This would mean that their off-diagonal elements cancel each other perfectly in the summation. However, the structure of the estimators considered in this thesis do not allow for this to happen since for all $m$, the noise and receive antenna covariances are the same.

Now, let us consider the first case. Recalling that $\boldsymbol{R} = \boldsymbol{U}\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{R}}\boldsymbol{U}^{\mathsf{H}}$,

$$\begin{aligned}\tilde{\boldsymbol{\Omega}}_{\Delta\underline{\boldsymbol{H}}} &= \tilde{\boldsymbol{\Omega}}_{\underline{\boldsymbol{H}}}\big[\tau_{\mathsf{tr}}\tilde{\boldsymbol{\Omega}}_{\underline{\boldsymbol{H}}} + (\boldsymbol{I}_N \otimes \tilde{\boldsymbol{C}})\big]^{-1}(\boldsymbol{I}_N \otimes \tilde{\boldsymbol{C}}) \\ &= \tilde{\boldsymbol{\Omega}}_{\underline{\boldsymbol{H}}} - \tilde{\boldsymbol{\Omega}}_{\underline{\boldsymbol{H}}}\big[\tilde{\boldsymbol{\Omega}}_{\underline{\boldsymbol{H}}} + (\boldsymbol{I}_N \otimes \tilde{\boldsymbol{C}}/\tau_{\mathsf{tr}})\big]^{-1}\tilde{\boldsymbol{\Omega}}_{\underline{\boldsymbol{H}}},\end{aligned} \tag{A.4}$$

and denoting $\hat{\boldsymbol{U}} = \boldsymbol{V}^{\mathsf{H}}\boldsymbol{U}$, we can write

$$\begin{aligned}&(\boldsymbol{I}_N \otimes \boldsymbol{V}^{\mathsf{H}})\tilde{\boldsymbol{\Omega}}_{\Delta\underline{\boldsymbol{H}}}(\boldsymbol{I}_N \otimes \boldsymbol{V}) \\ &\quad= (\tilde{\boldsymbol{T}} \otimes \hat{\boldsymbol{U}}\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{R}}\hat{\boldsymbol{U}}^{\mathsf{H}}) - (\tilde{\boldsymbol{T}} \otimes \hat{\boldsymbol{U}}\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{R}}\boldsymbol{U}^{\mathsf{H}})\big[(\tilde{\boldsymbol{T}} \otimes \boldsymbol{U}\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{R}}\boldsymbol{U}^{\mathsf{H}}) \\ &\qquad\qquad + (\boldsymbol{I}_N \otimes \boldsymbol{V}\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{C}}\boldsymbol{V}^{\mathsf{H}}/\tau_{\mathsf{tr}})\big]^{-1}(\tilde{\boldsymbol{T}} \otimes \boldsymbol{U}\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{R}}\hat{\boldsymbol{U}}^{\mathsf{H}}) \\ &\quad= (\tilde{\boldsymbol{T}} \otimes \hat{\boldsymbol{U}}\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{R}}\hat{\boldsymbol{U}}^{\mathsf{H}}) - (\tilde{\boldsymbol{T}} \otimes \hat{\boldsymbol{U}}^{\mathsf{H}})\Big\{(\boldsymbol{I}_M \otimes \tilde{\boldsymbol{\Lambda}}_{\boldsymbol{R}}) \\ &\qquad\times\big[(\tilde{\boldsymbol{T}} \otimes \tilde{\boldsymbol{\Lambda}}_{\boldsymbol{R}}) + (\boldsymbol{I}_N \otimes \hat{\boldsymbol{U}}^{\mathsf{H}}\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{C}}\hat{\boldsymbol{U}}/\tau_{\mathsf{tr}})\big]^{-1}(\boldsymbol{I}_M \otimes \tilde{\boldsymbol{\Lambda}}_{\boldsymbol{R}})\Big\}(\tilde{\boldsymbol{T}} \otimes \hat{\boldsymbol{U}}^{\mathsf{H}})\end{aligned} \tag{A.5}$$

By assumption, the $N \times N$ main diagonals of (A.5) have to be diagonal matrices, which is satisfied if and only if $\hat{\boldsymbol{U}}$ is diagonal. Since $\hat{\boldsymbol{U}}$ is unitary, $\hat{\boldsymbol{U}} = \boldsymbol{I}_N \implies \boldsymbol{V} = \boldsymbol{U}$. As a result, $\boldsymbol{U}$ simultaneously diagonalizes $\tilde{\boldsymbol{C}}$, $\tilde{\boldsymbol{R}}$, $\{\tilde{\boldsymbol{\Omega}}_{\Delta\boldsymbol{h}_m}\}_{m=1}^{M}$ and $\{\tilde{\boldsymbol{\Omega}}_{\langle\tilde{\boldsymbol{h}}_m\rangle}\}_{m=1}^{M}$.

Now, consider the linear MUD. Along the lines of Appendix D, we get from Section 3.2.3 that

$$\tilde{\boldsymbol{D}} = \tilde{\sigma}^2\boldsymbol{I}_N + \alpha\sum_{m=1}^{M}\mathbb{E}\big\{\tilde{\boldsymbol{D}}(\tilde{\boldsymbol{D}} + \tilde{\boldsymbol{\Omega}}_{\Delta\boldsymbol{v}_m})^{-1}(\langle\tilde{\boldsymbol{h}}_m\rangle\boldsymbol{m}_m^{\mathsf{H}}\tilde{\boldsymbol{D}} + \tilde{\boldsymbol{\Omega}}_{\Delta\boldsymbol{v}_m})\big\}, \tag{A.6}$$

where

$$\boldsymbol{m}_m^{\mathsf{H}} = \frac{\langle\tilde{\boldsymbol{h}}_m\rangle^{\mathsf{H}}(\tilde{\boldsymbol{D}} + \tilde{\boldsymbol{\Omega}}_{\Delta\boldsymbol{v}_m})^{-1}}{1 + \langle\tilde{\boldsymbol{h}}_m\rangle^{\mathsf{H}}(\tilde{\boldsymbol{D}} + \tilde{\boldsymbol{\Omega}}_{\Delta\boldsymbol{v}_m})^{-1}\langle\tilde{\boldsymbol{h}}_m\rangle}. \tag{A.7}$$

From the previous discussion, we know that if $\tilde{\boldsymbol{R}} = \boldsymbol{U}\tilde{\boldsymbol{\Lambda}}_{\boldsymbol{R}}\boldsymbol{U}^{\mathsf{H}}$ then $\boldsymbol{\Omega}_{\langle\tilde{\boldsymbol{h}}_m\rangle} = \boldsymbol{U}\tilde{\boldsymbol{\Lambda}}_{\langle\tilde{\boldsymbol{h}}_m\rangle}\boldsymbol{U}^{\mathsf{H}}$ and $\tilde{\boldsymbol{\Omega}}_{\Delta\boldsymbol{h}_m} = \boldsymbol{U}\tilde{\boldsymbol{\Lambda}}_{\Delta\boldsymbol{h}_m}\boldsymbol{U}^{\mathsf{H}}$. For the detectors considered in this thesis $\tilde{\boldsymbol{\Omega}}_{\Delta\boldsymbol{v}_m} = \boldsymbol{0}$ or $\tilde{\boldsymbol{\Omega}}_{\Delta\boldsymbol{v}_m} = \boldsymbol{\Omega}_{\langle\tilde{\boldsymbol{h}}_m\rangle}$. For the first case, let $\boldsymbol{g} \sim \mathsf{CN}(\boldsymbol{0};\, \boldsymbol{I}_N)$ so that $\boldsymbol{U}\sqrt{\boldsymbol{\Lambda}_{\langle\tilde{\boldsymbol{h}}_m\rangle}}\boldsymbol{g}$ has the same distribution as $\langle\tilde{\boldsymbol{h}}_m\rangle$. Then,

$$\tilde{\boldsymbol{D}} = \boldsymbol{U}\left(\tilde{\sigma}^2\boldsymbol{I}_N + \alpha\sum_{m=1}^{M}\mathbb{E}\Big\{\frac{\sqrt{\boldsymbol{\Lambda}_{\langle\tilde{\boldsymbol{h}}_m\rangle}}\boldsymbol{g}\boldsymbol{g}^{\mathsf{H}}\sqrt{\boldsymbol{\Lambda}_{\langle\tilde{\boldsymbol{h}}_m\rangle}}}{1 + \boldsymbol{g}^{\mathsf{H}}\sqrt{\boldsymbol{\Lambda}_{\langle\tilde{\boldsymbol{h}}_m\rangle}}\boldsymbol{U}\tilde{\boldsymbol{D}}^{-1}\boldsymbol{U}^{\mathsf{H}}\sqrt{\boldsymbol{\Lambda}_{\langle\tilde{\boldsymbol{h}}_m\rangle}}\boldsymbol{g}}\Big\}\right)\boldsymbol{U}^{\mathsf{H}}, \tag{A.8}$$

and, therefore, $\boldsymbol{U}$ diagonalizes $\tilde{\boldsymbol{D}}$. In the latter case

$$\tilde{\boldsymbol{D}} = \tilde{\sigma}^2 \boldsymbol{I}_N + \alpha \sum_{m=1}^{M} \mathbb{E}\Big\{ \frac{\tilde{\boldsymbol{D}}(\tilde{\boldsymbol{D}} + \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_m})^{-1} \langle \tilde{\boldsymbol{h}}_m \rangle \langle \tilde{\boldsymbol{h}}_m \rangle^{\mathsf{H}} (\tilde{\boldsymbol{D}} + \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_m})^{-1} \tilde{\boldsymbol{D}}}{1 + \langle \tilde{\boldsymbol{h}}_m \rangle^{\mathsf{H}} (\tilde{\boldsymbol{D}} + \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_m})^{-1} \langle \tilde{\boldsymbol{h}}_m \rangle} + \tilde{\boldsymbol{D}} - \tilde{\boldsymbol{D}}(\tilde{\boldsymbol{D}} + \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_m})^{-1} \tilde{\boldsymbol{D}} \Big\}. \tag{A.9}$$

If $\tilde{\boldsymbol{D}} = \boldsymbol{V} \tilde{\boldsymbol{\Lambda}}_{\boldsymbol{D}} \boldsymbol{V}^{\mathsf{H}}$, by similar arguments as before, we get a condition that $\hat{\boldsymbol{U}} = \boldsymbol{U}^{\mathsf{H}} \boldsymbol{V}$ has to be a diagonal unitary matrix and, therefore, $\boldsymbol{V} = \boldsymbol{U}$. The MAP-MUD can be handled in a similar manner

# *Appendix B*

# *Derivation of* (4.38)

Consider the problem of obtaining the APPs of the data symbol $x_{\xi,t}$ of the $\xi$th user at iteration $\ell = 1, 2 \ldots$ and time instant $t \in \mathcal{D}$. Assume that $\{\langle \tilde{\boldsymbol{h}}_{k,t} \rangle_{(\ell)}\}_{k \in \mathcal{K}}$ and $\{\langle \tilde{x}_{k,t} \rangle_{\mathsf{ext}}^{(\ell-1)}\}_{j \in \mathcal{K} \setminus \xi}$ defined in (4.33) and (4.34), respectively, are available at the receiver. The received signal (2.10) can then be written as

$$\boldsymbol{y}_t = \frac{1}{\sqrt{L}} \boldsymbol{S}_{\xi,t} \langle \tilde{\boldsymbol{h}}_{\xi,t} \rangle_{(\ell)} x_{\xi,t} + \frac{1}{\sqrt{L}} \sum_{j \in \mathcal{K} \setminus \xi} \boldsymbol{S}_{k,t} \langle \tilde{\boldsymbol{h}}_{k,t} \rangle_{(\ell)} x_{j,t} + \frac{1}{\sqrt{L}} \sum_{k=1}^{K} \boldsymbol{S}_{k,t} \Delta \boldsymbol{v}_{k,t} + \boldsymbol{w}_t \in \mathbb{C}^L, \tag{B.1}$$

where $\Delta \boldsymbol{v}_{k,t} = \Delta \boldsymbol{h}_{k,t} x_{k,t} \in \mathbb{C}^M \ \forall k \in \mathcal{K}$ was defined in (4.36) – (4.37). In the limit of large code word length and for fixed coherence time $T_{\mathsf{coh}}$, we can regard $\Delta \boldsymbol{h}_{k,t}$ and $x_{k,t}$ to be independent. Therefore, if $\Delta \boldsymbol{h}_{k,t} \sim \mathsf{CN}(\boldsymbol{0}; \boldsymbol{\Omega}_{\Delta \boldsymbol{h}_{k,t}}^{(\ell)})$ then $\Delta \boldsymbol{v}_{k,t} \sim \mathsf{CN}(\boldsymbol{0}; \boldsymbol{\Omega}_{\Delta \boldsymbol{h}_{k,t}}^{(\ell)})$. By (4.34) and (4.35), the data symbols of the interfering users can be written as

$$x_{j,t} = \langle \tilde{x}_{j,t} \rangle_{\mathsf{ext}}^{(\ell-1)} + \Delta x_{j,t}, \qquad j \in \mathcal{K} \setminus \xi, \tag{B.2}$$

where $\Delta x_{j,t} \in \mathbb{C}$ is a random variable with conditional mean and variance

$$\mu_{\Delta x_{j,t}}^{(\ell-1)} = \mathbb{E}\{\Delta x_{j,t} \mid \mathcal{I}_{\xi,t}^{(\ell)}\} = 0, \qquad j \in \mathcal{K} \setminus \xi, \tag{B.3}$$

$$\Omega_{\Delta x_{j,t}}^{(\ell-1)} = \mathbb{E}\{|\Delta x_{j,t}|^2 \mid \mathcal{I}_{\xi,t}^{(\ell)}\}, \qquad j \in \mathcal{K} \setminus \xi, \tag{B.4}$$

respectively. Note, however, that neither for the hard nor the soft feedback the estimation error $\Delta x_{j,t}$ is Gaussian.

Now, postulate the conditional Gaussian prior for the interfering users

$$\tilde{x}_{j,t} \mid \mathcal{I}_{\xi,t}^{(\ell)} \sim \mathsf{CN}\left(\langle \tilde{x}_{j,t} \rangle_{\mathsf{ext}}^{(\ell-1)}; \tilde{\Omega}_{\Delta x_{j,t}}^{(\ell-1)}\right), \qquad j \in \mathcal{K} \setminus \xi, \tag{B.5}$$

where the mean $\langle \tilde{x}_{j,t} \rangle_{\mathsf{ext}}^{(\ell-1)}$ is known and $\tilde{\Omega}_{\Delta x_{j,t}}^{(\ell-1)}$ represents the detectors knowledge about (B.4). Let $\Delta \tilde{\boldsymbol{h}}_{k,t} \sim \mathsf{CN}(\boldsymbol{0}; \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{h}_{k,t}}^{(\ell)})$ be the receiver's knowledge about the channel mismatch related to (4.36). Unfortunately, with the Gaussian priors, the term $\Delta \boldsymbol{v}_{k,t}$ in (B.1) makes the estimator non-linear. Thus, we further postulate that $\{\Delta \boldsymbol{v}_{k,t}\}_{k=1}^{K}$ are independent Gaussian RVs

$$\Delta \tilde{\boldsymbol{v}}_{k,t} \mid \mathcal{I}_{\xi,t}^{(\ell)} \sim \mathsf{CN}(\boldsymbol{0}; \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_{k,t}}^{(\ell)}), \tag{B.6}$$

uncorrelated with $\{\tilde{x}_{j,t}\}_{j \in \mathcal{K} \setminus \xi}$, and let the channel model at the receiver be

$$\begin{aligned} \tilde{\boldsymbol{y}}_t = \frac{1}{\sqrt{L}} \boldsymbol{S}_{\xi,t} \langle \tilde{\boldsymbol{h}}_{\xi,t} \rangle_{(\ell)} \tilde{x}_{\xi,t} + \frac{1}{\sqrt{L}} \sum_{j \in \mathcal{K} \setminus \xi} \boldsymbol{S}_{j,t} \langle \tilde{\boldsymbol{h}}_{j,t} \rangle_{(\ell)} \tilde{x}_{j,t} \\ + \frac{1}{\sqrt{L}} \sum_{k=1}^{K} \boldsymbol{S}_{k,t} \Delta \tilde{\boldsymbol{v}}_{k,t} + \tilde{\boldsymbol{w}}_t \in \mathbb{C}^L, \end{aligned} \tag{B.7}$$

where $\tilde{\boldsymbol{w}}_t \sim \mathsf{CN}(\boldsymbol{0}; \tilde{\sigma}^2 \boldsymbol{I}_L)$. The marginalized posterior probabilities of the data symbol $x_{\xi,t}$ based on the channel model (B.7) reads

$$\begin{aligned} &\mathbb{Q}^{(\ell)}(\tilde{x}_{\xi,t} \mid \mathcal{I}_{\xi,t}^{(\ell)}) \\ &= \frac{\mathbb{Q}(\tilde{x}_{\xi,t})}{\mathbb{E}_{\{\tilde{x}_{k,t}\}_{k \in \mathcal{K}}} \Big\{ \mathbb{E}_{\{\Delta \tilde{\boldsymbol{v}}_{k,t}\}_{k \in \mathcal{K}}} \big\{ \mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t \mid \{\tilde{x}_{k,t}\}_{k=1}^{K}, \{\Delta \tilde{\boldsymbol{v}}_{k,t}\}_{k=1}^{K}, \mathcal{I}_{\xi,t}^{(\ell)}) \big\} \Big\}} \\ &\quad \times \mathbb{E}_{\{\tilde{x}_{j,t}\}_{j \in \mathcal{K} \setminus \xi}} \Big\{ \mathbb{E}_{\{\Delta \tilde{\boldsymbol{v}}_{k,t}\}_{k \in \mathcal{K}}} \big\{ \mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t \mid \{\tilde{x}_{k,t}\}_{k=1}^{K}, \{\Delta \tilde{\boldsymbol{v}}_{k,t}\}_{k=1}^{K}, \mathcal{I}_{\xi,t}^{(\ell)}) \big\} \Big\} \\ &= \frac{\mathbb{Q}(\tilde{x}_{\xi,t}) \mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t \mid \tilde{x}_{\xi,t}, \mathcal{I}_{\xi,t}^{(\ell)})}{\mathbb{E}_{\tilde{x}_{\xi,t}} \big\{ \mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t \mid \tilde{x}_{\xi,t}, \mathcal{I}_{\xi,t}^{(\ell)}) \big\}}, \end{aligned} \tag{B.8}$$

where $\mathbb{Q}(\tilde{x}_{\xi,t})$ is the postulated prior of the desired user's data symbol $x_{\xi,t}$. The expectations with respect to (B.5) and (B.6) in (B.8) and leading to (B.8), were be calculated with the help of the Gaussian integral (C.22). Furthermore, the resulting distribution

$$\mathbb{Q}^{(\ell)}(\tilde{\boldsymbol{y}}_t = \boldsymbol{y}_t \mid \tilde{x}_{\xi,t}, \mathcal{I}_{\xi,t}^{(\ell)}) = \mathsf{CN}(\tilde{\boldsymbol{\mu}}_{\xi}^{\mathsf{pic},(\ell)}; \tilde{\boldsymbol{\Omega}}_{\xi}^{\mathsf{pic},(\ell)}), \tag{B.9}$$

is complex Gaussian with mean and variance given by

$$\tilde{\boldsymbol{\mu}}_{\xi}^{\mathsf{pic},(\ell)} = \frac{1}{\sqrt{L}} \boldsymbol{S}_{\xi,t} \langle \tilde{\boldsymbol{h}}_{\xi,t} \rangle_{(\ell)} \tilde{x}_{\xi,t} + \frac{1}{\sqrt{L}} \sum_{j \in \mathcal{K} \setminus \xi} \boldsymbol{S}_{j,t} \langle \tilde{\boldsymbol{h}}_{j,t} \rangle_{(\ell)} \langle \tilde{x}_{j,t} \rangle_{(\ell)}, \tag{B.10}$$

$$\begin{aligned} \tilde{\boldsymbol{\Omega}}_{\xi}^{\mathsf{pic},(\ell)} &= \tilde{\sigma}^2 \boldsymbol{I}_L + \frac{1}{L} \boldsymbol{S}_{\xi,t} \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_{\xi,t}}^{(\ell)} \boldsymbol{S}_{\xi,t}^{\mathsf{H}} \\ &\quad + \frac{1}{L} \sum_{j \in \mathcal{K} \setminus \xi} \boldsymbol{S}_{j,t} \Big( \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_{j,t}}^{(\ell)} + \langle \tilde{\boldsymbol{h}}_{j,t} \rangle_{(\ell)} \tilde{\Omega}_{\Delta x_{j,t}}^{(\ell-1)} \langle \tilde{\boldsymbol{h}}_{j,t}^{\mathsf{H}} \rangle_{(\ell)} \Big) \boldsymbol{S}_{j,t}^{\mathsf{H}}, \end{aligned} \tag{B.11}$$

respectively. Note that the second term on the RHS of (B.10) corresponds to the parallel interference cancellation employed by the MUD front-end.

# Appendix C

# Proof of Claim 4

The proof of Claim 4 is divided in two parts. The first one gives an informal proof by deriving the free energy for the system and concluding from its form that the system decouples to parallel single-user channels. The second part modifies the derivation to show that the joint moments of the single-user and the multiuser systems coincide. The reason for two different derivations is that the first approach gives a more pedagogical presentation of the replica method in the context of Bayesian estimation by omitting some extra variables present in the latter part. It also provides an example how to derive the mutual information for a given system via free energy. The derivations follow closely the approach of [85, 89, 92] and differ in some parts slightly from the presentation given in Section 2.6. Note that similar calculations were also performed by Tanaka in a slightly different context in [74, Proof of Lemma 1 and Appendix III]. Here we consider only the RS solution of the free energy and leave the investigation of RSB as a future topic.

## C.1 Derivation of the Free Energy

Consider the channel estimator defined by (4.9) and (4.26). Let

$$\left\{\vec{\boldsymbol{s}}_{k,t,l} = \begin{bmatrix} s_{k,t,l,1} & \cdots & s_{k,t,l,M} \end{bmatrix}\right\}_{l=1}^{L}, \tag{C.1}$$

be the rows of the spreading matrix $\boldsymbol{S}_{k,t} \in \mathcal{M}^{L\times M}$ in (2.10) that is modified according to the Assumption 1. Fix the time index $\vartheta \in \mathcal{D}$, and define for notational convenience two diagonal matrices $\boldsymbol{S}_{k,[1],l,m} \in \mathbb{C}^{\tau_{\mathsf{tr}}\times\tau_{\mathsf{tr}}}$, $\boldsymbol{S}_{k,[2],l,m} \in \mathbb{C}^{(\tau_{\mathsf{d}}-1)\times(\tau_{\mathsf{d}}-1)}$ as

$$\boldsymbol{S}_{k,[1],l,m} = \operatorname{diag}\big(\begin{bmatrix} s_{k,1,l,m} & \cdots & s_{k,\tau_{\mathsf{tr}},l,m} \end{bmatrix}\big), \tag{C.2}$$

$$\boldsymbol{S}_{k,[2],l,m} = \operatorname{diag}\big(\begin{bmatrix} s_{k,\tau_{\mathsf{tr}}+1,l,m} & \cdots & s_{k,\vartheta-1,l,m} & s_{k,\vartheta+1,l,m} & \cdots & s_{k,T_{\mathsf{coh}},l,m} \end{bmatrix}\big), \tag{C.3}$$

respectively. Assuming the $\ell$th iteration at the receiver, we may write the input-output relation of the DS-CDMA channel over multipath fading as

$$\boldsymbol{y}_{[1],l} = \frac{1}{\sqrt{L}} \sum_{k=1}^{K} \sum_{m=1}^{M} \boldsymbol{S}_{k,[1],l,m} h_{k,m} \boldsymbol{p}_k + \boldsymbol{w}_{\mathcal{T},l} \in \mathbb{C}^{\tau_{\mathsf{tr}}}, \tag{C.4}$$

$$\boldsymbol{y}_{[2],l} = \frac{1}{\sqrt{L}} \sum_{k=1}^{K} \boldsymbol{S}_{k,[2],l,m} \big( h_{k,m} \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}}^{(\ell)} + \Delta \boldsymbol{u}_{k,m} \big) + \boldsymbol{w}_{\mathcal{D}\setminus\vartheta,l} \in \mathbb{C}^{\tau_{\mathsf{d}}-1}, \tag{C.5}$$

where $\Delta \boldsymbol{u}_{k,m}$ is defined in (4.12) and $\mathbb{P}(\boldsymbol{w}_{\dots,l}) = \mathsf{CN}(\boldsymbol{0};\, \sigma^2 \boldsymbol{I})$. Now the set of vectors $\{\boldsymbol{y}_{[1],l}, \boldsymbol{y}_{[2],l}\}_{l=1}^{L}$ in (C.4) – (C.5) has the same information as $\mathcal{Y} \setminus \boldsymbol{y}_\vartheta$ in (2.10). Similarly, let the postulated channel information $\{\tilde{\boldsymbol{y}}_{\mathcal{T}}, \tilde{\boldsymbol{y}}_{\mathcal{D}\setminus\vartheta}\}$ in (4.19) – (4.20) be written as $\{\tilde{\boldsymbol{y}}_{[1],l}, \tilde{\boldsymbol{y}}_{[2],l}\}_{l=1}^{L}$, where

$$\tilde{\boldsymbol{y}}_{[1],l} = \frac{1}{\sqrt{L}} \sum_{k=1}^{K} \sum_{m=1}^{M} \boldsymbol{S}_{k,[1],l,m} \tilde{h}_{k,\vartheta,m} \boldsymbol{p}_k + \tilde{\boldsymbol{w}}_{\mathcal{T},l} \in \mathbb{C}^{\tau_{\mathsf{tr}}}, \tag{C.6}$$

$$\tilde{\boldsymbol{y}}_{[2],l} = \frac{1}{\sqrt{L}} \sum_{k=1}^{K} \sum_{m=1}^{M} \boldsymbol{S}_{k,[2],l,m} \big( \tilde{h}_{k,\vartheta,m} \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}}^{(\ell)} + \Delta \tilde{\boldsymbol{u}}_{k,m} \big) + \tilde{\boldsymbol{w}}_{\mathcal{D}\setminus\vartheta,l} \in \mathbb{C}^{\tau_{\mathsf{d}}-1}, \tag{C.7}$$

and $\mathbb{Q}(\tilde{\boldsymbol{w}}_{\dots,l}) = \mathsf{CN}(\boldsymbol{0};\, \tilde{\sigma}^2 \boldsymbol{I})$.

In the following, we shall associate the zeroth replica index with the channel variables in (C.4) – (C.5) and write

$$\mathcal{H}^{\{0\}} = \{ h_{k,m}^{\{0\}} = h_{k,m} \mid \forall k, m \}, \qquad \Delta\mathcal{U}^{\{0\}} = \{ \Delta \boldsymbol{u}_{k,m}^{\{0\}} = \Delta \boldsymbol{u}_{k,m} \mid \forall k, m \}. \tag{C.8}$$

Similarly, the replica indices $a = 1, \dots, n$ are connected to the postulated variables in (C.6) – (C.7) and we denote with a slight abuse of notation

$$\mathcal{H}^{\{a\}} = \{ h_{k,m}^{\{a\}} \mid \forall k, m \}, \qquad \Delta\mathcal{U}^{\{a\}} = \{ \Delta \boldsymbol{u}_{k,m}^{\{a\}} \mid \forall k, m \}, \tag{C.9}$$

where the replicated RVs are assumed to be IID and drawn according to the same distribution as the postulated RVs $\{\tilde{h}_{k,\vartheta,m} \mid \forall k, m\}$ and $\{\Delta \tilde{\boldsymbol{u}}_{k,m} \mid \forall k, m\}$ in Section 4.1, i.e., $h_{k,m}^{\{a\}} \sim \mathbb{Q}^{(\ell)}(\tilde{h}_{k,\vartheta,m})$ and $\Delta \boldsymbol{u}_{k,m}^{\{a\}} \sim \mathbb{Q}^{(\ell)}(\Delta \tilde{\boldsymbol{u}}_{k,m})$. For notational convenience, the iteration index is omitted in the following discussion and we define a set

$$\mathcal{X}_{\mathsf{tot}} = \{ \boldsymbol{p}_k, \boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta}, \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}} \mid \forall k \}, \tag{C.10}$$

related to all transmit symbols and their estimates present in the system.

The denominator in (4.26) is the partition function of our system of interest and will be denoted by $Z$ in the following. Just like earlier in Section 2.6, the partition function could in theory be used to calculate interesting macroscopic parameters of our system, e.g., the MSE of the channel estimates. As we learned though, direct computation of $Z$ is infeasible and we therefore resort to computing the free energy via the replica method.

Before proceeding to the actual calculation of the free energy for the system arising from (C.4) – (C.7), we present some assumptions made on the course of the following replica analysis. It should be remarked that what is given below should be proved and not simply postulated to be true. This is, however, out of the scope of the present dissertation.

**Assumption 6 (Self-averaging property and replica continuity).** The free energy at the thermodynamic equilibrium is self-averaging with respect to the quenched randomness of $\{\mathcal{Y}, \mathcal{S}, \Delta\mathcal{U}^{\{0\}}\}$ (see Assumption 4), and can be written in the form

$$F_{\mathsf{rm}} = -\lim_{n\to 0} \frac{\partial}{\partial n} \lim_{K=\alpha L\to\infty} \frac{1}{K} \log \Xi^{K,n}, \tag{C.11}$$

where $\Xi^{K,n} = \mathbb{E}\{Z^n \mid \mathcal{X}_{\mathsf{tot}}\}$ and the expectation is conditioned on the true and estimated information about the transmitted signal as defined in (C.10). The $n$th power of the "moment generating function" $\Xi^{K,n}$ is evaluated for positive integers $n$ and analytic continuity in the vicinity of 0 is assumed to hold. ♢

The $L$ channels in $\{\boldsymbol{y}_{[\nu],l}\}_{l=1}^{L}$ and $\{\tilde{\boldsymbol{y}}_{[\nu],l}\}_{l=1}^{L}$ that arise from the matched filtering of the spreading waveforms are conditionally IID, and we may thus write

$$\Xi^{K,n} = \mathbb{E}\Bigg\{ \Bigg[ \prod_{\nu=1}^{2} \int \mathrm{d}\boldsymbol{y}_{[\nu]} \times \mathbb{E}_{\mathcal{S}_{[\nu]}} \Bigg\{ \prod_{a=0}^{n} \frac{1}{(\pi\sigma_a^2)^{\tau_{[\nu]}}} \exp\left( -\frac{1}{\sigma_a^2} \left\| \boldsymbol{y}_{[\nu]} - \sqrt{\alpha}\boldsymbol{v}_{[\nu]}^{\{a\}} \right\|^2 \right) \Bigg\} \Bigg]^{L} \Bigg| \mathcal{X}_{\mathsf{tot}} \Bigg\}. \tag{C.12}$$

We denoted above $\tau_{[1]} = \tau_{\mathsf{tr}}, \tau_{[2]} = \tau_{\mathsf{d}} - 1$ and $\sigma_0^2 = \sigma^2$, $\sigma_a^2 = \tilde{\sigma}^2, a = 1, 2, \ldots, n$ so that $\boldsymbol{y}_{[\nu]} \in \mathbb{C}^{\tau_{[\nu]}}, \nu = 1, 2$. The random matrices in $\mathcal{S}_{[\nu]} = \{\boldsymbol{S}_{k,[\nu],m} \,|\, \forall k, m\}$, for $\nu = 1, 2$, are independent with IID elements that are distributed as the elements of (C.2) – (C.3) for any $l = 1, \ldots, L$. The random vectors $\{\boldsymbol{v}_{[\nu]}^{\{a\}}\}_{a=0}^{n}$ in (C.12) are

given by

$$\boldsymbol{v}_{[1]}^{\{a\}} = \frac{1}{\sqrt{K}} \sum_{k=1}^{K} \sum_{m=1}^{M} \boldsymbol{S}_{k,[1],m} h_{k,m}^{\{a\}} \boldsymbol{p}_k \in \mathbb{C}^{\tau_{\mathsf{tr}}}, \tag{C.13}$$

$$\boldsymbol{v}_{[2]}^{\{a\}} = \frac{1}{\sqrt{K}} \sum_{k=1}^{K} \sum_{m=1}^{M} \boldsymbol{S}_{k,[2],m} \left( h_{k,m}^{\{a\}} \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}} + \Delta \boldsymbol{u}_{k,m}^{\{a\}} \right) \in \mathbb{C}^{\tau_{\mathsf{d}}-1}. \tag{C.14}$$

Let us define for $m = 1, \ldots, M$ and $k \in \mathcal{K}$ the RVs $\boldsymbol{\omega}_{k,[1],m} \in \mathbb{C}^{(n+1)\tau_{\mathsf{tr}}}$ and $\boldsymbol{\omega}_{k,[2],m} \in \mathbb{C}^{(n+1)(\tau_{\mathsf{d}}-1)}$ as

$$\boldsymbol{\omega}_{k,[1],m} = \mathrm{vec}\big([\boldsymbol{p}_k h_{k,m}^{\{0\}} \ \cdots \ \boldsymbol{p}_k h_{k,m}^{\{n\}}]\big), \tag{C.15}$$

$$\boldsymbol{\omega}_{k,[2],m} = \mathrm{vec}\big([h_{k,m}^{\{0\}} \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}} + \Delta \boldsymbol{u}_{k,m}^{\{0\}} \ \cdots \ h_{k,m}^{\{n\}} \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}} + \Delta \boldsymbol{u}_{k,m}^{\{n\}}]\big), \tag{C.16}$$

respectively, so that in the large system limit and conditioned on the set

$$\{\mathcal{X}_{\mathsf{tot}}, \{\mathcal{H}^{\{a\}}\}_{a=0}^{n}, \{\Delta\mathcal{U}^{\{a\}}\}_{a=0}^{n}\}, \tag{C.17}$$

the vectors

$$\boldsymbol{v}_{[\nu]} = \mathrm{vec}\big([\boldsymbol{v}_{[\nu]}^{\{0\}} \quad \boldsymbol{v}_{[\nu]}^{\{1\}} \ \cdots \ \boldsymbol{v}_{[\nu]}^{\{n\}}]\big) \in \mathbb{C}^{(n+1)\tau_{[\nu]}}, \quad \nu = 1, 2, \tag{C.18}$$

converge by the central limit theorem to independent zero-mean Gaussian RVs with conditional covariance matrices

$$\boldsymbol{Q}_{[\nu]} = \lim_{K\to\infty} \boldsymbol{Q}_{[\nu]}^{K} = \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \boldsymbol{\omega}_{k,[\nu],m} \boldsymbol{\omega}_{k,[\nu],m}^{\mathsf{H}} \in \mathbb{C}^{(n+1)\tau_{[\nu]}}, \quad \nu = 1, 2. \tag{C.19}$$

Following [74, Appendix II] it can be shown that for finite $K$,

$$\Xi^{K,n} = \mathbb{E}\left\{ \prod_{\nu=1}^{2} \exp\left[ K\alpha^{-1} \left( G_{[\nu]}^{K,n}(\boldsymbol{Q}_{[\nu]}^{K}) + O(K^{-1}) \right) \right] \,\middle|\, \mathcal{X}_{\mathsf{tot}} \right\}, \tag{C.20}$$

where

$$\begin{aligned}
&\exp\left( G_{[\nu]}^{K,n}(\boldsymbol{Q}_{[\nu]}^{K}) \right) \\
&= \frac{(\pi\sigma^2)^{-\tau_{[\nu]}}}{(\pi\tilde{\sigma}^2)^{n\tau_{[\nu]}}} \mathbb{E}_{\boldsymbol{v}_{[\nu]}} \left\{ \exp\left[ -\frac{\alpha}{\sigma^2} \|\boldsymbol{v}_{[\nu]}^{\{0\}}\|^2 - \frac{\alpha}{\tilde{\sigma}^2} \sum_{a=1}^{n} \|\boldsymbol{v}_{[\nu]}^{\{a\}}\|^2 \right] \right. \\
&\qquad \times \int \exp\left[ -\left( \frac{1}{\sigma^2} + \frac{n}{\tilde{\sigma}^2} \right) \|\boldsymbol{y}_{[\nu]}\|^2 \right. \\
&\qquad\qquad \left. \left. + 2\Re\left\{ \sqrt{\alpha} \left( \frac{1}{\sigma^2} \boldsymbol{v}_{[\nu]}^{\{0\}} + \frac{1}{\tilde{\sigma}^2} \sum_{a=1}^{n} \boldsymbol{v}_{[\nu]}^{\{a\}} \right)^{\mathsf{H}} \boldsymbol{y}_{[\nu]} \right\} \right] \mathrm{d}\boldsymbol{y}_{[\nu]} \,\middle|\, \boldsymbol{Q}_{[\nu]}^{K} \right\},
\end{aligned} \tag{C.21}$$

and $\boldsymbol{v}^{[\nu]} \sim \mathsf{CN}(\mathbf{0};\, \boldsymbol{Q}^K_{[\nu]}), \nu = 1, 2$, are independent Gaussian RVs. Let $\boldsymbol{M}$ be a positive definite matrix and apply the vector form complex Gaussian integral

$$\int \mathrm{e}^{-\boldsymbol{y}^{\mathsf{H}}\boldsymbol{M}\boldsymbol{y}+2\Re\{\boldsymbol{b}^{\mathsf{H}}\boldsymbol{y}\}}\mathrm{d}\boldsymbol{y} = \frac{\pi^{T_{\mathsf{coh}}}}{\det(\boldsymbol{M})}\mathrm{e}^{\boldsymbol{b}^{\mathsf{H}}\boldsymbol{M}^{-1}\boldsymbol{b}}, \qquad \boldsymbol{M} > 0, \tag{C.22}$$

to (C.21), first for the integrals with respect to $\boldsymbol{y}_{[\nu]}$, then again to calculate the expectations with respect to $\boldsymbol{v}_{[\nu]} \sim \mathsf{CN}(\mathbf{0};\, \boldsymbol{Q}^K_{[\nu]})$. The end result reads

$$\exp\left(G^{K,n}_{[\nu]}(\boldsymbol{Q}^K_{[\nu]})\right) = \left(\frac{\tilde{\sigma}^2}{\tilde{\sigma}^2 + n\sigma^2}\right)^{\tau_{[\nu]}} \frac{(\pi\tilde{\sigma}^2)^{-n\tau_{[\nu]}}}{\det\left(\boldsymbol{Q}^K_{[\nu]}\right)\det\left((\boldsymbol{Q}^K_{[\nu]})^{-1} + \boldsymbol{A}_{[\nu]}\right)}, \tag{C.23}$$

where $\boldsymbol{A}_{[\nu]} \in \mathbb{R}^{(n+1)\tau_{[\nu]}\times(n+1)\tau_{[\nu]}}$ is a symmetric matrix defined as

$$\boldsymbol{A}_{[\nu]} = \frac{\alpha}{\tilde{\sigma}^2 + n\sigma^2}\left[\begin{array}{c|c} n & -\boldsymbol{e}_n^{\mathsf{T}} \\ \hline -\boldsymbol{e}_n & (1 + n\frac{\sigma^2}{\tilde{\sigma}^2})\boldsymbol{I}_n - \frac{\sigma^2}{\tilde{\sigma}^2}\boldsymbol{e}_n\boldsymbol{e}_n^{\mathsf{T}} \end{array}\right] \otimes \boldsymbol{I}_{\tau_{[\nu]}}, \tag{C.24}$$

and $\boldsymbol{e}_n$ denotes for the all-ones vector of length $n$. For later use, we write

$$\begin{aligned} -G^{K,n}_{[\nu]}(\boldsymbol{Q}^K_{[\nu]}) = \tau_{[\nu]}&\left[(n-1)\log(\tilde{\sigma}^2) + \log\left(\tilde{\sigma}^2 + n\sigma^2\right)\right] \\ &+ n\tau_{[\nu]}\log\pi + \log\det\left(\boldsymbol{I}_{(n+1)\tau_{[\nu]}} + \boldsymbol{A}_{[\nu]}\boldsymbol{Q}^K_{[\nu]}\right). \end{aligned} \tag{C.25}$$

Let $\mathcal{V}_{[\nu]}$ be the set of positive definite $(n+1)\tau_{[\nu]}\times(n+1)\tau_{[\nu]}$ Hermitian matrices for $\nu = 1, 2$, and define the conditional probability measure on $\mathcal{V}_{[1]} \times \mathcal{V}_{[2]}$ as

$$\mu^K(\mathcal{V}) = \mathbb{E}\left\{\prod_{\nu=1}^{2} \mathbb{1}_{\boldsymbol{V}_{[\nu]}}\left[K\boldsymbol{Q}_{[\nu]} = \sum_{k=1}^{K}\sum_{m=1}^{M}\boldsymbol{\omega}_{k,[\nu],m}\boldsymbol{\omega}^{\mathsf{H}}_{k,[\nu],m}\right] \,\middle|\, \mathcal{X}_{\mathsf{tot}}\right\}, \tag{C.26}$$

where $\mathcal{V} = (\boldsymbol{V}_{[1]}, \boldsymbol{V}_{[2]}) \subset \mathcal{V}_{[1]} \times \mathcal{V}_{[2]}$ and $\mathbb{1}$ is the indicator function. Since the users are assumed to have independent channels, the moment generating function induced by (C.26) reads

$$\begin{aligned} M^{K,n}(\tilde{\mathcal{Q}}) &= \prod_{k=1}^{K} M^n_k(\tilde{\mathcal{Q}}) \\ &= \prod_{k=1}^{K} \mathbb{E}\left\{\exp\left[\sum_{\nu=1}^{2}\sum_{m=1}^{M}\mathrm{tr}\left(\boldsymbol{\omega}_{k,[\nu],m}\boldsymbol{\omega}^{\mathsf{H}}_{k,[\nu],m}\tilde{\boldsymbol{Q}}_{[\nu]}\right)\right] \,\middle|\, \mathcal{X}_{\mathsf{tot}}\right\}, \end{aligned} \tag{C.27}$$

where $\tilde{\mathcal{Q}} = (\tilde{\boldsymbol{Q}}_{[1]}, \tilde{\boldsymbol{Q}}_{[2]}) \in \mathcal{V}_{[1]} \times \mathcal{V}_{[2]}$. The inverse Laplace transform and change of variables $K\boldsymbol{Q}_{[\nu]} \mapsto \boldsymbol{Q}_{[\nu]}$ yields with some abuse of notation[1]

$$\mu^K(\mathrm{d}\mathcal{Q}) = \left(\frac{K}{2\pi\mathrm{j}}\right)^{\kappa} \lim_{\Gamma\to\infty} \int_{\mathcal{J}_\Gamma^\kappa} \exp\left[-Kc^{K,n}(\mathcal{Q},\tilde{\mathcal{Q}})\right] \tilde{\mu}(\mathrm{d}\tilde{\mathcal{Q}}), \tag{C.28}$$

where $\kappa = \sum_{\nu=1}^{2}[(n+1)\tau_{[\nu]}]^2$, $\mathcal{Q} = (\boldsymbol{Q}_{[1]}, \boldsymbol{Q}_{[2]}) \in \mathcal{V}_{[1]} \times \mathcal{V}_{[2]}$, $\mathcal{J} = (-\mathrm{j}\Gamma,\ \mathrm{j}\Gamma)$ and

$$c^{K,n}(\mathcal{Q},\tilde{\mathcal{Q}}) = \sum_{\nu=1}^{2} \mathrm{tr}(\boldsymbol{Q}_{[\nu]}\tilde{\boldsymbol{Q}}_{[\nu]}) - \frac{1}{K}\sum_{k=1}^{K} \log M_k^n(\tilde{\mathcal{Q}}). \tag{C.29}$$

Using (C.28) we can write (C.20) as

$$\Xi^{K,n} = \int \exp\left[-K\Big(-\alpha^{-1}G^{K,n}(\mathcal{Q})\Big)\right] \mu^K(\mathrm{d}\mathcal{Q}) + O(K^{-1}) \tag{C.30}$$

$$= \int \mathrm{e}^{-K(-\alpha^{-1}G^{K,n}(\mathcal{Q}))} \left[\left(\frac{K}{2\pi j}\right)^{\kappa} \lim_{\Gamma\to\infty} \int_{\mathcal{J}_\Gamma^\kappa} \mathrm{e}^{-Kc^{K,n}(\mathcal{Q},\tilde{\mathcal{Q}})} \tilde{\mu}(\mathrm{d}\tilde{\mathcal{Q}})\right] \mathrm{d}\mathcal{Q}, \tag{C.31}$$

where we dropped the vanishing term in (C.31) and wrote

$$G^{K,n}(\mathcal{Q}) = \sum_{\nu=1}^{2} G_{[\nu]}^{K,n}(\boldsymbol{Q}_{[\nu]}). \tag{C.32}$$

Intuitively, if the exponents in (C.31) converge in the limit $K \to \infty$ as $c^{K,n}(\mathcal{Q},\tilde{\mathcal{Q}}) \to c^n(\mathcal{Q},\tilde{\mathcal{Q}})$ and $G^{K,n}(\mathcal{Q}) \to G^n(\mathcal{Q})$, the integrals are asymptotically dominated by the points in the neighborhood of the (local) minimas of $c^n(\mathcal{Q},\tilde{\mathcal{Q}})$ and $-G^n(\mathcal{Q})$. This is stated in more detail by the saddle point method (or Laplace's method) of integration, derived for the class of real-valued functions with complex arguments in Appendix F.

Now, let $K \to \infty$ and use (F.12) for the integral in the parenthesis on the RHS of (C.31) while the variables connected to $\mathcal{Q}$ are arbitrary and fixed. Then consider

[1]The Laplace transform is defined for functions with real arguments and, thus, for the inverse transform we represent the set of complex Hermitian matrices $\mathcal{Q} = (\boldsymbol{Q}_{[1]}, \boldsymbol{Q}_{[2]})$ by an equivalent set of $\kappa = \sum_{\nu=1}^{2}[(n+1)\tau_{[\nu]}]^2$ independent real variables. Similarly we represent $\tilde{\mathcal{Q}} = (\tilde{\boldsymbol{Q}}_{[1]}, \tilde{\boldsymbol{Q}}_{[2]})$ by $\kappa$ complex variables with fixed real part and let the integral measure $\tilde{\mu}$ be the corresponding $\kappa$ dimensional product measure. Since we are not interested in the exact evaluation of the integrals, we keep the same notation for the variables regardless how they are presented.

$\mathcal{Q}$ as a set of real variables and apply (F.11), that is,

$$\Xi^{K,n} = \int \exp\left\{-K\left[c^n(\mathcal{Q},\tilde{\mathcal{Q}}^{\mathsf{s}}) - \alpha^{-1}G^n(\mathcal{Q})\right]\right\} \times\left[\left(\frac{K}{2\pi}\right)^{\kappa}\left(\frac{(2\pi/K)^{\kappa}}{\det\left(\Re\{\nabla^2_{\Im}\, c^n(\mathcal{Q},\tilde{\mathcal{Q}}^{\mathsf{s}})\}\right)}\right)^{1/2}\right]\mathrm{d}\mathcal{Q} \tag{C.33}$$

$$= \exp\left\{-K\left[c^n(\mathcal{Q}^{\mathsf{s}},\tilde{\mathcal{Q}}^{\mathsf{s}}) - \alpha^{-1}G^n(\mathcal{Q}^{\mathsf{s}})\right]\right\} \times\left\{\det\left(\Re\{\nabla^2_{\Im}\, c^n(\mathcal{Q}^{\mathsf{s}},\tilde{\mathcal{Q}}^{\mathsf{s}})\}\right)\det\left[\nabla^2_{\Re}\left(c^n(\mathcal{Q}^{\mathsf{s}},\tilde{\mathcal{Q}}^{\mathsf{s}}) - \alpha^{-1}G^n(\mathcal{Q}^{\mathsf{s}})\right)\right]\right\}^{-1/2}, \tag{C.34}$$

where $\nabla^2_{\Im}\, c^n(\mathcal{Q}^{\mathsf{s}},\tilde{\mathcal{Q}}^{\mathsf{s}})$ and $\nabla^2_{\Re}\left(c^n(\mathcal{Q}^{\mathsf{s}},\tilde{\mathcal{Q}}^{\mathsf{s}}) - \alpha^{-1}G^n(\mathcal{Q}^{\mathsf{s}})\right)$ are complex and real Hessian matrices, independent of $K$ and defined in Appendix F. From (C.33) we get

$$\tilde{\mathcal{Q}}^{\mathsf{s}} = \inf_{\tilde{\mathcal{Q}}\in\mathcal{V}_{[1]}\times\mathcal{V}_{[2]}} c^n(\mathcal{Q},\tilde{\mathcal{Q}}), \tag{C.35}$$

$$\mathcal{Q}^{\mathsf{s}} = \inf_{\mathcal{Q}\in\mathcal{V}_{[1]}\times\mathcal{V}_{[2]}} \left\{c^n(\mathcal{Q},\tilde{\mathcal{Q}}^{\mathsf{s}}) - G^n(\mathcal{Q})\right\}, \tag{C.36}$$

where $\mathcal{Q}$ is arbitrary and fixed in the first optimization problem and

$$\lim_{K\to\infty}\frac{1}{K}\log\,\Xi^{K,n} = \alpha^{-1}G^n(\mathcal{Q}^{\mathsf{s}}) - c^n(\mathcal{Q}^{\mathsf{s}},\tilde{\mathcal{Q}}^{\mathsf{s}}). \tag{C.37}$$

Note that for $\tilde{\boldsymbol{Q}}^{\mathsf{s}}_{[\nu]}, \tilde{\boldsymbol{Q}}^{\mathsf{s}}_{[\nu]} \in \mathcal{V}_{[\nu]}, \nu = 1,2$, we get $G^n(\mathcal{Q}^{\mathsf{s}}) \in \mathbb{R}$ and $c^n(\mathcal{Q}^{\mathsf{s}},\tilde{\mathcal{Q}}^{\mathsf{s}}) \in \mathbb{R}$ so that (C.37) is real valued, as expected. With the help of [202], the extremas in (C.35) – (C.36) are found to satisfy the coupled equations

$$\boldsymbol{Q}^{\mathsf{s}}_{[\nu]} = \lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\frac{1}{M^n_k(\tilde{\mathcal{Q}})} \times\mathbb{E}\left\{\sum_{m=1}^{M}\boldsymbol{\omega}_{k,[\nu],m}\boldsymbol{\omega}^{\mathsf{H}}_{k,[\nu],m}\exp\left[\sum_{\nu=1}^{2}\sum_{m=1}^{M}\boldsymbol{\omega}^{\mathsf{H}}_{k,[\nu],m}\tilde{\boldsymbol{Q}}^{\mathsf{s}}_{[\nu]}\boldsymbol{\omega}_{k,[\nu],m}\right]\,\middle|\,\mathcal{X}_{\mathsf{tot}}\right\}, \tag{C.38}$$

$$\tilde{\boldsymbol{Q}}^{\mathsf{s}}_{[\nu]} = -\alpha^{-1}\left(\boldsymbol{I}_{(n+1)\tau_{[\nu]}} + \boldsymbol{A}_{[\nu]}\boldsymbol{Q}^{\mathsf{s}}_{[\nu]}\right)^{-1}\boldsymbol{A}_{[\nu]}. \tag{C.39}$$

To make the numerical evaluation of the saddle point equations feasible, we make next a simplifying assumption that limits the space of allowed saddle points drastically.

**Assumption 7 (Replica symmetry).** The saddle point solution to the system of equations (C.38) – (C.39) for $\nu = 1, 2$, is invariant under permutations of the replica indices, that is, we have the $(n+1)\tau_{[\nu]} \times (n+1)\tau_{[\nu]}$ Hermitian matrices

$$\boldsymbol{Q}^{\mathsf{s}}_{[\nu]} = \left[\begin{array}{c|c} \boldsymbol{Q}^{\{0,0\}}_{[\nu]} & \boldsymbol{e}_n^{\mathsf{T}} \otimes \boldsymbol{Q}^{\{0,1\}}_{[\nu]} \\ \hline \boldsymbol{e}_n \otimes (\boldsymbol{Q}^{\{0,1\}}_{[\nu]})^{\mathsf{H}} & \boldsymbol{I}_n \otimes (\boldsymbol{Q}^{\{1,1\}}_{[\nu]} - \boldsymbol{Q}^{\{1,2\}}_{[\nu]}) + \boldsymbol{e}_n\boldsymbol{e}_n^{\mathsf{T}} \otimes \boldsymbol{Q}^{\{1,2\}}_{[\nu]} \end{array}\right], \tag{C.40}$$

$$\tilde{\boldsymbol{Q}}^{\mathsf{s}}_{[\nu]} = \left[\begin{array}{c|c} \tilde{\boldsymbol{Q}}^{\{0,0\}}_{[\nu]} & \boldsymbol{e}_n^{\mathsf{T}} \otimes \tilde{\boldsymbol{Q}}^{\{0,1\}}_{[\nu]} \\ \hline \boldsymbol{e}_n \otimes (\tilde{\boldsymbol{Q}}^{\{0,1\}}_{[\nu]})^{\mathsf{H}} & \boldsymbol{I}_n \otimes (\tilde{\boldsymbol{Q}}^{\{1,1\}}_{[\nu]} - \tilde{\boldsymbol{Q}}^{\{1,2\}}_{[\nu]}) + \boldsymbol{e}_n\boldsymbol{e}_n^{\mathsf{T}} \otimes \tilde{\boldsymbol{Q}}^{\{1,2\}}_{[\nu]} \end{array}\right], \tag{C.41}$$

where $\{\boldsymbol{Q}^{\{0,0\}}_{[\nu]}, \tilde{\boldsymbol{Q}}^{\{0,0\}}_{[\nu]}, \boldsymbol{Q}^{\{1,1\}}_{[\nu]}, \tilde{\boldsymbol{Q}}^{\{1,1\}}_{[\nu]}, \boldsymbol{Q}^{\{1,2\}}_{[\nu]}, \tilde{\boldsymbol{Q}}^{\{1,2\}}_{[\nu]}\}$ are $\tau_{[\nu]} \times \tau_{[\nu]}$ Hermitian matrices. ♢

Under the Assumption 7,

$$\tilde{\boldsymbol{Q}}^{\{0,0\}}_{[\nu]} = -n(\tilde{\boldsymbol{C}}_{[\nu]} + n\boldsymbol{C}_{[\nu]})^{-1} \xrightarrow{n\to 0} \boldsymbol{0}, \tag{C.42}$$

$$\tilde{\boldsymbol{Q}}^{\{0,1\}}_{[\nu]} = (\tilde{\boldsymbol{C}}_{[\nu]} + n\boldsymbol{C}_{[\nu]})^{-1} \xrightarrow{n\to 0} \tilde{\boldsymbol{C}}^{-1}_{[\nu]}, \tag{C.43}$$

$$\tilde{\boldsymbol{Q}}^{\{1,2\}}_{[\nu]} = \tilde{\boldsymbol{C}}^{-1}_{[\nu]}\boldsymbol{C}_{[\nu]}(\tilde{\boldsymbol{C}}_{[\nu]} + n\boldsymbol{C}_{[\nu]})^{-1} \xrightarrow{n\to 0} \tilde{\boldsymbol{C}}^{-1}_{[\nu]}\boldsymbol{C}_{[\nu]}\tilde{\boldsymbol{C}}^{-1}_{[\nu]}, \tag{C.44}$$

$$\tilde{\boldsymbol{Q}}^{\{1,1\}}_{[\nu]} = \tilde{\boldsymbol{C}}^{-1}((1-n)\boldsymbol{C}_{[\nu]} - \tilde{\boldsymbol{C}}_{[\nu]}))(\tilde{\boldsymbol{C}}_{[\nu]} + n\boldsymbol{C}_{[\nu]})^{-1} = \tilde{\boldsymbol{Q}}^{\{1,2\}}_{[\nu]} - \tilde{\boldsymbol{C}}^{-1}, \tag{C.45}$$

where we defined for notational convenience the new matrices

$$\boldsymbol{C}_{[\nu]} = \sigma^2\boldsymbol{I}_{\tau_{[\nu]}} + \alpha\left(\boldsymbol{Q}^{\{0,0\}}_{[\nu]} - (\boldsymbol{Q}^{\{0,1\}}_{[\nu]} + (\boldsymbol{Q}^{\{0,1\}}_{[\nu]})^{\mathsf{H}}) + \boldsymbol{Q}^{\{1,2\}}_{[\nu]}\right), \tag{C.46}$$

$$\tilde{\boldsymbol{C}}_{[\nu]} = \tilde{\sigma}^2\boldsymbol{I}_{\tau_{[\nu]}} + \alpha\left(\boldsymbol{Q}^{\{1,1\}}_{[\nu]} - \boldsymbol{Q}^{\{1,2\}}_{[\nu]}\right). \tag{C.47}$$

With the assumption of replica symmetry, the first term on the RHS of (C.25) cancels and the remaining terms from the log-det yield

$$\begin{aligned} G^n(\mathcal{Q}^{\mathsf{s}}) = &- nT_{\mathsf{coh}}\log\pi \\ &- \sum_{\nu=1}^{2}\left[\log\det(\tilde{\boldsymbol{C}}_{[\nu]} + n\boldsymbol{C}_{[\nu]}) + (n-1)\log\det(\tilde{\boldsymbol{C}}_{[\nu]})\right] \in \mathbb{R}. \end{aligned} \tag{C.48}$$

The replica symmetric form of the trace in (C.29), on the other hand, reads

$$\begin{aligned} \operatorname{tr}(\boldsymbol{Q}^{\mathsf{s}}_{[\nu]}\tilde{\boldsymbol{Q}}^{\mathsf{s}}_{[\nu]}) = &\operatorname{tr}\left(\tilde{\boldsymbol{Q}}^{\{0,0\}}_{[\nu]}\boldsymbol{Q}^{\{0,0\}}_{[\nu]}\right) + n\left[\operatorname{tr}\left(\tilde{\boldsymbol{Q}}^{\{0,1\}}_{[\nu]}(\boldsymbol{Q}^{\{0,1\}}_{[\nu]})^{\mathsf{H}} + \boldsymbol{Q}^{\{0,1\}}_{[\nu]}\tilde{\boldsymbol{Q}}^{\{0,1\}}_{[\nu]}\right)\right] \\ &+ n\operatorname{tr}\left(\tilde{\boldsymbol{Q}}^{\{1,1\}}_{[\nu]}\boldsymbol{Q}^{\{1,1\}}_{[\nu]}\right) + n(n-1)\operatorname{tr}\left(\tilde{\boldsymbol{Q}}^{\{1,2\}}_{[\nu]}\boldsymbol{Q}^{\{1,2\}}_{[\nu]}\right) \in \mathbb{R}. \end{aligned} \tag{C.49}$$

Let us denote

$$\boldsymbol{\mu}_{k,[1],m}^{\{0\}} = \boldsymbol{p}_k h_{k,m}^{\{0\}}, \qquad \boldsymbol{\mu}_{k,[1],m}^{\{a\}} = \boldsymbol{p}_k h_{k,m}^{\{a\}} \tag{C.50}$$

$$\boldsymbol{\mu}_{k,[2],m}^{\{0\}} = \boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta} h_{k,m}^{\{0\}}, \qquad \boldsymbol{\mu}_{k,[2],m}^{\{a\}} = \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}}^{(\ell)} h_{k,m}^{\{a\}} + \Delta \boldsymbol{u}_{k,m}^{\{a\}}, \tag{C.51}$$

so that the replica symmetric moment generating function (C.27) is given by

$$\begin{aligned} M_k^n(\tilde{\mathcal{Q}}^{\mathsf{s}}) = \mathbb{E}\Bigg\{ \prod_{\nu=1}^{2} \prod_{m=1}^{M} \exp \Bigg[ & (\boldsymbol{\mu}_{k,[\nu],m}^{\{0\}})^{\mathsf{H}} \tilde{\boldsymbol{Q}}_{[\nu]}^{\{0,0\}} \boldsymbol{\mu}_{k,[\nu],m}^{\{0\}} \\ & + \sum_{a=1}^{n} (\boldsymbol{\mu}_{k,[\nu],m}^{\{a\}})^{\mathsf{H}} \tilde{\boldsymbol{Q}}_{[\nu]}^{\{1,1\}} \boldsymbol{\mu}_{k,[\nu],m}^{\{a\}} + \sum_{a=1}^{n} 2\Re\left\{ (\boldsymbol{\mu}_{k,[\nu],m}^{\{0\}})^{\mathsf{H}} \tilde{\boldsymbol{Q}}_{[\nu]}^{\{0,1\}} \boldsymbol{\mu}_{k,[\nu],m}^{\{a\}} \right\} \\ & + \sum_{a=1}^{n} \sum_{b\neq a} (\boldsymbol{\mu}_{k,[\nu],m}^{\{a\}})^{\mathsf{H}} \tilde{\boldsymbol{Q}}_{[\nu]}^{\{1,2\}} \boldsymbol{\mu}_{k,[\nu],m}^{\{b\}} \Bigg] \Bigg| \mathcal{X}_{\mathsf{tot}} \Bigg\}. \end{aligned} \tag{C.52}$$

Plugging

$$\tilde{\boldsymbol{Q}}_{[\nu]}^{\{0,1\}} = \boldsymbol{C}_{[\nu]}^{-1} (n\tilde{\boldsymbol{C}}_{[\nu]}^{-1} + \boldsymbol{C}_{[\nu]}^{-1})^{-1} \tilde{\boldsymbol{C}}_{[\nu]}^{-1}, \tag{C.53}$$

$$\tilde{\boldsymbol{Q}}_{[\nu]}^{\{1,2\}} = \tilde{\boldsymbol{C}}_{[\nu]}^{-1} (n\tilde{\boldsymbol{C}}_{[\nu]}^{-1} + \boldsymbol{C}_{[\nu]}^{-1})^{-1} \tilde{\boldsymbol{C}}_{[\nu]}^{-1}, \tag{C.54}$$

to (C.52) gives after some simplifications

$$\begin{aligned} & M_k^n(\tilde{\mathcal{Q}}^{\mathsf{s}}) \\ & = \mathbb{E}\Bigg\{ \prod_{\nu=1}^{2} \prod_{m=1}^{M} \exp \Bigg[ (\boldsymbol{C}_{[\nu]}^{-1} \boldsymbol{\mu}_{k,[\nu],m}^{\{0\}} + \tilde{\boldsymbol{C}}_{[\nu]}^{-1} \sum_{a=1}^{n} \boldsymbol{\mu}_{k,[\nu],m}^{\{a\}})^{\mathsf{H}} \\ & \qquad \times (n\tilde{\boldsymbol{C}}_{[\nu]}^{-1} + \boldsymbol{C}_{[\nu]}^{-1})^{-1} (\boldsymbol{C}_{[\nu]}^{-1} \boldsymbol{\mu}_{k,[\nu],m}^{\{0\}} + \tilde{\boldsymbol{C}}_{[\nu]}^{-1} \sum_{a=1}^{n} \boldsymbol{\mu}_{k,[\nu],m}^{\{a\}}) \Bigg] \\ & \qquad \times \exp \Bigg[ - (\boldsymbol{\mu}_{k,[\nu],m}^{\{0\}})^{\mathsf{H}} \boldsymbol{C}_{[\nu]}^{-1} \boldsymbol{\mu}_{k,[\nu],m}^{\{0\}} - \sum_{a=1}^{n} (\boldsymbol{\mu}_{k,[\nu],m}^{\{a\}})^{\mathsf{H}} \tilde{\boldsymbol{C}}_{[\nu]}^{-1} \boldsymbol{\mu}_{k,[\nu],m}^{\{a\}} \Bigg] \Bigg| \mathcal{X}_{\mathsf{tot}} \Bigg\}, \end{aligned} \tag{C.55}$$

and using (C.22) from right to left on the first exponential term in (C.55) yields

$$\begin{aligned} M_k^n(\tilde{\mathcal{Q}}^{\mathsf{s}}) = (C_{\mathsf{mgf}}^n)^M \mathbb{E}\Bigg\{ \int \prod_{\nu=1}^{2} \prod_{m=1}^{M} & f(\boldsymbol{z}_{k,[\nu],m} | \, \boldsymbol{\mu}_{k,[\nu],m}^{\{0\}} ; \, \boldsymbol{C}_{[\nu]}) \\ & \times \prod_{a=1}^{n} f(\boldsymbol{z}_{k,[\nu],m} | \, \boldsymbol{\mu}_{k,[\nu],m}^{\{a\}} ; \, \tilde{\boldsymbol{C}}_{[\nu]}) \mathrm{d}\boldsymbol{z}_{k,[\nu],m} \Bigg| \mathcal{X}_{\mathsf{tot}} \Bigg\}, \end{aligned} \tag{C.56}$$

where

$$C^n_{\mathsf{mgf}} = \pi^{nT_{\mathsf{coh}}} \prod_{\nu=1}^{2} \det(\tilde{\boldsymbol{C}}_{[\nu]})^{n-1} \det(\tilde{\boldsymbol{C}}_{[\nu]} + n\boldsymbol{C}_{[\nu]}), \tag{C.57}$$

is a normalization factor imposed by the introduction of the complex Gaussian densities

$$f\left(\boldsymbol{z} \mid \boldsymbol{\mu}; \boldsymbol{\Omega}\right) = \frac{\pi^{-\tau}}{\det(\boldsymbol{\Omega})} \exp\left(-(\boldsymbol{z}-\boldsymbol{\mu})^{\mathsf{H}} \boldsymbol{\Omega}^{-1} (\boldsymbol{z}-\boldsymbol{\mu})\right), \quad \boldsymbol{z}, \boldsymbol{\mu} \in \mathbb{C}^{\tau}, \boldsymbol{\Omega} > 0. \tag{C.58}$$

Since the replicas are assumed to be IID, we may write

$$\begin{aligned} M^n_k(\tilde{\mathcal{Q}}^{\mathsf{s}}) = \left(C^n_{\mathsf{mgf}}\right)^M \int \mathbb{E}\left\{ \prod_{\nu=1}^{2} \prod_{m=1}^{M} f\big(\boldsymbol{z}_{k,[\nu],m} \big| \, \boldsymbol{\mu}_{k,[\nu],m} \, ; \, \boldsymbol{C}_{[\nu]}\big) \,\middle|\, \mathcal{X}_{\mathsf{tot}} \right\} \\ \times \left( \mathbb{E}\left\{ \prod_{\nu=1}^{2} \prod_{m=1}^{M} f\big(\boldsymbol{z}_{k,[\nu],m} \mid \tilde{\boldsymbol{\mu}}_{k,[\nu],m} \, ; \, \tilde{\boldsymbol{C}}_{[\nu]}\big) \,\middle|\, \mathcal{X}_{\mathsf{tot}} \right\} \right)^n \prod_{\nu=1}^{2} \prod_{m=1}^{M} \mathrm{d}\boldsymbol{z}_{k,[\nu],m}, \end{aligned} \tag{C.59}$$

where

$$\boldsymbol{\mu}_{k,[1],m} = \boldsymbol{p}_k h_{k,m}, \qquad \tilde{\boldsymbol{\mu}}_{k,[1],m} = \boldsymbol{p}_k \tilde{h}_{k,\vartheta,m} \tag{C.60}$$

$$\boldsymbol{\mu}_{k,[2],m} = \boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta} h_{k,m}, \qquad \tilde{\boldsymbol{\mu}}_{k,[2],m} = \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle^{(\ell)}_{\mathsf{app}} \tilde{h}_{k,\vartheta,m} + \Delta\tilde{\boldsymbol{u}}_{k,m}. \tag{C.61}$$

When $n \to 0$, we get from (C.57) and (C.59) that $M^n_k(\tilde{\mathcal{Q}}^{\mathsf{s}}) \to 1 \ \forall k = 1, \ldots, K$. With some abuse of notation, the replica symmetric saddle point (C.38) becomes thus

$$\begin{aligned} \boldsymbol{Q}^{\mathsf{s}}_{[\nu]} = \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \mathbb{E}\Bigg\{ \int \mathrm{d}\boldsymbol{z}_{k,[\nu],m} \, \boldsymbol{\mu}^n_{k,[\nu],m} (\boldsymbol{\mu}^n_{k,[\nu],m})^{\mathsf{H}} \\ \times \prod_{\nu'=1}^{2} \prod_{m'=1}^{M} f\big(\boldsymbol{z}_{k,[\nu'],m'} \big| \, \boldsymbol{\mu}_{k,[\nu']m'}; \boldsymbol{C}_{[\nu']}\big) \\ \times \frac{f\big(\tilde{\boldsymbol{z}}_{k,[\nu'],m'} = \boldsymbol{z}_{k,[\nu'],m'} \big| \, \tilde{\boldsymbol{\mu}}_{k,[\nu'],m'} \, ; \tilde{\boldsymbol{C}}_{[\nu']}\big)}{\mathbb{E}\{ \prod_{\nu'=1}^{2} \prod_{m'=1}^{M} f\big(\tilde{\boldsymbol{z}}_{k,[\nu'],m'} = \boldsymbol{z}_{k,[\nu'],m'} \big| \, \tilde{\boldsymbol{\mu}}_{k,[\nu'],m'} \, ; \tilde{\boldsymbol{C}}_{[\nu']}\big) \mid \mathcal{X}_{\mathsf{tot}} \}} \,\Bigg|\, \mathcal{X}_{\mathsf{tot}} \Bigg\}, \end{aligned} \tag{C.62}$$

where

$$\boldsymbol{\mu}^n_{k,[\nu],m} = \mathrm{vec}\big( \begin{bmatrix} \boldsymbol{\mu}_{k,[\nu],m} & \boldsymbol{e}_n^{\mathsf{T}} \otimes \tilde{\boldsymbol{\mu}}_{k,[\nu],m} \end{bmatrix} \big) \in \mathbb{C}^{(n+1)\tau_{[\nu]}}. \tag{C.63}$$

Denoting

$$\langle \cdots \rangle_{k,m} = \frac{\mathbb{E}_{\tilde{h}_{k,\vartheta,m},\Delta\tilde{\boldsymbol{u}}_{k,m}} \left\{ \cdots \prod_{\nu=1}^{2} f\left(\boldsymbol{z}_{k,[\nu],m} \,|\, \tilde{\boldsymbol{\mu}}_{k,[\nu],m}\,; \tilde{\boldsymbol{C}}_{[\nu]}\right) \right\}}{\mathbb{E}_{\tilde{h}_{k,\vartheta,m},\Delta\tilde{\boldsymbol{u}}_{k,m}} \left\{ \prod_{\nu=1}^{2} f\left(\boldsymbol{z}_{k,[\nu],m} \,|\, \tilde{\boldsymbol{\mu}}_{k,[\nu],m}\,; \tilde{\boldsymbol{C}}_{[\nu]}\right) \right\}}, \tag{C.64}$$

for a single-user GPME indexed by $k$ and $m$, the elements of (C.40) are given by

$$\boldsymbol{Q}_{[\nu]}^{\{0,0\}} = \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \mathbb{E}\left\{ \boldsymbol{\mu}_{k,[\nu],m} \boldsymbol{\mu}_{k,[\nu],m}^{\mathsf{H}} \,|\, \mathcal{X}_{\mathsf{tot}} \right\}, \tag{C.65}$$

$$\boldsymbol{Q}_{[\nu]}^{\{0,1\}} = \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \mathbb{E}\left\{ \boldsymbol{\mu}_{k,[\nu],m} \langle \tilde{\boldsymbol{\mu}}_{k,[\nu],m} \rangle_{k,m}^{\mathsf{H}} \,|\, \mathcal{X}_{\mathsf{tot}} \right\}, \tag{C.66}$$

$$\boldsymbol{Q}_{[\nu]}^{\{1,1\}} = \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \mathbb{E}\left\{ \langle \tilde{\boldsymbol{\mu}}_{k,[\nu],m} \tilde{\boldsymbol{\mu}}_{k,[\nu],m}^{\mathsf{H}} \rangle_{k,m} \,|\, \mathcal{X}_{\mathsf{tot}} \right\}, \tag{C.67}$$

$$\boldsymbol{Q}_{[\nu]}^{\{1,2\}} = \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \mathbb{E}\left\{ \langle \tilde{\boldsymbol{\mu}}_{k,[\nu],m} \rangle_{k,m} \langle \tilde{\boldsymbol{\mu}}_{k,[\nu],m} \rangle_{k,m}^{\mathsf{H}} \,|\, \mathcal{X}_{\mathsf{tot}} \right\}. \tag{C.68}$$

Using (C.65) yields the following interpretation for the parameters $\boldsymbol{C}_{[\nu]}$ and $\tilde{\boldsymbol{C}}_{[\nu]}$

$$\begin{aligned} \boldsymbol{C}_{[\nu]} = \sigma^2 \boldsymbol{I}_{\tau_{[\nu]}} + \alpha \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \mathbb{E}\Big\{ & \left( \boldsymbol{\mu}_{k,[\nu],m} - \langle \tilde{\boldsymbol{\mu}}_{k,[\nu],m} \rangle_{k,m} \right) \\ & \times \left( \boldsymbol{\mu}_{k,[\nu],m} - \langle \tilde{\boldsymbol{\mu}}_{k,[\nu],m} \rangle_{k,m} \right)^{\mathsf{H}} \,|\, \mathcal{X}_{\mathsf{tot}} \Big\}, \end{aligned} \tag{C.69}$$

$$\begin{aligned} \tilde{\boldsymbol{C}}_{[\nu]} = \tilde{\sigma}^2 \boldsymbol{I}_{\tau_{[\nu]}} + \alpha \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \mathbb{E}\Big\{ & \left( \tilde{\boldsymbol{\mu}}_{k,[\nu],m} - \langle \tilde{\boldsymbol{\mu}}_{k,[\nu],m} \rangle_{k,m} \right) \\ & \left( \tilde{\boldsymbol{\mu}}_{k,[\nu],m} - \langle \tilde{\boldsymbol{\mu}}_{k,[\nu],m} \rangle_{k,m} \right)^{\mathsf{H}} \,|\, \mathcal{X}_{\mathsf{tot}} \Big\}. \end{aligned} \tag{C.70}$$

In order to evaluate the free energy (C.11) under the RS ansatz, we need to calculate

$$\begin{aligned} F_{\mathsf{rm-rs}} &= -\lim_{n\to 0} \frac{\partial}{\partial n} \left\{ \alpha^{-1} G^n(\mathcal{Q}^{\mathsf{s}}) - c^n(\mathcal{Q}^{\mathsf{s}}, \tilde{\mathcal{Q}}^{\mathsf{s}}) \right\} \\ &= \lim_{n\to 0} \frac{\partial}{\partial n} \left\{ \sum_{\nu=1}^{2} \mathrm{tr}(\boldsymbol{Q}_{[\nu]} \tilde{\boldsymbol{Q}}_{[\nu]}) - \frac{1}{\alpha} G^n(\mathcal{Q}^{\mathsf{s}}) - \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \log M_k^n(\tilde{\mathcal{Q}}) \right\}. \end{aligned} \tag{C.71}$$

After some algebra, we find[2]

$$\lim_{n\to 0}\frac{\partial}{\partial n}G^n(\mathcal{Q}^{\mathsf{s}}) = -\sum_{\nu=1}^{2}\left[\tau_{[\nu]}\log\pi + \operatorname{tr}\left(\tilde{\boldsymbol{C}}_{[\nu]}^{-1}\boldsymbol{C}_{[\nu]}\right) + \log\det(\tilde{\boldsymbol{C}}_{[\nu]})\right], \quad \text{(C.72)}$$

and

$$\begin{aligned}
&\lim_{n\to 0}\frac{\partial}{\partial n}\operatorname{tr}(\boldsymbol{Q}_{[\nu]}^{\mathsf{s}}\tilde{\boldsymbol{Q}}_{[\nu]}^{\mathsf{s}}) \\
&\quad= \operatorname{tr}\Big\{-\tilde{\boldsymbol{C}}_{[\nu]}^{-1}\left[\boldsymbol{Q}_{[\nu]}^{\{0,0\}} - (\boldsymbol{Q}_{[\nu]}^{\{0,1\}} + (\boldsymbol{Q}_{[\nu]}^{\{0,1\}})^{\mathsf{H}}) + \boldsymbol{Q}_{[\nu]}^{\{1,1\}}\right] \\
&\qquad\quad +\tilde{\boldsymbol{C}}_{[\nu]}^{-1}\boldsymbol{C}_{[\nu]}\tilde{\boldsymbol{C}}_{[\nu]}^{-1}\left(\boldsymbol{Q}_{[\nu]}^{\{1,1\}} - \boldsymbol{Q}_{[\nu]}^{\{1,2\}}\right)\Big\} \qquad \text{(C.73)}\\
&\quad= -\alpha^{-1}\operatorname{tr}\left[\boldsymbol{I}_{\tau_{[\nu]}} + \tilde{\sigma}^2\tilde{\boldsymbol{C}}_{[\nu]}^{-1}\boldsymbol{C}_{[\nu]}\tilde{\boldsymbol{C}}_{[\nu]}^{-1} - (\tilde{\sigma}^2+\sigma^2)\tilde{\boldsymbol{C}}_{[\nu]}^{-1}\right], \qquad \text{(C.74)}
\end{aligned}$$

where (C.74) follows by using (C.46) – (C.47) and simplifying. Similar calculations yield

$$\lim_{n\to 0}\frac{\partial}{\partial n}\,\log C_{\mathsf{mgf}}^n = MT_{\mathsf{coh}}\log\pi + M\sum_{\nu=1}^{2}\left[\operatorname{tr}\left(\tilde{\boldsymbol{C}}_{[\nu]}^{-1}\boldsymbol{C}_{[\nu]}\right) + \log\det(\tilde{\boldsymbol{C}}_{[\nu]})\right]. \quad \text{(C.75)}$$

If we exchange the limits in (C.71), the RS free energy finally reads

$$\begin{aligned}
F_{\mathsf{rm-rs}} = &-\lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}\Bigg\{\int\prod_{\nu=1}^{2}\prod_{m=1}^{M} f(\boldsymbol{z}_{k,[\nu],m}|\,\boldsymbol{\mu}_{k,[\nu]m};\boldsymbol{C}_{[\nu]}) \\
&\times\log\frac{f(\tilde{\boldsymbol{z}}_{k,[\nu],m}=\boldsymbol{z}_{k,[\nu],m}|\,\tilde{\boldsymbol{\mu}}_{k,[\nu],m}=\boldsymbol{\mu}_{k,[\nu]m};\tilde{\boldsymbol{C}}_{[\nu]})}{\mathbb{E}\{f(\tilde{\boldsymbol{z}}_{k,[\nu],m}=\boldsymbol{z}_{k,[\nu],m}|\,\tilde{\boldsymbol{\mu}}_{k,[\nu],m};\tilde{\boldsymbol{C}}_{[\nu]})\mid\mathcal{X}_{\mathsf{tot}}\}}\mathrm{d}\boldsymbol{z}_{k,[\nu],m}\,\Bigg|\,\mathcal{X}_{\mathsf{tot}}\Bigg\} \\
&+\left(\frac{1-\alpha M}{\alpha}\right)\sum_{\nu=1}^{2}\left[\tau_{[\nu]}\log\pi + \operatorname{tr}\left(\tilde{\boldsymbol{C}}_{[\nu]}^{-1}\boldsymbol{C}_{[\nu]}\right) + \log\det(\tilde{\boldsymbol{C}}_{[\nu]})\right] \\
&-\frac{1}{\alpha}\sum_{\nu=1}^{2}\operatorname{tr}\left[\boldsymbol{I}_{\tau_{[\nu]}} + \tilde{\sigma}^2\tilde{\boldsymbol{C}}_{[\nu]}^{-1}\boldsymbol{C}_{[\nu]}\tilde{\boldsymbol{C}}_{[\nu]}^{-1} - (\tilde{\sigma}^2+\sigma^2)\tilde{\boldsymbol{C}}_{[\nu]}^{-1}\right]. \qquad \text{(C.76)}
\end{aligned}$$

Comparing this to, e.g., [73, (229)] and [85, (180)] suggests that the decoupling result stated in Claim 4 is indeed correct.

## C.2 Sketch of a Derivation of the Joint Moments

For simplicity of notation, the channel coefficients are treated as real random variables in this section. Extension to the case of complex channel vectors is made along

[2]The following differentials are useful (see, e.g. [182, 186, 202]): $\partial(\log\det(\boldsymbol{X})) = \operatorname{tr}(\boldsymbol{X}^{-1}\partial(\boldsymbol{X}))$, $\partial(\operatorname{tr}(\boldsymbol{X})) = \operatorname{tr}(\partial(\boldsymbol{X}))$ and $\partial(\boldsymbol{X}^{-1}) = -\boldsymbol{X}^{-1}\partial(\boldsymbol{X})\boldsymbol{X}^{-1}$.

the lines of [92]. The calculations follow closely the ones taken in [85, 89, 92], with the modifications introduced in (C.2). Therefore, only a brief sketch of proof is given for completeness.

Let us start by defining a function

$$g(\mathcal{H}_k) = \prod_{m=1}^{M} \left(h_{k,m}^{\{0\}}\right)^{i_m} \left(h_{k,m}^{\{b_m\}}\right)^{l_m} \prod_{a_m=1}^{j_m} \left(h_{k,m}^{\{a_m\}}\right), \tag{C.77}$$

where $b_m > j_m \in \{1, 2, \ldots, n\}$ $\forall m = 1, \ldots, M$, and

$$\mathcal{H}_k = \{h_{k,m}^{\{a\}} \mid m = 1, \ldots, M, a = 0, 1, \ldots, n\}. \tag{C.78}$$

Fix the time index $\vartheta \in \mathcal{D}$ and denote the received signals in (4.16) – (4.17) by $\mathcal{Y}_{\setminus\vartheta} = \{\boldsymbol{y}_{\mathcal{T}}, \boldsymbol{y}_{\mathcal{D}\setminus\vartheta}\}$. Let the corresponding postulated received signals (4.19) – (4.20) be $\tilde{\mathcal{Y}}_{\setminus\vartheta} = \{\tilde{\boldsymbol{y}}_{\mathcal{T}}, \tilde{\boldsymbol{y}}_{\mathcal{D}\setminus\vartheta}\}$, and denote the set (4.3) for some fixed iteration and block indices $\ell$ and $c$ by $\mathcal{I}_\vartheta$. The postulated conditional distribution of the $a$th channel replica reads then

$$\mathbb{Q}(\tilde{\mathcal{Y}}_{\setminus\vartheta} \mid \mathcal{I}_\vartheta, \mathcal{H}^{\{a\}}, \Delta\mathcal{U}^{\{a\}}) = \mathbb{Q}(\tilde{\boldsymbol{y}}_{\mathcal{T}} \mid \mathcal{I}_\vartheta, \mathcal{H}^{\{a\}}) \mathbb{Q}(\tilde{\boldsymbol{y}}_{\mathcal{D}\setminus\vartheta} \mid \mathcal{I}_\vartheta, \mathcal{H}^{\{a\}}, \Delta\mathcal{U}^{\{a\}}). \tag{C.79}$$

Let $\{\mathcal{K}_u\}_{u=1}^{U}$ be a partition of $\mathcal{K}$ so that $K_u = |\mathcal{K}_u|$ is the number of users belonging to the $u$th group, as stated in Assumption 5. We also write $\beta_u = K_u/K \in (0, 1)$ so that $\sum_u \beta_u = 1$. For further development, a free energy like quantity related to (C.11) for the users in the $u$th group is defined as

$$\tilde{f} = \lim_{K=\alpha L=K_u/\beta_u\to\infty} \frac{1}{K_u} \log \tilde{\Xi}^{K_u,n}, \tag{C.80}$$

where

$$\tilde{\Xi}^{K_u,n}(r) = \mathbb{E}\left\{ \exp\left( r \sum_{\xi\in\mathcal{K}_u} g(\mathcal{H}_\xi) \right) \times \prod_{a=1}^{n} \mathbb{Q}(\tilde{\mathcal{Y}}_{\setminus\vartheta} = \mathcal{Y}_{\setminus\vartheta} \mid \mathcal{I}_\vartheta, \mathcal{H}^{\{a\}}, \Delta\mathcal{U}^{\{a\}}) \,\middle|\, \mathcal{X}_{\mathsf{tot}} \right\}, \tag{C.81}$$

and $r$ is a real variable. Note that

$$\lim_{r\to 0} \tilde{\Xi}^{K_u,n}(r) = \Xi^{K,n}, \tag{C.82}$$

where $\Xi^{K,n}$ is given in (C.12).

We first consider obtaining the moments of the channel inputs and the estimator outputs for the multiuser system described in Section 4.1.3. A bit of calculus gives

$$\begin{aligned}
&\frac{\partial}{\partial r}\tilde{f}\bigg|_{r=0} \\
&= \lim_{\substack{K=\alpha L\to\infty \\ K=K_u/\beta_u\to\infty}} \frac{1}{K_u}\sum_{\xi\in\mathcal{K}_u}\mathbb{E}\left\{ g(\mathcal{H}_\xi)\prod_{a=1}^{n}\mathbb{Q}(\tilde{\mathcal{Y}}_{\setminus\vartheta}=\mathcal{Y}_{\setminus\vartheta}\mid\mathcal{I}_\vartheta,\mathcal{H}^{\{a\}},\Delta\mathcal{U}^{\{a\}})\,\middle|\,\mathcal{X}_{\mathsf{tot}}\right\} \\
&= \lim_{\substack{K=\alpha L\to\infty \\ K=K_u/\beta_u\to\infty}} \frac{1}{K_u}\sum_{\xi\in\mathcal{K}_u}\mathbb{E}\left\{ \prod_{m=1}^{M}\left(h_{\xi,m}^{\{0\}}\right)^{i_m}\left(h_{\xi,m}^{\{b_m\}}\right)^{l_m}\right. \\
&\qquad\times\left[\prod_{a_m=1}^{j_m}\prod_{a=1}^{n}\left(h_{\xi,m}^{\{a_m\}}\right)\mathbb{Q}(\mathcal{H}^{\{a\}}\mid\tilde{\boldsymbol{y}}_{\mathcal{T}},\mathcal{I}_\vartheta)\mathbb{Q}(\mathcal{H}^{\{a\}}\mid\tilde{\boldsymbol{y}}_{\mathcal{D}\setminus\vartheta},\mathcal{I}_\vartheta,\Delta\mathcal{U}^{\{a\}})\right] \\
&\qquad\left.\times\left[\prod_{a=1}^{n}\frac{\mathbb{E}_{\mathcal{H}^{\{a\}}}\{\mathbb{Q}(\tilde{\mathcal{Y}}_{\setminus\vartheta}=\mathcal{Y}_{\setminus\vartheta}\mid\mathcal{I}_\vartheta,\mathcal{H}^{\{a\}},\Delta\mathcal{U}^{\{a\}})\}}{\mathbb{Q}(\mathcal{H}^{\{a\}})}\right]\,\middle|\,\mathcal{X}_{\mathsf{tot}}\right\}.
\end{aligned} \tag{C.83}$$

Recalling the definition of the multiuser GPME (4.9) and (4.26), using the fact that the users within the $u$th group have IID channels with uncorrelated multipaths, and taking the limit $n\to 0$ in (C.83) yields

$$\lim_{n\to 0}\left[\frac{\partial}{\partial r}\tilde{f}\bigg|_{r=0}\right]=\prod_{m=1}^{M}\mathbb{E}\left\{ h_{\xi,m}^{i_m}\tilde{h}_{\xi,\vartheta,m}^{l_m}\langle\tilde{h}_{\xi,\vartheta,m}\rangle^{j_m}\,\middle|\,\mathcal{X}_{\mathsf{tot}}\right\},\qquad \xi\in\mathcal{K}_u, \tag{C.84}$$

where $\langle\tilde{h}_{\xi,\vartheta,m}\rangle$ are the posterior mean estimates given by (4.9) and (4.26) for a user $\xi$ in the $u$th group. Hence, the moments on the RHS of (C.84) can be expressed in terms of (C.80) by first differentiating with respect to $r$ and then taking the limits $r\to 0$ and $n\to 0$, in this order.

We now turn to deriving the joint moments of the decoupled channel given in Section 4.2.1. Recall the definition of $G^{K,n}(\mathcal{Q})$ from (C.25) and (C.32), where $\mathcal{Q}=(\boldsymbol{Q}_{[1]},\boldsymbol{Q}_{[2]})$ and $\boldsymbol{Q}_{[\nu]},\nu=1,2$, are defined in (C.19). Let the probability measure akin to (C.26), but modified for (C.81), be given by

$$\begin{aligned}
\tilde{\mu}^K(\mathcal{V};\,r)=\mathbb{E}\left\{\exp\left(r\sum_{\xi\in\mathcal{K}_u}g(\mathcal{H}_\xi)\right)\right. \\
\left.\times\prod_{\nu=1}^{2}\mathbb{1}_{\boldsymbol{V}_{[\nu]}}\left[K\boldsymbol{Q}_{[\nu]}=\sum_{k=1}^{K}\sum_{m=1}^{M}\boldsymbol{\omega}_{k,[\nu],m}\boldsymbol{\omega}_{k,[\nu],m}^{\mathsf{H}}\right]\,\middle|\,\mathcal{X}_{\mathsf{tot}}\right\},
\end{aligned} \tag{C.85}$$

where $\mathcal{V} = (\boldsymbol{V}_{[1]}, \boldsymbol{V}_{[2]}) \subset \mathcal{V}_{[1]} \times \mathcal{V}_{[2]}$. The moment generating function induced by (C.85) reads thus

$$\tilde{M}^{K,n}(\tilde{\mathcal{Q}};\, r) = \prod_{\xi \in \mathcal{K}_u} \tilde{M}^n_\xi(\tilde{\mathcal{Q}};\, r) \prod_{j \in \mathcal{K} \setminus \mathcal{K}_u} M^n_j(\tilde{\mathcal{Q}}), \tag{C.86}$$

where $\tilde{\mathcal{Q}} = (\tilde{\boldsymbol{Q}}_{[1]}, \tilde{\boldsymbol{Q}}_{[2]}) \in \mathcal{V}_{[1]} \times \mathcal{V}_{[2]}$, $M^n_j(\tilde{\mathcal{Q}})$ are given in (C.27) and

$$\tilde{M}^n_\xi(\tilde{\mathcal{Q}};\, r) = \mathbb{E}\left\{ \exp\left[ r g(\mathcal{H}_\xi) \right] \exp\left[ \sum_{\nu=1}^{2} \sum_{m=1}^{M} \operatorname{tr}\left( \boldsymbol{\omega}_{\xi,[\nu],m} \boldsymbol{\omega}^{\mathsf{H}}_{\xi,[\nu],m} \tilde{\boldsymbol{Q}}_{[\nu]} \right) \right] \,\middle|\, \mathcal{X}_{\mathsf{tot}} \right\}. \tag{C.87}$$

Furthermore, define function

$$\begin{aligned} &\tilde{c}^{K,n}(\mathcal{Q}, \tilde{\mathcal{Q}};\, r) \\ &\quad = \sum_{\nu=1}^{2} \operatorname{tr}(\boldsymbol{Q}_{[\nu]} \tilde{\boldsymbol{Q}}_{[\nu]}) - \frac{1}{K} \left[ \sum_{\xi \in \mathcal{K}_u} \log \tilde{M}^n_\xi(\tilde{\mathcal{Q}};\, r) + \sum_{j \in \mathcal{K} \setminus \mathcal{K}_u} \log M^n_j(\tilde{\mathcal{Q}}) \right], \end{aligned} \tag{C.88}$$

related to (C.29) and let $\tilde{c}^{K,n}(\mathcal{Q}, \tilde{\mathcal{Q}};\, r) \to \tilde{c}^n(\mathcal{Q}, \tilde{\mathcal{Q}};\, r)$ and $G^{K,n}(\mathcal{Q}) \to G^n(\mathcal{Q})$, in the limit $K \to \infty$. Then,

$$\tilde{f} = \lim_{K \to \infty} \frac{1}{K_u} \log \tilde{\Xi}^{K_u,n} = \frac{1}{\beta_u} \left[ \alpha^{-1} G^n(\mathcal{Q}^{\mathsf{s}}) - \tilde{c}^n(\mathcal{Q}^{\mathsf{s}}, \tilde{\mathcal{Q}}^{\mathsf{s}};\, r) \right], \tag{C.89}$$

where

$$\tilde{\mathcal{Q}}^{\mathsf{s}} = \inf_{\tilde{\mathcal{Q}} \in \mathcal{V}_{[1]} \times \mathcal{V}_{[2]}} \tilde{c}^n(\mathcal{Q}, \tilde{\mathcal{Q}};\, r), \tag{C.90}$$

$$\mathcal{Q}^{\mathsf{s}} = \inf_{\mathcal{Q} \in \mathcal{V}_{[1]} \times \mathcal{V}_{[2]}} \left\{ \tilde{c}^n(\mathcal{Q}, \tilde{\mathcal{Q}}^{\mathsf{s}};\, r) - G^n(\mathcal{Q}) \right\}, \tag{C.91}$$

coincide with the solutions of (C.35) and (C.36), respectively, when $r \to 0$ by (C.82). Since $G^n(\mathcal{Q})$ does not depend on $r$, we have the partial derivative

$$\begin{aligned} \frac{\partial}{\partial r} \tilde{f} \bigg|_{r=0} &= -\frac{1}{\beta_u} \left[ \frac{\partial}{\partial r} \tilde{c}^n(\mathcal{Q}^{\mathsf{s}}, \tilde{\mathcal{Q}}^{\mathsf{s}};\, r) \bigg|_{r=0} \right] \\ &= \frac{1}{\beta_u K} \sum_{\xi \in \mathcal{K}_u} \left[ \frac{1}{\tilde{M}^n_\xi(\tilde{\mathcal{Q}}^{\mathsf{s}};\, r)} \frac{\partial}{\partial r} \tilde{M}^n_\xi(\tilde{\mathcal{Q}}^{\mathsf{s}};\, r) \bigg|_{r=0} \right], \end{aligned} \tag{C.92}$$

where we omitted the terms in (C.88) that do not depend on $r$ in the second equality.

Recalling the definition of $g(\mathcal{H}_k)$ in (C.77), taking the partial derivative with respect to $r$ in in (C.87) and finally the limit $r \to 0$ yields

$$\frac{\partial}{\partial r}\tilde{M}_{\xi}^{n}(\tilde{\mathcal{Q}};\, r)\bigg|_{r=0} = \mathbb{E}\left\{ \prod_{m=1}^{M} \left(h_{k,m}^{\{0\}}\right)^{i_m} \left(h_{k,m}^{\{b_m\}}\right)^{l_m} \prod_{a_m=1}^{j_m} \left(h_{k,m}^{\{a_m\}}\right) \times \exp\left[ \sum_{\nu=1}^{2} \sum_{m=1}^{M} \operatorname{tr}\big(\boldsymbol{\omega}_{\xi,[\nu],m} \boldsymbol{\omega}_{\xi,[\nu],m}^{\mathsf{H}} \tilde{\boldsymbol{Q}}_{[\nu]}\big) \right] \,\middle|\, \mathcal{X}_{\mathsf{tot}} \right\}. \tag{C.93}$$

By the same techniques as used in the derivations of (C.59) and (C.62), and using the assumption that the users in $\mathcal{K}_u$ are IID with uncorrelated multipaths, we get

$$\begin{aligned} &\lim_{n\to 0}\left[\frac{\partial}{\partial r}\tilde{M}_{\xi}^{n}(\tilde{\mathcal{Q}};\, r)\bigg|_{r=0}\right] \\ &= \int \prod_{m=1}^{M}\prod_{\nu=1}^{2} \mathrm{d}\boldsymbol{z}_{\xi,[\nu],m} \mathbb{E}\left\{ \left[ h_{\xi,m}^{i_m} \prod_{\nu=1}^{2} f\big(\boldsymbol{z}_{\xi,[\nu],m} \big|\, \boldsymbol{\mu}_{\xi,[\nu],m}\,;\, \boldsymbol{C}_{[\nu]}\big)\right] \right. \\ &\quad \times \left[ \tilde{h}_{\xi,m}^{l_m} \frac{\prod_{\nu=1}^{2} f\big(\tilde{\boldsymbol{z}}_{\xi,[\nu],m} = \boldsymbol{z}_{\xi,[\nu],m} \big|\, \tilde{\boldsymbol{\mu}}_{\xi,[\nu],m}\,;\, \tilde{\boldsymbol{C}}_{[\nu]}\big)}{\mathbb{E}\big\{ \prod_{\nu=1}^{2} f\big(\tilde{\boldsymbol{z}}_{\xi,[\nu],m} = \boldsymbol{z}_{\xi,[\nu],m} \big|\, \tilde{\boldsymbol{\mu}}_{\xi,[\nu],m}\,;\, \tilde{\boldsymbol{C}}_{[\nu]}\big) \mid \mathcal{X}_{\mathsf{tot}}\big\}} \right] \\ &\quad \left. \times \left[ \prod_{a_m=1}^{j_m} \frac{\tilde{h}_{\xi,\vartheta,m} \prod_{\nu=1}^{2} f\big(\tilde{\boldsymbol{z}}_{\xi,[\nu],m} = \boldsymbol{z}_{\xi,[\nu],m} \big|\, \tilde{\boldsymbol{\mu}}_{\xi,[\nu],m}\,;\, \tilde{\boldsymbol{C}}_{[\nu]}\big)}{\mathbb{E}\big\{ \prod_{\nu=1}^{2} f\big(\tilde{\boldsymbol{z}}_{\xi,[\nu],m} = \boldsymbol{z}_{\xi,[\nu],m} \big|\, \tilde{\boldsymbol{\mu}}_{\xi,[\nu],m}\,;\, \tilde{\boldsymbol{C}}_{[\nu]}\big) \mid \mathcal{X}_{\mathsf{tot}}\big\}} \right] \,\middle|\, \mathcal{X}_{\mathsf{tot}} \right\}. \end{aligned} \tag{C.94}$$

From the definition for the single-user GPME (C.64) and the fact that $a_m$ is a dummy variable, we finally get

$$\lim_{n\to 0}\left[\frac{\partial}{\partial r}\tilde{f}\,\bigg|_{r=0}\right] = \prod_{m=1}^{M} \mathbb{E}\left\{ h_{\xi,m}^{i_m} \tilde{h}_{\xi,\vartheta,m}^{l_m} \langle \tilde{h}_{\xi,\vartheta,m}\rangle_{\xi,m}^{j_m} \,\middle|\, \mathcal{X}_{\mathsf{tot}}\right\}, \qquad \xi \in \mathcal{K}_u, \tag{C.95}$$

which coincides with (C.84).

From above we may conclude that the joint moments for the user $\xi \in \mathcal{K}_u$, given by the multiuser and the single-user characterizations (C.84) and (C.95), respectively, coincide, that is,

$$\prod_{m=1}^{M} \mathbb{E}\left\{ h_{\xi,m}^{i_m} \tilde{h}_{\xi,\vartheta,m}^{l_m} \langle \tilde{h}_{\xi,\vartheta,m}\rangle^{j_m} \,\middle|\, \mathcal{X}_{\mathsf{tot}}\right\} = \prod_{m=1}^{M} \mathbb{E}\left\{ h_{\xi,m}^{i_m} \tilde{h}_{\xi,\vartheta,m}^{l_m} \langle \tilde{h}_{\xi,\vartheta,m}\rangle_{\xi,m}^{j_m} \,\middle|\, \mathcal{X}_{\mathsf{tot}}\right\}, \tag{C.96}$$

where $\langle \cdots \rangle$ and $\langle \cdots \rangle_{\xi,m}$ are the multiuser and single-user GPMEs defined in Section 4.1.3 and Section 4.2.1, respectively. Assuming that the joint distributions are fully determined by their moments (see [89, Section 1.4.3]), we get the Claim 4.

# *Appendix D*

# *Proof of Propositions 8 and 9*

Let us assume that the replica assumptions are valid so that the Claim 4 holds. Consider iteration $\ell = 1, 2, \ldots$ and time instant $\vartheta \in \mathcal{D}$. In the following we drop the iteration index and write $\langle \cdots \rangle = \langle \cdots \rangle_{k,m}^{(\ell)}$ for notational convenience. The proper indexing should be always clear from the context.

Solving the Gaussian integrals in (4.63) by using (C.22) and the identity

$$\int \boldsymbol{y} e^{-\boldsymbol{y}^{\mathsf{H}}\boldsymbol{A}\boldsymbol{y}+2\Re\{\boldsymbol{b}^{\mathsf{H}}\boldsymbol{y}\}} \mathrm{d}\boldsymbol{y} = \frac{\pi^{\tau}}{\det(\boldsymbol{A})} \boldsymbol{A}^{-1}\boldsymbol{b} e^{\boldsymbol{b}^{\mathsf{H}}\boldsymbol{A}^{-1}\boldsymbol{b}}, \quad \boldsymbol{y} \in \mathbb{C}^{\tau}, \boldsymbol{A} > 0, \tag{D.1}$$

gives the channel estimate

$$\langle \tilde{h}_{k,\vartheta,m} \rangle = \boldsymbol{m}_{k,\mathcal{T},m}^{\mathsf{H}} \boldsymbol{z}_{k,\mathcal{T},m} + \boldsymbol{m}_{k,\mathcal{D}\setminus\vartheta,m}^{\mathsf{H}} \boldsymbol{z}_{k,\mathcal{D}\setminus\vartheta,m}, \tag{D.2}$$

where

$$\boldsymbol{m}_{k,\mathcal{T},m}^{\mathsf{H}} = \frac{\bar{t}_{k,m}}{\tilde{\Gamma}_{k,\vartheta,m}} \boldsymbol{p}_k^{\mathsf{H}} \tilde{\boldsymbol{C}}_{\mathcal{T}}^{-1} \tag{D.3}$$

$$\boldsymbol{m}_{k,\mathcal{D}\setminus\vartheta,m}^{\mathsf{H}} = \boldsymbol{m}_{k,\mathcal{D}\setminus\vartheta,m}^{\mathsf{H}} = \frac{\bar{t}_{k,m}}{\tilde{\Gamma}_{k,\vartheta,m}} \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}}^{\mathsf{H}} \big(\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{u}_{k,m}} + \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}\big)^{-1}, \tag{D.4}$$

and we denoted

$$\tilde{\Gamma}_{k,\vartheta,m} = 1 + \bar{t}_{k,m} \big[\boldsymbol{p}_k^{\mathsf{H}} \tilde{\boldsymbol{C}}_{\mathcal{T}}^{-1} \boldsymbol{p}_k + \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}}^{\mathsf{H}} \big(\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{u}_{k,m}} + \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}\big)^{-1} \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}} \big]. \tag{D.5}$$

Let us also define for the following

$$\underline{\boldsymbol{x}}_{k,\vartheta} = \mathrm{vec}\big(\begin{bmatrix}\boldsymbol{p}_k & \boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta}\end{bmatrix}\big), \tag{D.6}$$

$$\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta} = \mathrm{vec}\big(\begin{bmatrix}\boldsymbol{p}_k & \langle\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}\rangle_{\mathsf{app}}\end{bmatrix}\big), \tag{D.7}$$

$$\underline{\Delta\tilde{\boldsymbol{u}}}_{k,m} = \mathrm{vec}\big(\begin{bmatrix}\boldsymbol{0}_{\tau_{\mathsf{tr}}\times 1} & \Delta\tilde{\boldsymbol{u}}_{k,m}\end{bmatrix}\big), \tag{D.8}$$

$$\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}} = \mathrm{diag}\big(\boldsymbol{0}_{\tau_{\mathsf{tr}}\times\tau_{\mathsf{tr}}}, \tilde{\boldsymbol{\Omega}}_{\Delta\boldsymbol{u}_{k,m}}\big), \tag{D.9}$$

$$\boldsymbol{C} = \mathrm{diag}\big(\boldsymbol{C}_{\mathcal{T}}, \boldsymbol{C}_{\mathcal{D}\setminus\vartheta}\big), \tag{D.10}$$

$$\tilde{\boldsymbol{C}} = \mathrm{diag}\big(\tilde{\boldsymbol{C}}_{\mathcal{T}}, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}\big), \tag{D.11}$$

$$\underline{\boldsymbol{z}}_{k,\vartheta,m} = \mathrm{vec}\big(\begin{bmatrix}\boldsymbol{z}_{k,\mathcal{T},m} & \boldsymbol{z}_{k,\mathcal{D}\setminus\vartheta,m}\end{bmatrix}\big), \tag{D.12}$$

$$\underline{\boldsymbol{m}}^{\mathsf{H}}_{k,\vartheta,m} = \begin{bmatrix}\boldsymbol{m}^{\mathsf{H}}_{k,\mathcal{T},m} & \boldsymbol{m}^{\mathsf{H}}_{k,\mathcal{D}\setminus\vartheta,m}\end{bmatrix}, \tag{D.13}$$

so that by using the above notation, the per-path MSE reads

$$\begin{aligned}\mathsf{mse}_{k,\vartheta,m} &= \mathbb{E}\left\{|h_{k,m} - \langle\tilde{h}_{k,\vartheta,m}\rangle|^2\right\} \\ &= \bar{t}_{k,m}\mathbb{E}\left\{\left|1 - \underline{\boldsymbol{m}}^{\mathsf{H}}_{k,\vartheta,m}\underline{\boldsymbol{x}}_{k,\vartheta}\right|^2\right\} + \mathbb{E}\left\{\underline{\boldsymbol{m}}^{\mathsf{H}}_{k,\vartheta,m}\boldsymbol{C}\underline{\boldsymbol{m}}^{\mathsf{H}}_{k,\vartheta,m}\right\}. \end{aligned}\tag{D.14}$$

However, in order to calculate (D.14) we first need to obtain the expressions for the noise covariance matrices given in (4.53).

Using (4.61) – (4.62) and the notation defined above, we write

$$\underline{\boldsymbol{u}}_{k,\vartheta,m} = \underline{\boldsymbol{x}}_{k,\vartheta}h_{k,m}, \qquad \underline{\tilde{\boldsymbol{u}}}_{k,\vartheta,m} = \underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}\tilde{h}_{k,\vartheta,m} + \underline{\Delta\tilde{\boldsymbol{u}}}_{k,m}, \tag{D.15}$$

and

$$\underline{\boldsymbol{m}}^{\mathsf{H}}_{k,\vartheta,m} = \bar{t}_{k,m}\tilde{\Gamma}^{-1}_{k,\vartheta,m}\underline{\tilde{\boldsymbol{x}}}^{\mathsf{H}}_{k,\vartheta}(\tilde{\boldsymbol{C}} + \tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}})^{-1}, \tag{D.16}$$

$$\tilde{\Gamma}_{k,\vartheta,m} = 1 + \bar{t}_{k,m}\underline{\tilde{\boldsymbol{x}}}^{\mathsf{H}}_{k,\vartheta}(\tilde{\boldsymbol{C}} + \tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}})^{-1}\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}. \tag{D.17}$$

With the help of (C.22) and (D.1) we get from (4.61) – (4.63) after some simplifications[1]

$$\begin{aligned}\langle\underline{\tilde{\boldsymbol{u}}}_{k,\vartheta,m}\rangle &= \underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}\langle\tilde{h}_{k,\vartheta,m}\rangle + \langle\underline{\Delta\tilde{\boldsymbol{u}}}_{k,m}\rangle \\ &= \left[\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}\underline{\boldsymbol{m}}^{\mathsf{H}}_{k,\vartheta,m} + \tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}}\left(\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}} + \tilde{\boldsymbol{C}} + \bar{t}_{k,m}\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}\underline{\tilde{\boldsymbol{x}}}^{\mathsf{H}}_{k,\vartheta}\right)^{-1}\right]\underline{\boldsymbol{z}}_{k,\vartheta,m} \\ &= \tilde{\boldsymbol{C}}(\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}} + \tilde{\boldsymbol{C}})^{-1}[\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}\underline{\boldsymbol{m}}^{\mathsf{H}}_{k,\vartheta,m} + \tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}}\tilde{\boldsymbol{C}}^{-1}]\underline{\boldsymbol{z}}_{k,\vartheta,m}. \end{aligned}\tag{D.18}$$

[1] The following identities are helpful (see, e.g. [182] and the references therein). Let $\boldsymbol{u}, \boldsymbol{v}$ be complex vectors and $\boldsymbol{A}, \boldsymbol{B}$ invertible complex matrices. Then $(\boldsymbol{A}+\boldsymbol{u}\boldsymbol{v}^{\mathsf{H}})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}\boldsymbol{A}^{-1}/(1+\boldsymbol{v}^{\mathsf{H}}\boldsymbol{A}^{-1}\boldsymbol{u})$ and $(\boldsymbol{A}+\boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1}\boldsymbol{A}^{-1}$.

The matrices (4.57) – (4.60) are then the block diagonals of

$$\begin{aligned}\boldsymbol{\Sigma}_{k,\vartheta,m} &= \mathbb{E}_k^{\mathsf{c}}\left\{(\underline{\boldsymbol{u}}_{k,\vartheta,m} - \langle\underline{\tilde{\boldsymbol{u}}}_{k,\vartheta,m}\rangle)(\underline{\boldsymbol{u}}_{k,\vartheta,m} - \langle\underline{\tilde{\boldsymbol{u}}}_{k,\vartheta,m}\rangle)^{\mathsf{H}}\right\} = \tilde{\boldsymbol{C}}(\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}} + \tilde{\boldsymbol{C}})^{-1} \\ &\times\Big[(\boldsymbol{I}_{\underline{T}_{\mathsf{coh}}} - \underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}\underline{\boldsymbol{m}}_{k,\vartheta,m}^{\mathsf{H}})(\bar{t}_{k,m}\underline{\boldsymbol{x}}_{k,\vartheta}\underline{\boldsymbol{x}}_{k,\vartheta}^{\mathsf{H}})(\boldsymbol{I}_{\underline{T}_{\mathsf{coh}}} - \underline{\boldsymbol{m}}_{k,\vartheta,m}\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}^{\mathsf{H}}) \\ &\quad+(\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}\underline{\boldsymbol{m}}_{k,\vartheta,m}^{\mathsf{H}} + \tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}}\tilde{\boldsymbol{C}}^{-1})\boldsymbol{C}(\underline{\boldsymbol{m}}_{k,\vartheta,m}\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}^{\mathsf{H}} + \tilde{\boldsymbol{C}}^{-1}\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}})\Big] \\ &\quad\times(\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}} + \tilde{\boldsymbol{C}})^{-1}\tilde{\boldsymbol{C}}, \end{aligned} \tag{D.19}$$

$$\begin{aligned}\tilde{\boldsymbol{\Sigma}}_{k,\vartheta,m} &= \mathbb{E}_k^{\mathsf{c}}\left\{(\underline{\tilde{\boldsymbol{u}}}_{k,\vartheta,m} - \langle\underline{\tilde{\boldsymbol{u}}}_{k,\vartheta,m}\rangle)(\underline{\tilde{\boldsymbol{u}}}_{k,\vartheta,m} - \langle\underline{\tilde{\boldsymbol{u}}}_{k,\vartheta,m}\rangle)^{\mathsf{H}}\right\} = \tilde{\boldsymbol{C}}(\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}} + \tilde{\boldsymbol{C}})^{-1} \\ &\times\Big[(\boldsymbol{I}_{\underline{T}_{\mathsf{coh}}} - \underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}\underline{\boldsymbol{m}}_{k,\vartheta,m}^{\mathsf{H}})(\bar{t}_{k,m}\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}^{\mathsf{H}} + \tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}})(\boldsymbol{I}_{\underline{T}_{\mathsf{coh}}} - \underline{\boldsymbol{m}}_{k,\vartheta,m}\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}^{\mathsf{H}}) \\ &\quad+(\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}\underline{\boldsymbol{m}}_{k,\vartheta,m}^{\mathsf{H}} + \tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}}\tilde{\boldsymbol{C}}^{-1})\tilde{\boldsymbol{C}}(\underline{\boldsymbol{m}}_{k,\vartheta,m}\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}^{\mathsf{H}} + \tilde{\boldsymbol{C}}^{-1}\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}})\Big] \\ &\quad\times(\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}} + \tilde{\boldsymbol{C}})^{-1}\tilde{\boldsymbol{C}}, \end{aligned} \tag{D.20}$$

where $\mathbb{E}_k^{\mathsf{c}}$ was defined in (4.52) and, with some foresight, we used a shorthand notation $\underline{T}_{\mathsf{coh}} = T_{\mathsf{coh}} - 1$. Since $\tilde{\boldsymbol{\Sigma}}_{k,\vartheta,m}$ depends only on the postulated variables, (D.20) can be further simplified by using the identity $\tilde{\Gamma}_{k,\vartheta,m}^{-1} = 1 - \underline{\boldsymbol{m}}_{k,\vartheta,m}^{\mathsf{H}}\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}$ to

$$\begin{aligned}\tilde{\boldsymbol{\Sigma}}_{k,\vartheta,m} &= \tilde{\boldsymbol{C}}(\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}} + \tilde{\boldsymbol{C}})^{-1}\Big[\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}}\tilde{\boldsymbol{C}}^{-1}(\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}} + \tilde{\boldsymbol{C}}) \\ &\quad+\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}\left(\bar{t}_{k,m}|1 - \underline{\boldsymbol{m}}_{k,\vartheta,m}^{\mathsf{H}}\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}|^2 + \underline{\boldsymbol{m}}_{k,\vartheta,m}^{\mathsf{H}}(\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}} + \tilde{\boldsymbol{C}})\underline{\boldsymbol{m}}_{k,\vartheta,m}\right)\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}^{\mathsf{H}}\Big] \\ &\quad\times(\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}} + \tilde{\boldsymbol{C}})^{-1}\tilde{\boldsymbol{C}} \\ &= \tilde{\boldsymbol{C}}(\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}} + \tilde{\boldsymbol{C}})^{-1}(\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}} + \underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}\underline{\boldsymbol{m}}_{k,\vartheta,m}^{\mathsf{H}}\tilde{\boldsymbol{C}}). \end{aligned} \tag{D.21}$$

On the other hand, using (4.53) – (4.60) and cancelling the common terms yields the following relation between the matrices $\boldsymbol{C}$ and $\tilde{\boldsymbol{C}}$,

$$\begin{aligned}\boldsymbol{C}^{-1}\lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}&\Bigg\{\sigma^2\boldsymbol{I}_{\underline{T}_{\mathsf{coh}}} + \alpha\sum_{m=1}^{M}\tilde{\boldsymbol{C}}(\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}} + \tilde{\boldsymbol{C}})^{-1}(\boldsymbol{I} - \underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}\underline{\boldsymbol{m}}_{k,\vartheta,m}^{\mathsf{H}}) \\ &\times(\bar{t}_{k,m}\underline{\boldsymbol{x}}_{k,\vartheta}\underline{\boldsymbol{x}}_{k,\vartheta}^{\mathsf{H}})(\boldsymbol{I} - \underline{\boldsymbol{m}}_{k,\vartheta,m}\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}^{\mathsf{H}})(\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}} + \tilde{\boldsymbol{C}})^{-1}\tilde{\boldsymbol{C}}\Bigg\} \\ = \tilde{\boldsymbol{C}}^{-1}\lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}&\Bigg\{\tilde{\sigma}^2\boldsymbol{I}_{\underline{T}_{\mathsf{coh}}} + \alpha\sum_{m=1}^{M}\tilde{\boldsymbol{C}}(\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}} + \tilde{\boldsymbol{C}})^{-1}(\boldsymbol{I} - \underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}\underline{\boldsymbol{m}}_{k,\vartheta,m}^{\mathsf{H}}) \\ &\times(\bar{t}_{k,m}\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}^{\mathsf{H}} + \tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}})(\boldsymbol{I} - \underline{\boldsymbol{m}}_{k,\vartheta,m}\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}^{\mathsf{H}})(\tilde{\boldsymbol{\Omega}}_{\underline{\Delta\boldsymbol{u}}_{k,m}} + \tilde{\boldsymbol{C}})^{-1}\tilde{\boldsymbol{C}}\Bigg\}. \end{aligned} \tag{D.22}$$

Since we consider the symbols in $\underline{\boldsymbol{x}}_{k,\vartheta}$ and in $\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}$ to be uncorrelated, the true and postulated noise covariance matrices simplify as $\boldsymbol{C} = \mathrm{diag}(C_{\mathsf{tr}}\boldsymbol{I}_{\tau_{\mathsf{tr}}}, C_{\mathsf{d}}\boldsymbol{I}_{\tau_{\mathsf{d}}-1})$ and

$\tilde{\boldsymbol{C}} = \mathrm{diag}(\tilde{C}_{\mathsf{tr}}\boldsymbol{I}_{\tau_{\mathsf{tr}}}, \tilde{C}_{\mathsf{d}}\boldsymbol{I}_{\tau_{\mathsf{d}}-1})$. To further simplify (D.22), however, we need to specify the operator $\varphi_{\mathsf{app}}$ for the feedback (cf. Section 4.1.1), and the parameters that define the channel estimator in Section 4.1.3.

Consider the approximate ML estimator defined in Example 9, that is, $\tilde{\boldsymbol{\Omega}}_{\underline{\Delta \boldsymbol{u}}_{k,m}} = \mathbf{0}$ and $\|\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}\|^2 = \underline{T}_{\mathsf{coh}}$. Using (4.55) – (4.56) and (D.21) reveals that $\tilde{C}_{\mathsf{d}} = \tilde{C}_{\mathsf{tr}}$, where

$$\tilde{C}_{\mathsf{tr}} = \tilde{\sigma}^2 + \alpha \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \frac{\bar{t}_{k,m}\tilde{C}_{\mathsf{tr}}}{\tilde{C}_{\mathsf{tr}} + \bar{t}_{k,m}\underline{T}_{\mathsf{coh}}}. \tag{D.23}$$

If $\underline{T}_{\mathsf{coh}} > \alpha M$, as usually is the case[2], we get from (D.23) that $\tilde{\sigma}^2 \to 0 \implies \tilde{C}_{\mathsf{tr}} \to 0$. Furthermore, by (D.23) the RHS of (D.22) simplifies as

$$\left(\frac{\tilde{\sigma}^2}{\tilde{C}_{\mathsf{tr}}} + \alpha \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \frac{\bar{t}_{k,m}\tilde{C}_{\mathsf{tr}}}{\left(\tilde{C}_{\mathsf{tr}} + \bar{t}_{k,m}\underline{T}_{\mathsf{coh}}\right)^2}\right) \boldsymbol{I}_{\underline{T}_{\mathsf{coh}}} \xrightarrow{\tilde{\sigma}^2\to 0} \left(1 - \frac{\alpha M}{\underline{T}_{\mathsf{coh}}}\right) \boldsymbol{I}_{\underline{T}_{\mathsf{coh}}}. \tag{D.24}$$

In order to tackle the LHS of (D.22), we present a small lemma related to the statistics of the hard feedback.

**Lemma 1.** *Recall from Section 4.3.1 that for the approximate APP based hard feedback with Gray mapped QPSK symbols,*

$$\Pr\left(\Re\{x_{k,t}\} \neq \Re\{\langle \tilde{x}_{k,t}\rangle_{\mathsf{app}}\}\right) = \Pr\left(\Im\{x_{k,t}\} \neq \Im\{\langle \tilde{x}_{k,t}\rangle_{\mathsf{app}}\}\right) = \varepsilon_k^{\mathsf{app}}, \tag{D.25}$$

*for any given data symbol $x_{k,t}, t \in \mathcal{D} \setminus \vartheta$. Let us also assume that the bits corresponding to the real and imaginary parts of the QPSK symbol, and the feedback symbols for time instances $t, t' \in \mathcal{D} \setminus \vartheta$ with $t \neq t'$ are independent. Then for all $k = 1, \ldots, K$,*

$$\mathbb{E}\{\Re\{\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}^{\mathsf{H}}\underline{\boldsymbol{x}}_{k,\vartheta}\}\} = \underline{T}_{\mathsf{coh}} - 2\varepsilon_k^{\mathsf{app}}(\tau_{\mathsf{d}} - 1), \tag{D.26}$$

*so that*

$$\begin{aligned}
&\left[\mathbb{E}\{\Re\{\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}(\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}^{\mathsf{H}}\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta})\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}^{\mathsf{H}}\}\}\right]_{t,t} \\
&= \begin{cases} \underline{T}_{\mathsf{coh}} - 2\varepsilon_k^{\mathsf{app}}(\tau_{\mathsf{d}} - 1), & t \in \mathcal{T}, \\ \underline{T}_{\mathsf{coh}} - 2\varepsilon_k^{\mathsf{app}}(\tau_{\mathsf{d}} - 1) - 2\varepsilon_k^{\mathsf{app}}\left[\underline{T}_{\mathsf{coh}} - 2\left(1 + \varepsilon_k^{\mathsf{app}}(\tau_{\mathsf{d}} - 2)\right)\right], & t \in \mathcal{D} \setminus \vartheta, \end{cases}
\end{aligned} \tag{D.27}$$

[2]Much like we did in Chapter 3 in the case of decorrelating MUD, one could define and analyze a similar estimator also for $\underline{T}_{\mathsf{coh}} < \alpha M$. We do not consider this in our analysis since such a case arises rarely in channel estimation.

*and*

$$\mathbb{E}\{|\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}^{\mathsf{H}}\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}|^2\} = \underline{T}_{\mathsf{coh}}^2 - 4\varepsilon_k^{\mathsf{app}}(\tau_{\mathsf{d}}-1)\left[\underline{T}_{\mathsf{coh}} - \varepsilon_k^{\mathsf{app}}(\tau_{\mathsf{d}}-2) - 1\right]. \quad \text{(D.28)}$$

**Proof:** By using (D.25), we find that

$$\begin{aligned}&\mathbb{E}\{\Re\{\langle\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}\rangle_{\mathsf{app}}^{\mathsf{H}}\boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta}\}\}\\ &\quad= \sum_{t=0}^{2(\tau_{\mathsf{d}}-1)}(\tau_{\mathsf{d}}-1-t)\binom{2(\tau_{\mathsf{d}}-1)}{t}(\varepsilon_k^{\mathsf{app}})^t(1-\varepsilon_k^{\mathsf{app}})^{2(\tau_{\mathsf{d}}-1)-t}. \quad \text{(D.29)}\end{aligned}$$

Since the last three terms on the RHS of (D.29) is just the binomial distribution with $2(\tau_{\mathsf{d}}-1)$ trials, each having a success probability of $\varepsilon_k^{\mathsf{app}}$, we immediately get (D.26) and the first part of (D.27). For $t \in \mathcal{D}\setminus\vartheta$, note that

$$\begin{aligned}&\mathbb{E}\{\Re\{\langle\tilde{x}_{k,t}\rangle_{\mathsf{app}}^{*}(\tau_{\mathsf{tr}} + \langle\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}\rangle_{\mathsf{app}}^{\mathsf{H}}\boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta})x_{k,t}\}\}\\ &= 1 + (1-2\varepsilon_k^{\mathsf{app}})\\ &\quad\times\left[\tau_{\mathsf{tr}} + \frac{1}{2}\sum_{t=0}^{2(\tau_{\mathsf{d}}-2)}[2(\tau_{\mathsf{d}}-2)-2t]\binom{2(\tau_{\mathsf{d}}-2)}{t}(1-\varepsilon_k^{\mathsf{app}})^{2(\tau_{\mathsf{d}}-2)-t}(\varepsilon_k^{\mathsf{app}})^t\right]\\ &= 1 + (1-2\varepsilon_k^{\mathsf{app}})\left[\tau_{\mathsf{tr}} + (\tau_{\mathsf{d}}-2)(1-2\varepsilon_k^{\mathsf{app}})\right], \quad \text{(D.30)}\end{aligned}$$

which gives the second part of (D.27) after re-arranging the terms. Proof of (D.28) is similar and therefore omitted. □

Let us consider the first $\tau_{\mathsf{tr}}$ diagonal terms of the user $k$ in (D.22). Using Lemma 1, we get

$$\begin{aligned}&\left[\mathbb{E}\{\left[\boldsymbol{I} - \underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}\underline{\boldsymbol{m}}_{k,\vartheta,m}^{\mathsf{H}}\right)\boldsymbol{x}_k\boldsymbol{x}_k^{\mathsf{H}}\left(\boldsymbol{I} - \underline{\boldsymbol{m}}_{k,\vartheta,m}\underline{\tilde{\boldsymbol{x}}}_{k,\vartheta}^{\mathsf{H}}\right)\}\right]_{t,t}\\ &\quad= 1 - 2\bar{t}_{k,m}\frac{\underline{T}_{\mathsf{coh}} - 2\varepsilon_k^{\mathsf{app}}(\tau_{\mathsf{d}}-1)}{\tilde{C}_{\mathsf{tr}} + \bar{t}_{k,m}\underline{T}_{\mathsf{coh}}}\\ &\qquad+\bar{t}_{k,m}^2\frac{\underline{T}_{\mathsf{coh}}^2 - 4\varepsilon_k^{\mathsf{app}}(\tau_{\mathsf{d}}-1)\left[\underline{T}_{\mathsf{coh}} - \varepsilon_k^{\mathsf{app}}(\tau_{\mathsf{d}}-2) - 1\right]}{\left(\tilde{C}_{\mathsf{tr}} + \bar{t}_{k,m}\underline{T}_{\mathsf{coh}}\right)^2}, \quad \text{(D.31)}\end{aligned}$$

for all $t \in \mathcal{T}$. Thus, the first $\tau_{\mathsf{tr}}$ diagonal terms on the LHS of (D.22) are

$$\begin{aligned}&\frac{1}{C_{\mathsf{tr}}}\Bigg[\sigma^2 + \alpha\lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\sum_{m=1}^{M}\frac{1}{\left(\tilde{C}_{\mathsf{tr}} + \bar{t}_{k,m}\underline{T}_{\mathsf{coh}}\right)^2}\\ &\qquad\times\bar{t}_{k,m}\left(\tilde{C}_{\mathsf{tr}}^2 + 4\bar{t}_{k,m}\varepsilon_k^{\mathsf{app}}(\tau_{\mathsf{d}}-1)\left[\tilde{C}_{\mathsf{tr}} + \bar{t}_{k,m}(1+\varepsilon_k^{\mathsf{app}}(\tau_{\mathsf{d}}-2))\right]\right)\Bigg]. \quad \text{(D.32)}\end{aligned}$$

Equating (D.24) and (D.32), and taking the limit $\tilde{\sigma}^2 \to 0 \implies \tilde{C}_{\mathsf{tr}} \to 0$ yields

$$C_{\mathsf{tr}} \xrightarrow{\tilde{\sigma}^2 \to 0} \frac{\sigma^2 \underline{T}_{\mathsf{coh}}}{\underline{T}_{\mathsf{coh}} - \alpha M} + \alpha \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \frac{4\bar{t}_k \varepsilon_k^{\mathsf{app}} (\tau_{\mathsf{d}} - 1)\left(1 + \varepsilon_k^{\mathsf{app}}(\tau_{\mathsf{d}} - 2)\right)}{(\underline{T}_{\mathsf{coh}} - \alpha M)\underline{T}_{\mathsf{coh}}}. \tag{D.33}$$

Similarly, from the lower $\tau_{\mathsf{d}}$ diagonal terms of (D.22) we get with the help of Lemma 1,

$$C_{\mathsf{d}} \xrightarrow{\tilde{\sigma}^2 \to 0} C_{\mathsf{tr}} + \alpha \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \frac{4\bar{t}_k \varepsilon_k^{\mathsf{app}} \left[\underline{T}_{\mathsf{coh}} - 2\left(1 + \varepsilon_k^{\mathsf{app}}(\tau_{\mathsf{d}} - 2)\right)\right]}{\underline{T}_{\mathsf{coh}} - \alpha M}, \tag{D.34}$$

Plugging (D.33) – (D.34) to (D.14) and using once more Lemma 1 finally yields

$$\begin{aligned}
&\mathbb{E}\left\{ \underline{\boldsymbol{m}}_{k,\vartheta,m}^{\mathsf{H}} \boldsymbol{C} \underline{\boldsymbol{m}}_{k,\vartheta,m} \right\} \xrightarrow{\tilde{\sigma}^2 \to 0} \frac{1}{\underline{T}_{\mathsf{coh}}^2} (\tau_{\mathsf{tr}} C_{\mathsf{tr}} + (\tau_{\mathsf{d}} - 1) C_{\mathsf{d}}) \\
&= \frac{\sigma^2}{\underline{T}_{\mathsf{coh}} - \alpha M} + \alpha \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \frac{4\bar{t}_k \varepsilon_k^{\mathsf{app}} (\tau_{\mathsf{d}} - 1)\left[\underline{T}_{\mathsf{coh}} - \left(1 + \varepsilon_k^{\mathsf{app}}(\tau_{\mathsf{d}} - 2)\right)\right]}{\underline{T}_{\mathsf{coh}}^2 (\underline{T}_{\mathsf{coh}} - \alpha M)},
\end{aligned} \tag{D.35}$$

$$\begin{aligned}
&\bar{t}_{k,m} \mathbb{E}\left\{ \left| 1 - \underline{\boldsymbol{m}}_{k,\vartheta,m}^{\mathsf{H}} \underline{\boldsymbol{x}}_{k,\vartheta} \right|^2 \right\} \xrightarrow{\tilde{\sigma}^2 \to 0} \frac{\bar{t}_{k,m}}{\underline{T}_{\mathsf{coh}}^2} \mathbb{E}\left\{ \left| (\tau_{\mathsf{d}} - 1) - \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}}^{\mathsf{H}} \boldsymbol{x}_{k,\mathcal{D}\setminus\vartheta} \right|^2 \right\} \\
&= \frac{4\bar{t}_{k,m} \varepsilon_k^{\mathsf{app}} (\tau_{\mathsf{d}} - 1)\left[1 + \varepsilon_k^{\mathsf{app}}(\tau_{\mathsf{d}} - 2)\right]}{\underline{T}_{\mathsf{coh}}^2},
\end{aligned} \tag{D.36}$$

where the fact that $\underline{T}_{\mathsf{coh}} = \tau_{\mathsf{tr}} + \tau_{\mathsf{d}} - 1$ was used to simplify the last equation. This concludes the proof of Proposition 9.

Consider next the iterative LMMSE channel estimator with soft feedback, described in Example 8. By definition, $\tilde{\sigma}^2 = \sigma^2$,

$$\mathbb{E}\{ \underline{\boldsymbol{u}}_{k,\vartheta,m} \underline{\boldsymbol{u}}_{k,\vartheta,m}^{\mathsf{H}} \mid \tilde{\underline{\boldsymbol{x}}}_{k,\vartheta} \} = \bar{t}_{k,m} \tilde{\underline{\boldsymbol{x}}}_{k,\vartheta} \tilde{\underline{\boldsymbol{x}}}_{k,\vartheta}^{\mathsf{H}} + \boldsymbol{\Omega}_{\underline{\Delta \boldsymbol{u}}_{k,m}}, \tag{D.37}$$

where $\boldsymbol{\Omega}_{\underline{\Delta \boldsymbol{u}}_{k,m}} = \mathrm{diag}(\mathbf{0}_{\tau_{\mathsf{tr}} \times \tau_{\mathsf{tr}}}, \boldsymbol{\Omega}_{\Delta \boldsymbol{u}_{k,m}})$, and $\mathbb{E}\{ \tilde{\underline{\boldsymbol{x}}}_{k,\vartheta}^{\mathsf{H}} \underline{\boldsymbol{x}}_{k,\vartheta} | \tilde{\boldsymbol{x}}_k \} = \tau_{\mathsf{tr}} + \| \langle \tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta} \rangle_{\mathsf{app}}^{\mathsf{H}} \|^2$. This implies by (D.22) that $\tilde{\boldsymbol{C}} = \boldsymbol{C}$ and, thus,

$$\tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)} = \boldsymbol{C}_{\mathcal{T}}^{(\ell)}, \qquad \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)} = \boldsymbol{C}_{\mathcal{D}\setminus\vartheta}^{(\ell)}, \tag{D.38}$$

in (4.53) – (4.56) and

$$\mathbb{E}\{ \boldsymbol{\Sigma}_{k,\mathcal{T},m}(\boldsymbol{C}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)}) \} = \mathbb{E}\{ \tilde{\boldsymbol{\Sigma}}_{k,\mathcal{T},m}(\tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)}) \}, \tag{D.39}$$

$$\mathbb{E}\{ \boldsymbol{\Sigma}_{k,\mathcal{D}\setminus\vartheta,m}(\boldsymbol{C}_{\mathcal{D}\setminus\vartheta}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)}) \} = \mathbb{E}\{ \tilde{\boldsymbol{\Sigma}}_{k,\mathcal{D}\setminus\vartheta,m}(\tilde{\boldsymbol{C}}_{\mathcal{T}}^{(\ell)}, \tilde{\boldsymbol{C}}_{\mathcal{D}\setminus\vartheta}^{(\ell)}) \}, \tag{D.40}$$

in (4.57) – (4.60). Plugging this to (D.14) yields

$$\mathsf{mse}_{k,\vartheta,m} = \mathbb{E}\left\{\frac{\bar{t}_{k,m}}{1+\bar{t}_{k,m}\big[\boldsymbol{p}_k^{\mathsf{H}}\boldsymbol{C}_{\mathcal{T}}^{-1}\boldsymbol{p}_k + \langle\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}\rangle_{\mathsf{app}}^{\mathsf{H}}\big(\boldsymbol{\Omega}_{\Delta\boldsymbol{u}_{k,m}} + \boldsymbol{C}_{\mathcal{D}\setminus\vartheta}\big)^{-1}\langle\tilde{\boldsymbol{x}}_{k,\mathcal{D}\setminus\vartheta}\rangle_{\mathsf{app}}\big]}\right\}, \tag{D.41}$$

where the noise covariances are given in (D.21). A little bit more algebra results to (4.94) – (4.96).

# *Appendix E*

# *Proof of Propositions 10 and 12*

## E.1 Derivation of (4.118) and (4.119)

Consider the set of single-user channels defined in (4.65), and assume the receiver postulates the channels (4.66). Let $\xi \in \mathcal{K}$ be the user of interest and consider the GPME defined in (4.74). The CSI is provided by the LMMSE estimator of Example 8 for all users $k = 1, \ldots, K$ in the form

$$\mathbb{Q}^{(\ell)}(\tilde{h}_{k,t,m} \mid \mathcal{I}_t^{(\ell)}) = \mathsf{CN}(\langle \tilde{h}_{k,t,m} \rangle_{(\ell)};\, \mathsf{mse}_{k,t,m}^{(\ell)}), \tag{E.1}$$

and the (unconditional) posterior mean of the channel has a complex Gaussian distribution

$$\langle \tilde{h}_{k,t,m} \rangle_{(\ell)} \sim \mathsf{CN}(0;\, \bar{t}_{k,m} - \mathsf{mse}_{k,t,m}^{(\ell)}), \tag{E.2}$$

in the large system limit, as found in Proposition 8. We also make an assumption that due to bit-interleaving and coding over several fading blocks, the data symbols and the channel estimation errors are independent.

Recall from Sections 2.3.2 and 4.3.1 that for Gray encoded QPSK signaling the extrinsic probabilities factor as given in (4.83). We postulate the priors for the data symbols

$$x_{j,t} = \frac{1}{\sqrt{2}}(a_{j,t,1} + \mathrm{j} a_{j,t,2}) \in \mathcal{M}, \qquad t \in \mathcal{D}, \tag{E.3}$$

of the interfering users $j \in \mathcal{K} \setminus \xi$ as $\mathbb{Q}(x_{j,t}) = \mathbb{Q}(a_{j,t,1})\mathbb{Q}(a_{j,t,2})$, where

$$\mathbb{Q}(a_{k,t,q}) = \frac{1 + \tanh(\lambda_{a_{j,t,q}}^{(\ell-1)}/2)}{2}\delta_{a_{j,t,q}}(+1) + \frac{1 - \tanh(\lambda_{a_{j,t,q}}^{(\ell-1)}/2)}{2}\delta_{a_{j,t,q}}(-1), \tag{E.4}$$

for $q = 1, 2$, are the probabilities and

$$\lambda_{a_{j,t,q}}^{(\ell-1)} = \log\Big(\mathbb{P}_{\mathsf{ext}}^{(\ell-1)}(a_{j,t,q} = +1)\Big) - \log\Big(\mathbb{P}_{\mathsf{ext}}^{(\ell-1)}(a_{j,t,q} = -1)\Big), \tag{E.5}$$

the extrinsic log-likelihood ratios of $a_{j,t,q} \in \{\pm 1\}, q = 1, 2$, obtained by the single-user decoders. Note that we assumed in (E.5) that $\varphi_{\mathsf{ext}}$ is an identity operator. With the help of (C.22) and (D.1), some (tedious) algebra gives the posterior mean estimates of the data symbols of interfering users

$$\langle \tilde{x}_{j,t} \rangle_j^{(\ell)} = \frac{\sum_{\tilde{x}_{j,t} \in \mathcal{M}} \tilde{x}_{j,t} \mathbb{Q}(\tilde{x}_{j,t}) \prod_{m=1}^{M} f\big(z_{j,t,m} \mid \langle \tilde{h}_{j,t,m} \rangle_{(\ell)} \tilde{x}_{j,t};\, D_t + \mathsf{mse}_{j,t,m}^{(\ell)}\big)}{\sum_{\tilde{x}_{j,t} \in \mathcal{M}} \mathbb{Q}(\tilde{x}_{j,t}) \prod_{m=1}^{M} f\big(z_{j,t,m} \mid \langle \tilde{h}_{j,t,m} \rangle_{(\ell)} \tilde{x}_{j,t};\, D_t + \mathsf{mse}_{j,t,m}^{(\ell)}\big)} \tag{E.6}$$

$$= \frac{1}{\sqrt{2}} \Bigg[ \tanh\left( \frac{\lambda_{a_{j,t,1}}^{(\ell-1)}}{2} + \sum_{m=1}^{M} \frac{\sqrt{2}}{D_t + \mathsf{mse}_{j,t,m}^{(\ell)}} \Re\Big\{ \langle \tilde{h}_{j,t,m} \rangle_{(\ell)}^{*} z_{j,t,m} \Big\} \right) + \mathrm{j} \tanh\left( \frac{\lambda_{a_{j,t,2}}^{(\ell-1)}}{2} + \sum_{m=1}^{M} \frac{\sqrt{2}}{D_t + \mathsf{mse}_{j,t,m}^{(\ell)}} \Im\Big\{ \langle \tilde{h}_{j,t,m} \rangle_{(\ell)}^{*} z_{j,t,m} \Big\} \right) \Bigg], \tag{E.7}$$

where $f\,(z \mid \mu;\, \Omega)$ is the complex Gaussian density (C.58). Similarly we get

$$\langle \tilde{h}_{j,t,m} \tilde{x}_{j,t} \rangle_j^{(\ell)} = \frac{D_t}{D_t + \mathsf{mse}_{j,t,m}^{(\ell)}} \big( \langle \tilde{h}_{j,t,m} \rangle_{(\ell)} \langle \tilde{x}_{j,t} \rangle_j^{(\ell)} + \mathsf{mse}_{j,t,m}^{(\ell)} D_t^{-1} z_{j,t,m} \big). \tag{E.8}$$

The corresponding terms for the user of interest $\xi$ are obtained from (E.7) – (E.8) by setting $\lambda_{a_{\xi,t,q}}^{(\ell-1)} = 0, q = 1, 2$. It is interesting to note that (E.8) and (D.18) have similar form.

Let us now concentrate on the case $\tilde{\sigma}^2 = \sigma^2$. Assuming the replica symmetry holds, the noise variance $D_t$ is given by the solution to the fixed point equation

$$\begin{aligned} \Sigma_{k,t,m}(D_t, \tilde{D}_t = D_t) &= \tilde{\Sigma}_{k,t,m}(D_t, \tilde{D}_t = D_t) \\ &= \mathbb{E}_k^{\mathsf{d}}\{ |h_{k,m} x_{k,t} - \langle \tilde{h}_{k,t,m} \tilde{x}_{k,t} \rangle_{(\ell)}|^2 \}, \quad \forall k \in \mathcal{K}. \end{aligned} \tag{E.9}$$

where $\mathbb{E}_k^{\mathsf{d}}\{\,\cdot\,\}$ is defined in (4.67). Denoting $\Sigma_{k,t,m}(D_t) = \Sigma_{k,t,m}(D_t, \tilde{D}_t = D_t)$,

and simplifying yields

$$\begin{aligned}
&\Sigma_{k,t,m}(D_t) = \mathbb{E}_k^{\mathsf{d}}\{|h_{k,m}x_{k,t} - \langle \tilde{h}_{k,t,m}\tilde{x}_{k,t}\rangle_k^{(\ell)}|^2\} \\
&= \frac{D_t^2}{(D_t + \mathsf{mse}_{k,t,m})^2}\Big[|h_{k,m}|^2 + D_t^{-1}\mathsf{mse}_{k,t,m}^2 \\
&\quad + |\langle \tilde{h}_{k,t,m}\rangle_{(\ell)}|^2\left(\mathbb{E}_k^{\mathsf{d}}\{|\langle \tilde{x}_{k,t}\rangle_k^{(\ell)}|^2\} - 2\Re\left\{\mathbb{E}_k^{\mathsf{d}}\{x_{k,t}^*\langle \tilde{x}_{k,t}\rangle_k^{(\ell)}\}\right\}\right) \\
&\quad - 2\Re\left\{\Delta h_{k,t,m}^*\langle \tilde{h}_{k,t,m}\rangle_{(\ell)}\mathbb{E}_k^{\mathsf{d}}\{\tilde{x}_{k,t}^*\langle \tilde{x}_{k,t}\rangle_k^{(\ell)}\}\right\}\Big] \qquad \text{(E.10)} \\
&= \frac{D_t^2}{(D_t + \mathsf{mse}_{k,t,m})^2}\left(|h_{k,m}|^2 + D_t^{-1}\mathsf{mse}_{k,t,m}^2 - |\langle \tilde{h}_{k,t,m}\rangle_{(\ell)}|^2\mathbb{E}_k^{\mathsf{d}}\{|\langle \tilde{x}_{k,t}\rangle_k^{(\ell)}|^2\}\right), \qquad \text{(E.11)}
\end{aligned}$$

where $\Delta h_{k,t,m} = h_{k,m} - \langle \tilde{h}_{k,t,m}\rangle_{(\ell)}$. The last term of (E.10) vanishes in (4.70) for the LMMSE channel estimator as $K \to \infty$, and is therefore is omitted in (E.11). We also used the first equality in (by symmetry same holds for the imaginary part of the signal)

$$\begin{aligned}
&\sum_{a_{k,t,1}\in\{\pm 1\}} a_{k,t,1}\mathbb{Q}(a_{k,t,1})\Re\{\langle \tilde{x}_{k,t}\rangle_k^{(\ell)}\} = \sum_{a_{k,t,1}\in\{\pm 1\}} \mathbb{Q}(a_{k,t,1})\Re\{\langle \tilde{x}_{k,t}\rangle_k^{(\ell)}\}^2 \qquad \text{(E.12)} \\
&= \sum_{a_{k,t,1}\in\{\pm 1\}} \frac{1 + a_{k,t,1}\tanh(\lambda_{a_{k,t,1}}^{(\ell-1)}/2)}{2} \\
&\times \int \tanh\left(a_{k,t,1}\frac{\lambda_{a_{k,t,1}}^{(\ell-1)}}{2} + \nu_k\sqrt{\sum_{m=1}^{M}\frac{|\langle \tilde{h}_{k,t,m}\rangle_{(\ell)}|^2}{D_t + \mathsf{mse}_{k,t,m}}} + \sum_{m=1}^{M}\frac{|\langle \tilde{h}_{k,t,m}\rangle_{(\ell)}|^2}{D_t + \mathsf{mse}_{k,t,m}}\right)\mathrm{D}\nu_k, \qquad \text{(E.13)}
\end{aligned}$$

where we used the fact that for $\lambda \in \mathbb{R}$ and $E > 0$

$$\begin{aligned}
&\int \frac{1 + \tanh(\lambda/2)}{2}\tanh\left(E + \nu\sqrt{E} + \frac{\lambda}{2}\right) \\
&\qquad - \frac{1 - \tanh(\lambda/2)}{2}\tanh\left(-E + \nu\sqrt{E} + \frac{\lambda}{2}\right)\mathrm{D}\nu \\
&= \int \frac{1 + \tanh(\lambda/2)}{2}\tanh\left(E + \nu\sqrt{E} + \frac{\lambda}{2}\right) \\
&\qquad + \frac{1 - \tanh(\lambda/2)}{2}\tanh\left(E + \nu\sqrt{E} - \frac{\lambda}{2}\right)\mathrm{D}\nu, \qquad \text{(E.14)}
\end{aligned}$$

where $\mathrm{D}\nu$ is defined in (2.44). Using (E.13) and the knowledge that $\{a_{k,t,q}\}_{q=1}^{2}$ are IID yields

$$\Sigma_{k,t,m}(D_t) = \frac{D_t^2}{(D_t+\mathsf{mse}_{k,t,m})^2}\Big[|h_{k,m}|^2 + D_t^{-1}\mathsf{mse}_{k,t,m}^2 - |\langle\tilde{h}_{k,t,m}\rangle_{(\ell)}|^2\,\mathbb{E}_{\lambda_{a_{k,t,1}}^{(\ell-1)}}\Bigg\{\sum_{a_{k,t,1}\in\{\pm1\}}\frac{1+a_{k,t,1}\tanh(\lambda_{a_{k,t,1}}^{(\ell-1)}/2)}{2} \times\int\tanh\left(a_{k,t,1}\frac{\lambda_{a_{k,t,1}}^{(\ell-1)}}{2}+w\sqrt{\sum_{m=1}^{M}\frac{|\langle\tilde{h}_{k,t,m}\rangle_{(\ell)}|^2}{D_t+\mathsf{mse}_{k,t,m}}} + \sum_{m=1}^{M}\frac{|\langle\tilde{h}_{k,t,m}\rangle_{(\ell)}|^2}{D_t+\mathsf{mse}_{k,t,m}}\right)\mathrm{D}w\Bigg]\Bigg\}. \tag{E.15}$$

Little bit of algebra and re-organizing the terms completes the proof.

## E.2 Derivation of (4.110) and (4.120).

Consider the set of single-user channels defined in (4.65), and assume the receiver postulates the channels (4.76). The GPME is defined in (4.81) and the posterior mean estimates $\{\langle\tilde{h}_{k,t,m}\rangle_{(\ell)}\}_{m=1}^{M}$ of the channel are provided by the LMMSE estimator of of Example 8 for all users $k=1,\ldots,K$. The (unconditional) posterior mean of the channel has a complex Gaussian distribution (E.2) in the large system limit, as found in Proposition 8. We also define the vectors

$$\boldsymbol{z}_{k,t} = [z_{k,t,1}\ \cdots\ z_{k,t,M}]^{\mathsf{T}}, \tag{E.16}$$

$$\langle\tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} = [\langle\tilde{h}_{k,t,1}\rangle_{(\ell)}\ \cdots\ \langle\tilde{h}_{k,t,M}\rangle_{(\ell)}]^{\mathsf{T}}, \tag{E.17}$$

$$\tilde{\boldsymbol{v}}_{k,t} = \langle\tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)}\tilde{x}_{k,t} + \Delta\tilde{\boldsymbol{v}}_{k,t}, \tag{E.18}$$

where $\Delta\tilde{\boldsymbol{v}}_{k,t}\in\mathbb{C}^M$ is given in (4.80). For notational convenience the iteration index will be omitted in the following.

Let the user of interest be $\xi\in\mathcal{K}$. For the interfering users $j\in\mathcal{K}\setminus\xi$, the postulated a priori probability is given in (4.82). Performing the Gaussian integrals in (4.81) by using (C.22) and (D.1) yields

$$\langle\tilde{x}_{j,t}\rangle_j^{(\ell)} = \boldsymbol{m}_{j,t}^{\mathsf{H}}\boldsymbol{z}_{j,t} + \tilde{\Gamma}_{j,t}^{-1}\langle\tilde{x}_{j,t}\rangle_{\mathsf{ext}}^{(\ell-1)} \tag{E.19}$$

$$\boldsymbol{m}_{j,t}^{\mathsf{H}} = \frac{\tilde{\Omega}_{\Delta x_{j,t}}^{(\ell-1)}}{\tilde{\Gamma}_{j,t}}\langle\tilde{\boldsymbol{h}}_{j,t}\rangle_{(\ell)}^{\mathsf{H}}(\tilde{D}_t\boldsymbol{I}_M+\tilde{\boldsymbol{\Omega}}_{\Delta\boldsymbol{v}_{j,t}}^{(\ell)})^{-1}, \tag{E.20}$$

$$\tilde{\Gamma}_{j,t} = 1+\tilde{\Omega}_{\Delta x_{j,t}}^{(\ell)}\langle\tilde{\boldsymbol{h}}_{j,t}\rangle_{(\ell)}^{\mathsf{H}}(\tilde{D}_t\boldsymbol{I}_M+\tilde{\boldsymbol{\Omega}}_{\Delta\boldsymbol{v}_{j,t}}^{(\ell)})^{-1}\langle\tilde{\boldsymbol{h}}_{j,t}\rangle_{(\ell)}. \tag{E.21}$$

A little bit more algebra gives

$$\langle \Delta\tilde{\boldsymbol{v}}_{j,t}\rangle_j^{(\ell)} = \tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta \boldsymbol{v}_{j,t}} \left[\tilde{D}_t \boldsymbol{I}_M + \tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta \boldsymbol{v}_{j,t}} + \tilde{\Omega}^{(\ell-1)}_{\Delta x_{j,t}} \langle \tilde{\boldsymbol{h}}_{j,t}\rangle_{(\ell)} \langle \tilde{\boldsymbol{h}}_{j,t}\rangle_{(\ell)}^{\mathsf{H}}\right]^{-1} \times \left(\boldsymbol{z}_{j,t} - \langle \tilde{\boldsymbol{h}}_{j,t}\rangle_{(\ell)} \langle \tilde{x}_{j,t}\rangle_{\mathsf{ext}}^{(\ell-1)}\right), \tag{E.22}$$

and, thus,

$$\begin{aligned}\langle \Delta\tilde{\boldsymbol{v}}_{j,t}\rangle_j^{(\ell)} &= \langle \tilde{\boldsymbol{h}}_{j,t}\rangle_{(\ell)} \langle \tilde{x}_{j,t}\rangle_j^{(\ell)} + \langle \Delta\tilde{\boldsymbol{v}}_{j,t}\rangle_j^{(\ell)} \\ &= \tilde{D}_t (\tilde{D}_t \boldsymbol{I}_M + \tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta \boldsymbol{v}_{j,t}})^{-1} \\ &\quad \times \left[ (\langle \tilde{\boldsymbol{h}}_{j,t}\rangle_{(\ell)} \boldsymbol{m}_{j,t}^{\mathsf{H}} + \tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta \boldsymbol{v}_{j,t}} \tilde{D}_t^{-1}) \boldsymbol{z}_{j,t} + \tilde{\Gamma}_{j,t}^{-1} \langle \tilde{\boldsymbol{h}}_{j,t}\rangle_{(\ell)} \langle \tilde{x}_{j,t}\rangle_{\mathsf{ext}}^{(\ell-1)} \right] \\ &= \tilde{D}_t (\tilde{D}_t \boldsymbol{I}_M + \tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta \boldsymbol{v}_{j,t}})^{-1} \left( \langle \tilde{\boldsymbol{h}}_{j,t}\rangle_{(\ell)} \langle \tilde{x}_{j,t}\rangle_j^{(\ell)} + \tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta \boldsymbol{v}_{j,t}} \tilde{D}_t^{-1} \boldsymbol{z}_{j,t} \right), \end{aligned} \tag{E.23}$$

where (E.23) is similar to the expressions (D.18) and (E.8). Like in the previous section, we can obtain the desired user's equation by setting $\tilde{\Omega}^{(\ell)}_{\Delta x_{\xi,t}} = 1$ and $\langle \tilde{x}_{\xi,t}\rangle_{\mathsf{ext}}^{(\ell)} = 0$. Since

$$\langle \tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} \tilde{\Gamma}_{k,t}^{-1} = (\boldsymbol{I}_M - \langle \tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} \boldsymbol{m}_{k,t}^{\mathsf{H}}) \langle \tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)}, \tag{E.24}$$

we have

$$\begin{aligned} \boldsymbol{h}_k x_{k,t} - \langle \tilde{\boldsymbol{v}}_{k,t}\rangle_k^{(\ell)} &= \tilde{D}_t (\tilde{D}_t \boldsymbol{I}_M + \tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta \boldsymbol{v}_{k,t}})^{-1} \\ &\quad \times \Big[ (\boldsymbol{I}_M - \langle \tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} \boldsymbol{m}_{k,t}^{\mathsf{H}}) (\boldsymbol{h}_k x_{k,t} - \langle \tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} \langle \tilde{x}_{k,t}\rangle_{\mathsf{ext}}^{(\ell-1)}) \\ &\qquad - (\langle \tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} \boldsymbol{m}_{k,t}^{\mathsf{H}} + \tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta \boldsymbol{v}_{k,t}} \tilde{D}_t^{-1}) \boldsymbol{w}_{k,t} \Big], \end{aligned} \tag{E.25}$$

$$\begin{aligned} \tilde{\boldsymbol{v}}_{k,t} - \langle \tilde{\boldsymbol{v}}_{k,t}\rangle_k^{(\ell)} &= \tilde{D}_t (\tilde{D}_t \boldsymbol{I}_M + \tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta \boldsymbol{v}_{k,t}})^{-1} \\ &\quad \times \Big[ (\boldsymbol{I}_M - \langle \tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} \boldsymbol{m}_{k,t}^{\mathsf{H}}) (\Delta\tilde{\boldsymbol{v}}_{k,t} + \langle \tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} (\tilde{x}_{k,t} - \langle \tilde{x}_{k,t}\rangle_{\mathsf{ext}}^{(\ell-1)})) \\ &\qquad - (\langle \tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} \boldsymbol{m}_{k,t}^{\mathsf{H}} + \tilde{\boldsymbol{\Omega}}^{(\ell)}_{\Delta \boldsymbol{v}_{k,t}} \tilde{D}_t^{-1}) \tilde{\boldsymbol{w}}_{k,t} \Big]. \end{aligned} \tag{E.26}$$

Recalling the definitions (4.35) – (4.37) we get

$$\Delta \boldsymbol{v}_{k,t} + \langle \tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} \Delta x_{k,t} = \boldsymbol{h}_k x_{k,t} - \langle \tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} \langle \tilde{x}_{k,t}\rangle_{\mathsf{ext}}^{(\ell-1)}, \tag{E.27}$$

so that writing $\Delta\tilde{x}_{k,t} = \tilde{x}_{k,t} - \langle \tilde{x}_{k,t}\rangle_{\mathsf{ext}}^{(\ell-1)}$, by (4.70) and (4.77) the connection

between the true and postulated noise variances is obtained

$$
\begin{aligned}
D_t^{-1}\sigma^2 + \lim_{K\to\infty} \frac{\alpha}{KD_t} \sum_{k=1}^{K} \operatorname{tr}\Big[\mathbb{E}_k^{\mathsf{d}}\Big\{ & \tilde{D}_t(\tilde{D}_t \boldsymbol{I}_M + \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_{k,t}})^{-1}(\boldsymbol{I}_M - \langle \tilde{\boldsymbol{h}}_{k,t}\rangle \boldsymbol{m}_{k,t}^{\mathsf{H}}) \\
& \times (\Delta\tilde{\boldsymbol{v}}_{k,t} + \langle \tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} \Delta\tilde{x}_{k,t})(\Delta\tilde{\boldsymbol{v}}_{k,t} + \langle \tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} \Delta\tilde{x}_{k,t})^{\mathsf{H}} \\
& \times (\boldsymbol{I}_M - \boldsymbol{m}_{k,t}\langle \tilde{\boldsymbol{h}}_{k,t}\rangle^{\mathsf{H}})(\tilde{D}_t \boldsymbol{I}_M + \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_{k,t}})^{-1} \tilde{D}_t \Big\}\Big] \\
= \tilde{D}_t^{-1}\tilde{\sigma}^2 + \lim_{K\to\infty} \frac{\alpha}{K\tilde{D}_t} \sum_{k=1}^{K} \operatorname{tr}\Big[\mathbb{E}_k^{\mathsf{d}}\Big\{ & \tilde{D}_t(\tilde{D}_t \boldsymbol{I}_M + \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_{k,t}})^{-1}(\boldsymbol{I}_M - \langle \tilde{\boldsymbol{h}}_{k,t}\rangle \boldsymbol{m}_{k,t}^{\mathsf{H}}) \\
& \times (\Delta\boldsymbol{v}_{k,t} + \langle \tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} \Delta x_{k,t})(\Delta\boldsymbol{v}_{k,t} + \langle \tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} \Delta x_{k,t})^{\mathsf{H}} \\
& \times (\boldsymbol{I}_M - \boldsymbol{m}_{k,t}\langle \tilde{\boldsymbol{h}}_{k,t}\rangle^{\mathsf{H}})(\tilde{D}_t \boldsymbol{I}_M + \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_{k,t}})^{-1} \tilde{D}_t \Big\}\Big].
\end{aligned}
\tag{E.28}
$$

For the postulated noise covariance, on the other hand,

$$
\begin{aligned}
&\mathbb{E}_k^{\mathsf{d}}\{(\tilde{\boldsymbol{v}}_{k,t} - \langle \tilde{\boldsymbol{v}}_{k,t}\rangle)(\tilde{\boldsymbol{v}}_{k,t} - \langle \tilde{\boldsymbol{v}}_{k,t}\rangle)^{\mathsf{H}}\} \\
&\quad = \tilde{D}_t(\tilde{D}_t \boldsymbol{I}_M + \tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_{k,t}})^{-1} \left(\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_{k,t}} + \langle \tilde{\boldsymbol{h}}_{k,t}\rangle \boldsymbol{m}_{k,t}^{\mathsf{H}} \tilde{D}_t\right),
\end{aligned}
\tag{E.29}
$$

which has the same form as (D.21).

Let us now consider the special case of SUMF-based receiver described in Example 12. Given arbitrary $\tilde{\Omega}_{\Delta x_{k,t}} > 0$, we get from (4.71) and (E.29) $\tilde{\sigma}^2 \to \infty \implies \tilde{D}_t \to \infty$ and $\tilde{\sigma}^2/\tilde{D}_t \to 1$. On the other hand, postulating first $\tilde{\Omega}_{\Delta x_{k,t}} = 0$ and taking then the limit $\tilde{\sigma}^2 \to \infty$ gives also $\tilde{D}_t \to \infty$ and $\tilde{\sigma}^2/\tilde{D}_t \to 1$, as expected. Since

$$
\lim_{\tilde{D}_t\to\infty} (\boldsymbol{I}_M - \langle \tilde{\boldsymbol{h}}_{k,t}\rangle \boldsymbol{m}_{k,t}^{\mathsf{H}}) = \boldsymbol{I}_M,
\tag{E.30}
$$

the general expression for the decoupled noise variance of the data detector in Example 12 reads

$$
\begin{aligned}
D_t = \sigma^2 + \alpha \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \operatorname{tr}\Big[\mathbb{E}_k^{\mathsf{d}}\Big\{ & (\boldsymbol{h}_k x_{k,t} - \langle \tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} \langle \tilde{x}_{k,t}\rangle_{\mathsf{ext}}^{(\ell-1)}) \\
& \times (\boldsymbol{h}_k x_{k,t} - \langle \tilde{\boldsymbol{h}}_{k,t}\rangle_{(\ell)} \langle \tilde{x}_{k,t}\rangle_{\mathsf{ext}}^{(\ell-1)})^{\mathsf{H}} \Big\}\Big],
\end{aligned}
\tag{E.31}
$$

which completes the proof of Proposition 10.

Now, recall that the multipaths are uncorrelated and let the CSI be provided by the LMMSE CE defined in Example 8, so that $\mathbb{E}\{\Delta\boldsymbol{h}_k \langle \tilde{\boldsymbol{h}}_{k,t}\rangle^{\mathsf{H}}\} = \boldsymbol{0}$. For the LMMSE-PIC MUDD described in Example 11, $\tilde{\sigma}^2 = \sigma^2$ and $\tilde{\boldsymbol{\Omega}}_{\Delta \boldsymbol{v}_{k,t}} = \boldsymbol{\Omega}_{\Delta \boldsymbol{v}_{k,t}} =$

$\mathrm{diag}\big([\mathsf{mse}_{k,t,1},\ldots,\mathsf{mse}_{k,t,M}]\big)\ \forall k$, where $\{\mathsf{mse}_{k,t,m}\}_{m=1}^{M}$ are the per-path MSEs obtained by the channel estimator. This implies due to (E.28) that $\tilde{D}_t = D_t$. Using (4.71), (4.78) and (E.29) gives after some algebra (4.118) and (4.120).

# *Appendix F*

# *Saddle Point Integration for Multivariate Functions with Complex Arguments*

Consider calculating an integral of the form

$$I_K = \int_C h(\boldsymbol{z}) e^{-Kg(\boldsymbol{z})} \mathrm{d}\boldsymbol{z}, \tag{F.1}$$

where the integral is along the curve $C$, $\boldsymbol{z} \in \mathbb{C}^d$, $g$ is a real valued function and $h$ changes slowly compared to $g$. With some abuse of notation, let

$$\boldsymbol{z} = \boldsymbol{x} + j\boldsymbol{y}, \text{ where } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d \iff \boldsymbol{c} = \begin{bmatrix} \boldsymbol{z} \\ \boldsymbol{z}^* \end{bmatrix} \in \mathbb{C}^{2d} \text{ so that } g(\boldsymbol{z}) = g(\boldsymbol{c}), \tag{F.2}$$

be two equivalent representations of the function $g$ with complex argument $\boldsymbol{z}$. We define

$$\frac{\partial g}{\partial \boldsymbol{c}} = \begin{bmatrix} \frac{\partial g}{\partial \boldsymbol{z}} & \frac{\partial g}{\partial \boldsymbol{z}^*} \end{bmatrix}, \qquad \frac{\partial g}{\partial \boldsymbol{z}} = \frac{1}{2}\left(\frac{\partial g}{\partial \boldsymbol{x}} - \mathrm{j}\frac{\partial g}{\partial \boldsymbol{y}}\right), \qquad \frac{\partial g}{\partial \boldsymbol{z}^*} = \frac{1}{2}\left(\frac{\partial g}{\partial \boldsymbol{x}} + \mathrm{j}\frac{\partial g}{\partial \boldsymbol{y}}\right), \tag{F.3}$$

and expand $g(\boldsymbol{z})$ as Taylor series around the point $\boldsymbol{z}_0 \in \mathbb{C}^d$ $(= \boldsymbol{c}_0 \in \mathbb{C}^{2d})$

$$\begin{aligned} g(\boldsymbol{c}) = g(\boldsymbol{c}_0) + \underbrace{\frac{\partial g(\boldsymbol{c}_0)}{\partial \boldsymbol{c}}}_{=\nabla_{\boldsymbol{c}}\, g(\boldsymbol{c}_0)} (\boldsymbol{c} - \boldsymbol{c}_0) \\ + \frac{1}{2}(\boldsymbol{c} - \boldsymbol{c}_0)^{\mathsf{H}} \Bigg[ \underbrace{\frac{\partial}{\partial \boldsymbol{c}} \left( \frac{\partial g(\boldsymbol{c}_0)}{\partial \boldsymbol{c}} \right)^{\mathsf{H}}}_{=\nabla^2_{\boldsymbol{c}\boldsymbol{c}}\, g(\boldsymbol{c}_0)} \Bigg] (\boldsymbol{c} - \boldsymbol{c}_0) + \cdots \end{aligned} \tag{F.4}$$

where $\nabla_{\boldsymbol{c}}\, g(\boldsymbol{c}_0)$ is the complex gradient and

$$\nabla^2_{\boldsymbol{cc}}\, g(\boldsymbol{c}_0) = \begin{bmatrix} \nabla^2_{\boldsymbol{zz}}\, g(\boldsymbol{c}_0) & \nabla^2_{\boldsymbol{z}^*\boldsymbol{z}}\, g(\boldsymbol{c}_0) \\ \left(\nabla^2_{\boldsymbol{z}^*\boldsymbol{z}}\, g(\boldsymbol{c}_0)\right)^* & \left(\nabla^2_{\boldsymbol{zz}}\, g(\boldsymbol{c}_0)\right)^* \end{bmatrix}, \tag{F.5}$$

the complex Hessian of $g$ at $\boldsymbol{c}_0$ [203]. Note that the Hessian is a Hermitian matrix and, thus, the block matrices satisfy $\nabla^2_{\boldsymbol{zz}}\, g(\boldsymbol{c}_0) = \left(\nabla^2_{\boldsymbol{zz}}\, g(\boldsymbol{c}_0)\right)^{\mathsf{H}}$ and $\nabla^2_{\boldsymbol{z}^*\boldsymbol{z}}\, g(\boldsymbol{c}_0) = \left(\nabla^2_{\boldsymbol{z}^*\boldsymbol{z}}\, g(\boldsymbol{c}_0)\right)^{\mathsf{T}}$.

Now, let $\boldsymbol{z}_0$ be a local extrema[1] so that $\nabla_{\boldsymbol{c}}\, g(\boldsymbol{c}_0) = \boldsymbol{0}$, and approximate $g(\boldsymbol{z})$ near $\boldsymbol{z}_0$ by the linear and second order terms on the right hand side of (F.4). Write $\boldsymbol{z} - \boldsymbol{z}_0 = \boldsymbol{\Phi}\boldsymbol{t}$ where $\boldsymbol{\Phi} = \operatorname{diag}\left([e^{\mathrm{j}\phi_1} \cdots e^{\mathrm{j}\phi_d}]\right) \in \mathbb{C}^{d\times d}$ and $\boldsymbol{t} = [t_1 \cdots t_d]^{\mathsf{T}} \in \mathbb{R}^d$. Fix the angle $\boldsymbol{\Phi}$ and the integral in (F.1) becomes

$$I_K \approx h(\boldsymbol{z}_0)e^{-Kg(\boldsymbol{z}_0)} \int_{\mathbb{R}^d} \exp\left[-\frac{1}{2}K\boldsymbol{t}^{\mathsf{T}}\left(2\Re\{\nabla^2_{\boldsymbol{\Phi}}\, g(\boldsymbol{c}_0)\}\right)\boldsymbol{t}\right] \det(\boldsymbol{\Phi})\mathrm{d}\boldsymbol{t}, \tag{F.6}$$

where

$$\nabla^2_{\boldsymbol{\Phi}}\, g(\boldsymbol{c}_0) = \boldsymbol{\Phi}^{\mathsf{H}}(\nabla^2_{\boldsymbol{zz}}\, g(\boldsymbol{c}_0))\boldsymbol{\Phi} + \boldsymbol{\Phi}^{\mathsf{H}}(\nabla^2_{\boldsymbol{z}^*\boldsymbol{z}}\, g(\boldsymbol{c}_0))\boldsymbol{\Phi}^{\mathsf{H}} \in \mathbb{C}^{d\times d}, \tag{F.7}$$

is a Hermitian matrix. Performing the Gaussian integral over $\boldsymbol{t}$ yields the final result

$$I_K \approx h(\boldsymbol{z}_0)e^{-Kg(\boldsymbol{z}_0)} \det(\boldsymbol{\Phi}) \sqrt{\frac{(\pi/K)^d}{\det\left(\Re\{\nabla^2_{\boldsymbol{\Phi}}\, g(\boldsymbol{c}_0)\}\right)}}, \qquad \boldsymbol{z} \in \mathbb{C}^d. \tag{F.8}$$

Two special cases

- $\boldsymbol{z} = \boldsymbol{x} \in \mathbb{R}^d$, $\boldsymbol{\Phi} = \boldsymbol{I}$;
- $\boldsymbol{z} = \mathrm{j}\boldsymbol{y} \in \mathbb{C}^d$, $\boldsymbol{y} \in \mathbb{R}^d$, $\boldsymbol{\Phi} = \mathrm{j}\boldsymbol{I}$;

are obtained from (F.3) and (F.7) by noticing that

$$\nabla^2_{\boldsymbol{I}}\, g(\boldsymbol{c}_0) = \frac{1}{2}\nabla^2_{\Re}\, g(\boldsymbol{x}_0) = \frac{1}{4}\cdot 2\frac{\partial}{\partial \boldsymbol{x}}\left(\frac{\partial g(\boldsymbol{x}_0)}{\partial \boldsymbol{x}}\right)^{\mathsf{T}}, \qquad \boldsymbol{x} \in \mathbb{R}^d, \tag{F.9}$$

$$\nabla^2_{\mathrm{j}\boldsymbol{I}}\, g(\boldsymbol{c}_0) = \frac{1}{2}\nabla^2_{\Im}\, g(\mathrm{j}\boldsymbol{y}_0) = \frac{1}{4}\cdot 2\frac{\partial}{\partial \boldsymbol{y}}\left(\frac{\partial g(\mathrm{j}\boldsymbol{y}_0)}{\partial \boldsymbol{y}}\right)^{\mathsf{H}}, \qquad \boldsymbol{y} \in \mathbb{R}^d, \tag{F.10}$$

and, therefore,

$$I_K \approx h(\boldsymbol{x}_0)e^{-Kg(\boldsymbol{x}_0)} \sqrt{\frac{(2\pi/K)^d}{\det\left(\nabla^2_{\Re}\, g(\boldsymbol{x}_0)\right)}} \qquad \boldsymbol{x}_0 \in \mathbb{R}^d, \tag{F.11}$$

$$I_K \approx h(\mathrm{j}\boldsymbol{y}_0)e^{-Kg(\mathrm{j}\boldsymbol{y}_0)} \det(\mathrm{j}\boldsymbol{I}_d) \sqrt{\frac{(2\pi/K)^d}{\det\left(\Re\{\nabla^2_{\Im}\, g(\mathrm{j}\boldsymbol{y}_0)\}\right)}}, \qquad \boldsymbol{y}_0 \in \mathbb{R}^d. \tag{F.12}$$

[1] In fact, it has to be a saddle point in this case.

# *Bibliography*


[1] R. Prasad and T. Ojanperä.
An overview of CDMA evolution toward wideband CDMA.
*IEEE Commun. Surv. Tutorials*, 1(1):2–29, 1998.

[2] H. Holma and A. Toskala, editors.
*WCDMA for UMTS*.
John Wiley & Sons, Chichester, West Sussex, UK, 3rd edition, 2004.

[3] A. J. Viterbi.
*CDMA: Principles of Spread Spectrum Communication*.
Addison-Wesley Wireless Communications Series. Addison-Wesley, Reading, MA, USA, 1995.

[4] S. Moshavi.
Multi-user detection for DS-CDMA communications.
*IEEE Commun. Mag.*, 34(10):124–137, October 1996.

[5] S. Verdú.
*Multiuser Detection*.
Cambridge University Press, Cambridge, UK, 1998.

[6] J. G. Proakis.
*Digital Communications*.
McGraw-Hill, Inc., New York, USA, 4th edition, 2001.

[7] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bölcskei.
An overview of MIMO communications: A key to gigabit wireless.
*Proc. IEEE*, 92(2):198–218, February 2004.

[8] A. Goldsmith.
*Wireless Communications*.
Cambridge University Press, Cambridge, UK, 2005.

[9] D. Tse and P. Viswanath.
*Fundamentals of Wireless Communication*.
Cambridge University Press, Cambridge, UK, 2005.

[10] S. Verdú.
Recent progress in multiuser detection.
In W. A. Porter and S. C. Kak, editors, *Advances in Communitions and Signal Processing*, Lecture Notes in Control and Information Sciences, pages 27–38. Springer-Verlag, Berlin, 1989.

[11] T. M. Cover and J. A. Thomas.
*Elements of Information Theory*.
Wiley-Interscience, New York, USA, 2nd edition, 2006.

[12] A. M. Tulino and S. Verdú.
Random matrix theory and wireless communications.
*Foundations and Trends in Communications and Information Theory*, 1(1): 1–182, 2004.

[13] V. Dotsenko.
*Introduction to the Replica Theory of Disordered Statistical Systems*.
Cambridge University Press, New York, 2001.

[14] H. Nishimori.
*Statistical Physics of Spin Glasses and Information Processing*.
Oxford University Press, New York, 2001.

[15] M. Mézard and A. Montanari.
*Information, Physics, and Computation*.
Oxford University Press, New York, 2009.

[16] Iterative, soft signal processing for digital communications.
A. Kavcic, J. M. F. Moura, and V. Bhagavatula, Eds., *IEEE Signal Processing Mag.*, vol. 21, no. 1, 2004.

[17] A. J. Viterbi.
Wireless digital communication: A view based on three lessons learned.
*IEEE Commun. Mag.*, 29(9):33–36, September 1991.

[18] B. Sklar.
A primer on turbo code concepts.

*IEEE Commun. Mag.*, 35:94–102, December 1997.

[19] C. Luschi, M. Sandell, P. Strauch, J.-J. Wu, C. Ilas, P.-W. Ong, R. Baeriswyl, F. Battaglia, S. Karageorgis, and R.-H. Yan.
Advanced signal-processing algorithms for energy-efficient wireless communications.
*Proc. IEEE*, 88(10):1633–1650, October 2000.

[20] L. Hanzo, J. P. Woodard, and P. Robertson.
Turbo decoding and detection for wireless applications.
*Proc. IEEE*, 95(6):1178–1200, June 2007.

[21] D. J. C. MacKay.
*Information Theory, Inference, and Learning Algorithms*.
Cambridge University Press, Cambridge, UK, 2003.

[22] E. Biglieri.
*Coding for Wireless Channels*.
Springeri-Science+Business Media, New York, 2005.

[23] A. Montanari and R. Urbanke.
Modern coding theory: The statistical mechanics and computer science point of view.
Lecture notes for Les Houches Summer School on Complex Systems, arXiv:0704.2857v1 [cs.IT], April 2007.

[24] T. Richardson and R. Urbanke.
*Modern Coding Theory*.
Cambridge University Press, Cambridge, UK, 2008.

[25] H.-A. Loeliger.
An introduction to factor graphs.
*IEEE Signal Processing Mag.*, 21(1):28–41, January 2004.

[26] H.-A. Loeliger, J. Dauwels, Junli Hu, S. Korl, Li Ping, and F. R. Kschischang.
The factor graph approach to model-based signal processing.
*Proc. IEEE*, 95(6):1295–1322, June 2007.

[27] B. Vucetic and J. Yuan.
*Space-Time Coding*.
John Wiley & Sons Ltd, Chichester, West Sussex, UK, 2003.

[28] K. S. Schneider.
Optimum detection of code division multiplexed signals.
*IEEE Trans. Aerosp. Electron. Syst.*, 15(1):181–185, January 1979.

[29] S. Verdú.
*Optimum Multiuser Signal Detection*.
PhD thesis, University of Illinois, Urbana-Champaign, IL, USA, 1984.

[30] S. Verdú.
Optimum multiuser asymptotic efficiency.
*IEEE Trans. Commun.*, 34(9):890–897, September 1986.

[31] S. Verdú.
Minimum probability of error for asynchronous Gaussian multiple-access channels.
*IEEE Trans. Inform. Theory*, 32(1):85–96, January 1986.

[32] S. Verdú.
Computational complexity of optimum multiuser detection.
*Algorithmica*, 4(3):303–312, 1989.

[33] R. Lupas.
*Near-Far Resistant Linear Multiuser Detection*.
PhD thesis, Princeton University, Princeton, NJ, 1989.

[34] R. Lupas and S. Verdú.
Linear multiuser detectors for synchronous code-division multiple-access channels.
*IEEE Trans. Inform. Theory*, 34(1):123–136, January 1989.

[35] R. Lupas and S. Verdú.
Near-far resistance of multiuser detectors in asynchronous channels.
*IEEE Trans. Commun.*, 38(4):496–508, April 1990.

[36] Z. Xie, R. T. Short, and C. K. Rushforth.
A family of suboptimum detectors for coherent multiuser communications.
*IEEE J. Select. Areas Commun.*, 8(4):683–690, May 1990.

[37] A. J. Viterbi.
Very low rate convolutional codes for maximum theoretical performance of spread-spectrum multiple-access channels.

*IEEE J. Select. Areas Commun.*, 8(4):641–649, May 1990.

[38] A. Duel-Hallen.
Decorrelating decision-feedback multiuser detector for synchronous code-division multiple-access channel.
*IEEE Trans. Commun.*, 41(2):285–290, February 1993.

[39] M. K. Varanasi and B. Aazhang.
Multistage detection in asynchronous code-division multiple-access communications.
*IEEE Trans. Commun.*, 38(4):509–519, April 1990.

[40] M. K. Varanasi and B. Aazhang.
Near-optimum detection in synchronous code-division multiple-access systems.
*IEEE Trans. Commun.*, 39(5):725–736, May 1991.

[41] Z. Xie, C. K. Rushforth, and R. T. Short.
Multiuser signal detection using sequential decoding.
*IEEE Trans. Commun.*, 38(5):578–583, May 1990.

[42] R. K. Morrow, Jr. and J. S. Lehnert.
Bit-to-bit error dependence in slotted DS/SSMA packet systems with random signature sequences.
*IEEE Trans. Commun.*, 37(10):1052–1061, October 1989.

[43] D. N. C. Tse and S. Verdú.
Optimum asymptotic multiuser efficiency of randomly spread CDMA.
*IEEE Trans. Inform. Theory*, 46(7):2718–2722, November 2000.

[44] A. J. Grant and P. D. Alexander.
Random sequence multisets for synchronous code-division multiple-access channels.
*IEEE Trans. Inform. Theory*, 44(7):2832–2836, November 1998.

[45] U. Madhow and M. L. Honig.
On the average near-far resistance for MMSE detection of direct sequence CDMA signals with random spreading.
*IEEE Trans. Inform. Theory*, 45(6):2039–2045, September 1999.

[46] R. R. Müller.
Multiuser receivers for randomly spread signals: Fundamental limits with and without decision-feedback.
*IEEE Trans. Inform. Theory*, 47(1):268–283, January 2001.

[47] S. Verdú and S. Shamai.
Spectral efficiency of CDMA with random spreading.
*IEEE Trans. Inform. Theory*, 45(2):622–640, March 1999.

[48] S. Shamai and S. Verdú.
The impact of frequency-flat fading on the spectral efficiency of CDMA.
*IEEE Trans. Inform. Theory*, 47(4):1302–1327, May 2001.

[49] D. N. C. Tse and S. V. Hanly.
Linear multiuser receivers: Effective interference, effective bandwidth and user capacity.
*IEEE Trans. Inform. Theory*, 45(2):641–657, March 1999.

[50] S. V. Hanly and D. N. C. Tse.
Resource pooling and effective bandwidths in CDMA networks with multiuser receivers and spatial diversity.
*IEEE Trans. Inform. Theory*, 47(4):1328–1351, May 2001.

[51] I. E. Telatar.
Capacity of multi-antenna Gaussian channels.
Technical report, Bell Laboratories, Lucent Technologies, October 1995.

[52] I. E. Telatar.
Capacity of multi-antenna Gaussian channels.
*European Trans. Telecommun.*, 10(6):585–595, November-December 1999.

[53] G. J. Foschini.
Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas.
*Bell Labs Tech. J.*, 1(2):41–59, aug 1996.

[54] G. J. Foschini and M. J. Gans.
On limits of wireless communications in a fading environment when using multiple antennas.
*Wireless Pers. Commun., Kluwer*, 6:311–335, 1998.

[55] D. N. C. Tse and O. Zeitouni.
Linear multiuser receivers in random environments.
*IEEE Trans. Inform. Theory*, 46(1):171–188, January 2000.

[56] Junshan Zhang, E. K. P. Chong, and D. N. C. Tse.
Output MAI distributions of linear MMSE multiuser receivers in DS-CDMA systems.
*IEEE Trans. Inform. Theory*, 47(3):1128–1144, March 2001.

[57] D. Guo, S. Verdú, and L. K. Rasmussen.
Asymptotic normality of linear multiuser receiver outputs.
*IEEE Trans. Inform. Theory*, 48(12):3080–3095, December 2002.

[58] Kiran and D. N. C. Tse.
Effective interference and effective bandwidth of linear multiuser receivers in asynchronous CDMA systems.
*IEEE Trans. Inform. Theory*, 46(4):1426–1447, July 2000.

[59] 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, September 2000.

[60] B. M. Zaidel, S. Shamai, and S. Verdú.
Multicell uplink spectral efficiency of coded DS-CDMA with random signatures.
*IEEE J. Select. Areas Commun.*, 19(8):1556–1569, August 2001.

[61] 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):1293–1311, May 2003.

[62] A. M. Tulino, L. Li, and S. Verdú.
Spectral efficiency of multicarrier CDMA.
*IEEE Trans. Inform. Theory*, 51(2):479–505, February 2005.

[63] R. R Müller.
A random matrix model of communication via antenna arrays.
*IEEE Trans. Inform. Theory*, 48(9):2495–2506, September 2002.

[64] R. R Müller.
On the asymptotic eigenvalue distribution of concatenated vector-valued fading channels.
*IEEE Trans. Inform. Theory*, 48(7):2086–2091, July 2002.

[65] A. M. Tulino, A. Lozano, and S. Verdú.
Impact of antenna correlation on the capacity of multiantenna channels.
*IEEE Trans. Inform. Theory*, 51(7):2491–2509, July 2005.

[66] M. Debbah and R. R. Müller.
MIMO channel modeling and the principle of maximum entropy.
*IEEE Trans. Inform. Theory*, 51(5):1667–1690, May 2005.

[67] R. R. Müller and S. Verdú.
Design and analysis of low-complexity interference mitigation on vector channels.
*IEEE J. Select. Areas Commun.*, 19(8):1429–1441, August 2001.

[68] A. M. Tulino and S. Verdú.
Asymptotic analysis of improved linear receivers for BPSK-CDMA subject to fading.
*IEEE J. Select. Areas Commun.*, 19(8):1544–1555, August 2001.

[69] L. G. F. Trichard, J. S. Evans, and I. B. Collings.
Large system analysis of linear multistage parallel interference cancellation.
*IEEE Trans. Commun.*, 50(11):1778–1786, November 2002.

[70] L. Li, A. M. Tulino, and S. Verdú.
Design of reduced-rank MMSE multiuser detectors using random matrix methods.
*IEEE Trans. Inform. Theory*, 50(6):986–1008, June 2004.

[71] L. Cottatellucci and R. R. Müller.
A systematic approach to multistage detectors in multipath fading channels.
*IEEE Trans. Inform. Theory*, 51(9):3146–3158, September 2005.

[72] M. R. McKay.
*Random Matrix Theory Analysis of Multiple Antenna Communication Systems*.
PhD thesis, The University of Sydney, Sydney, Australia, October 2006.

[73] R. R. Müller.
Applications of large random matrices in communications engineering.
In *Proc. Int. Conf. on Adv. in the Internet, Processing, Systems, and Interdisc. Research*, Sveti Stefan, Montenegro, October 5–11 2003.

[74] T. Tanaka.
A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors.
*IEEE Trans. Inform. Theory*, 48(11):2888–2910, November 2002.

[75] T. Tanaka.
Statistical mechanics of CDMA multiuser demodulation.
*Europhys. Lett.*, 54(4):540–546, 2001.

[76] H. Nishimori.
Comment on 'Statistical mechanics of CDMA multiuser demodulation'.
*Europhys. Lett.*, 57(2):302–303, 2002.

[77] R. G. Gallager.
*Low Density Parity Check Codes*.
M.I.T Press, Cambridge, MA, USA, 1963.

[78] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman.
Improved low-density parity-check codes using irregular graphs.
*IEEE Trans. Inform. Theory*, 47(2):585–598, February 2001.

[79] R. Vicente, D. Saad, and Y. Kabashima.
Statistical physics of irregular low-density parity-check codes.
*J. Phys. A: Math. Gen.*, 33(37):6527, 2000.

[80] Tatsuto Murayama, Yoshiyuki Kabashima, David Saad, and Renato Vicente.
Statistical physics of regular low-density parity-check error-correcting codes.
*Phys. Rev. E*, 62(2):1577–1591, August 2000.

[81] C. Berrou and A. Glavieux.
Near optimum error correcting coding and decoding: Turbo codes.
*IEEE Trans. Commun.*, 44(10):1261–1271, October 1996.

[82] A. Montanari and N. Sourlas.
The statistical mechanics of turbo codes.
*Eur. Phys. J. B*, 18(1):107–119, 2000.

[83] A. Montanari.
The glassy phase of Gallager codes.
*Eur. Phys. J. B*, 23(1):121–136, 2001.

[84] Y. Kabashima and D. Saad.
Statistical mechanics of low-density parity-check codes.
*J. Phys. A: Math. Gen.*, 37:R1–R43, February 2004.

[85] D. Guo and S. Verdú.
Randomly spread CDMA: Asymptotics via statistical physics.
*IEEE Trans. Inform. Theory*, 51(6):1983–2010, June 2005.

[86] D. Guo and S. Verdú.
Multiuser detection and statistical physics.
In V Bhargava, H. V. Poor, V. Tarokh, and S. Yoon, editors, *Communications on Information and Network Security*, chapter 13, pages 229–277. Kluwer Academic Publishers, 2002.

[87] A. Montanari.
Estimating random variables from random sparse observations.
*European Trans. Telecommun.*, 19(9):385–403, 2008.

[88] T. Tanaka and K. Nakamura.
Mean field theory of communication.
*Eur. Phys. J. B*, 64(3):625—631, 2008.

[89] D. Guo and T. Tanaka.
Generic multiuser detection and statistical physics.
In M. Honig, editor, *Advances in Multiuser Detection*. Wiley-IEEE Press, 2009.

[90] P. Ting, C.-K. Wen, J.-C. Chen, and J.-T. Chen.
BER analysis of the optimum multiuser detection with channel mismatch in MC-CDMA systems.
*IEEE J. Select. Areas Commun.*, 24(6):1221–1235, June 2006.

[91] D. Guo.
Performance of multicarrier CDMA in frequency-selective fading via statistical physics.
*IEEE Trans. Inform. Theory*, 52(4):1765–1774, April 2006.

[92] K. Takeuchi, T. Tanaka, and T. Yano.
Asymptotic analysis of general multiuser detectors in MIMO DS-CDMA channels.
*IEEE J. Select. Areas Commun.*, 26(3):486–496, April 2008.

[93] K. Takeuchi and T. Tanaka.
Statistical-mechanics-based analysis of multiuser MIMO channels with linear dispersion codes.
*J. Phys.: Conf. Ser.*, 95(1):012008–1–11, 2008.

[94] R. R. Müller.
Channel capacity and minimum probability of error in large dual antenna array systems with binary modulation.
*IEEE Trans. Signal Processing*, 51(11):2821–2828, November 2003.

[95] K. Takeda, A. Hatabu, and Y. Kabashima.
Statistical mechanical analysis of the linear vector channel in digital communication.
*J. Phys. A: Math. Theor.*, 40(47):14085–14098, 2007.

[96] R. R. Müller, D. Guo, and A. Moustakas.
Vector precoding for wireless MIMO systems and its replica analysis.
*IEEE J. Select. Areas Commun.*, 26(3):530–540, April 2008.

[97] C.-K. Wen, Y.-N. Lee, J.-T. Chen, and P. Ting.
Asymptotic spectral efficiency of MIMO multiple-access wireless systems exploring only channel spatial correlations.
*IEEE Trans. Signal Processing*, 53(6):2059–2073, June 2005.

[98] C.-K. Wen, P. Ting, and J.-T. Chen.
Asymptotic analysis of MIMO wireless systems with spatial correlation at the receiver.
*IEEE Trans. Commun.*, 54(2):349–363, February 2006.

[99] C.-K. Wen and K.-K. Wong.
Asymptotic analysis of spatially correlated MIMO multiple-access channels with arbitrary signaling inputs for joint and separate decoding.
*IEEE Trans. Inform. Theory*, 53(1):252–268, January 2007.

[100] A. L. Moustakas, S. H. Simon, and A. M. Sengupta.

MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large $N$ analysis.
*IEEE Trans. Inform. Theory*, 49(10):2545–2561, October 2003.

[101] A. L. Moustakas and S. H. Simon.
On the outage capacity of correlated multiple-path MIMO channels.
*IEEE Trans. Inform. Theory*, 53(11):3887–3903, November 2007.

[102] F. L. Toninelli.
*Rigorous results for mean field spin glasses: thermodynamic limit and sum rules for the free energy*.
PhD thesis, Scuola Normale Superiore, 2002.

[103] M. Talagrand.
*Spin Glasses: A Challenge for Mathematicians, Cavity and Mean Field Models*.
Springer-Verlag, Berlin Heidelberg, 2003.

[104] M. Talagrand.
*Mean Field Models for Spin Glasses*, volume 1.
Springer-Verlag, Berlin Heidelberg, preliminary version available at http://people.math.jussieu.fr/~talagran/challenge/index.html, 2010.

[105] A. Montanari and D. N. C. Tse.
Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove Tanaka's formula).
In *Proc. IEEE Inform. Theory Workshop*, pages 160–164, Punta del Este, Uruguay, March 13–17 2006.

[106] S. B. Korada and N. Macris.
Tight bounds on the capacity of binary input random CDMA systems.
arXiv:0803.1454v1 [cs.IT], March 2008.

[107] K. Alishahi, F. Marvasti, V. Aref, and P. Pad.
Bounds on the sum capacity of synchronous binary CDMA channels.
arXiv:0806.1659v3 [cs.IT], May 2009.

[108] B. Zaidel, R. R. Müller, A. Moustakas, and R. de Miguel.
Vector precoding for Gaussian MIMO broadcast channels: Impact of replica symmetry breaking.
arXiv:1001.3790v1 [cs.IT], Jan. 2010.

[109] M. K. Varanasi and T. Guess.
Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple-access channel.
In *Proc. Annual Asilomar Conf. Signals, Syst., Comp.*, volume 2, pages 1405–1409, Pacific Grove, CA, USA, November 2–5 1997.

[110] R. R. Müller and W. H Gerstacker.
On the capacity loss due to separation of detection and decoding.
*IEEE Trans. Inform. Theory*, 50(8):1769–1778, August 2004.

[111] T. Tanaka.
Replica analysis of performance loss due to separation of detection and decoding in CDMA channels.
In *Proc. IEEE Int. Symp. Inform. Theory*, pages 2368–2372, Seattle, Washington, USA, July 9–14 2006.

[112] G. Caire, S. Guemghar, A. Roumy, and S. Verdú.
Maximizing the spectral efficiency of coded CDMA under successive decoding.
*IEEE Trans. Inform. Theory*, 50(1):152–164, January 2004.

[113] P. D. Alexander, A. J. Grant, and M. C. Reed.
Iterative detection in code-division multiple-access with error control coding.
*European Trans. Telecommun.*, 9(5):419–525, September/October 1998.

[114] M. C. Reed, C. B. Schlegel, P. D. Alexander, and J. A. Asenstorfer.
Iterative multiuser detection for CDMA with FEC: Near-single-user performance.
*IEEE Trans. Commun.*, 46(12):1693–1699, December 1998.

[115] M. Moher.
An iterative multiuser decoder for near-capacity communications.
*IEEE Trans. Commun.*, 46(7):870–880, July 1998.

[116] P. D. Alexander, M. C. Reed, J. A. Asenstorfer, and C. B. Schlegel.
Iterative multiuser interference reduction: Turbo CDMA.
*IEEE Trans. Commun.*, 47(7):1008–1014, July 1999.

[117] X. Wang and H. V. Poor.

Iterative (turbo) soft interference cancellation and decoding for coded CDMA.
*IEEE Trans. Commun.*, 47(7):1046–1061, July 1999.

[118] H. El Gamal and E. Geraniotis.
Iterative multiuser detection for coded CDMA signals in AWGN and fading channels.
*IEEE J. Select. Areas Commun.*, 18(1):30–41, January 2000.

[119] S. L. Ariyavisitakul.
Turbo space-time processing to improve wireless channel capacity.
*IEEE Trans. Commun.*, 48(8):1347–1359, August 2000.

[120] M. Sellathurai and S. Haykin.
TURBO-BLAST for wireless communications: Theory and experiments.
*IEEE Trans. Signal Processing*, 50(10):2538–2546, October 2002.

[121] C. Douillard, C.B. Michel Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glavieux.
Iterative correction of intersymbol interference: Turbo-equalisation.
*European Trans. Telecommun.*, 6(5):507–511, September 1995.

[122] M. Tüchler, R. Kötter, and A. C. Singer.
Turbo equalisation: Principles and new results.
*IEEE Trans. Commun.*, 50(5):754–767, May 2002.

[123] M. Tüchler, A. C. Singer, and R. Kötter.
Minimum mean squared error equalisation using a priori information.
*IEEE Trans. Signal Processing*, 50(3):673–683, March 2002.

[124] F. R. Kschischang, B. J. Frey, and H. A. Loeliger.
Factor graphs and the sum-product algorithm.
*IEEE Trans. Inform. Theory*, 47(2):498–519, February 2001.

[125] J. Boutros and G. Caire.
Iterative multiuser joint decoding: Unified framework and asymptotic analysis.
*IEEE Trans. Inform. Theory*, 48(7):1772–1793, July 2002.

[126] G. Caire, R. R. Müller, and T. Tanaka.

Iterative multiuser joint decoding: Optimal power allocation and low-complexity implementation.
*IEEE Trans. Inform. Theory*, 50(9):1950–1973, September 2004.

[127] A.P. Worthen and W.E. Stark.
Unified design of iterative receivers using factor graphs.
*IEEE Trans. Inform. Theory*, 47(2):843–849, February 2001.

[128] M. L. Honig and R. Ratasuk.
Large-system performance of iterative multiuser decision-feedback detection.
*IEEE Trans. Commun.*, 51(8):1368–1377, August 2003.

[129] M. V. Burnashev, C. B. Schlegel, W. A. Krzymien, and Z. Shi.
Analysis of the dynamics of iterative interference cancellation in iterative decoding.
*Problems of Information Transmission*, 40(4):297–317, October 2004.

[130] C. Schlegel, Zhenning Shi, and M. Burnashev.
Optimal power/rate allocation and code selection for iterative joint detection of coded random CDMA.
*IEEE Trans. Inform. Theory*, 52(9):4286–4294, September 2006.

[131] D. Truhachev, C. Schlegel, and L. Krzymien.
A two-stage capacity-achieving demodulation/decoding method for random matrix channels.
*IEEE Trans. Inform. Theory*, 55(1):136–146, January 2009.

[132] T. J. Richardson and R. L. Urbanke.
The capacity of low-density parity-check codes under message-passing decoding.
*IEEE Trans. Inform. Theory*, 47(2):599–618, February 2001.

[133] S-Y. Chung, T. J. Richardson, and R. L. Urbanke.
Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation.
*IEEE Trans. Inform. Theory*, 47(2):657–670, February 2001.

[134] A. Sanderovich, M. Peleg, and S. Shamai.
LDPC coded MIMO multiple access with iterative joint decoding.
*IEEE Trans. Inform. Theory*, 51(4):1437–1450, April 2005.

[135] B. Hu, I. Land, L. Rasmussen, R. Piton, and B. H. Fleury.
A divergence minimization approach to joint multiuser decoding for coded CDMA.
*IEEE J. Select. Areas Commun.*, 26(3):432, April 2008.

[136] D. D. Lin and T. J. Lim.
Bit-level equalization and soft detection for Gray-coded multilevel modulation.
*IEEE Trans. Inform. Theory*, 54(10):4731–4742, October 2008.

[137] H. Li and H. V. Poor.
Impact of channel estimation errors on multiuser detection via the replica method.
*EURASIP J. Wirel. Commun. Netw.*, 2005(2):175–186, April 2005.

[138] H. Li, S. M. Betz, and H. V. Poor.
Performance analysis of iterative channel estimation and multiuser detection in multipath DS-CDMA channels.
*IEEE Trans. Signal Processing*, 55(5):1981–1993, May 2007.

[139] M. Médard.
The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel.
*IEEE Trans. Inform. Theory*, 46(3):933–946, May 2000.

[140] B. Hassibi and B. M. Hochwald.
How much training is needed in multiple-antenna wireless links?
*IEEE Trans. Inform. Theory*, 49(4):951–963, April 2003.

[141] A. Lapidoth and S. Shamai.
Fading channels: How perfect need 'perfect side information' be?
*IEEE Trans. Inform. Theory*, 48(5):1118–1134, May 2002.

[142] H. Weingarten, Y. Steinberg, and S. Shamai.
Gaussian codes and weighted nearest neighbor decoding in fading multiple-antenna channels.
*IEEE Trans. Inform. Theory*, 50(8):1665–1686, August 2004.

[143] P. D. Alexander and A. J. Grant.
Iterative channel and information sequence estimation in CDMA.

In *Proc. IEEE Int. Symp. Spread Spectrum Techniques and Applications*, volume 2, pages 593–597, New Jersey, USA, September 6–8 2000.

[144] G. Caire, A. Tulino, and E. Biglieri.
Iterative multiuser joint decoding and parameter estimation: A factor-graph approach.
In *Proc. IEEE Inform. Theory Workshop*, pages 36–38, Cairns, Australia, September 2–7 2001.

[145] J. Thomas and E. Geraniotis.
Soft iterative multisensor multiuser detection in coded dispersive CDMA wireless channels.
*IEEE J. Select. Areas Commun.*, 19(7):1334–1351, July 2001.

[146] M. Kobayashi, J. Boutros, and G. Caire.
Successive interference cancellation with SISO decoding and EM channel estimation.
*IEEE J. Select. Areas Commun.*, 19(8):1450–1460, August 2001.

[147] M. C. Valenti and B. D. Woerner.
Iterative channel estimation and decoding of pilot symbol assisted turbo codes over flat-fading channels.
*IEEE J. Select. Areas Commun.*, 19(9):1697–1705, September 2001.

[148] M. Lončar, R. R. Müller, J. Wehinger, C. Mecklenbräuker, and T. Abe.
Iterative channel estimation and data detection in frequency-selective fading MIMO channels.
*European Trans. Telecommun.*, 15(5):459–470, September/October 2004.

[149] J. Wehinger and C. F. Mecklenbräuker.
Iterative CDMA multiuser receiver with soft decision-directed channel estimation.
*IEEE Trans. Signal Processing*, 54(10):3922–3934, October 2006.

[150] M. Nicoli, S. Ferrara, and U. Spagnolini.
Soft-iterative channel estimation: Methods and performance analysis.
*IEEE Trans. Signal Processing*, 55(6):2993–3006, June 2007.

[151] D. D. Lin and Teng Joon Lim.
A variational inference framework for soft-in soft-out detection in multiple-access channels.

*IEEE Trans. Inform. Theory*, 55(5):2345–2364, May 2009.

[152] E. Panayırcı, H. Doğan, H. A. Çirpan, A. Kocian, and B. H. Fleury.
Iterative joint data detection and channel estimation for uplink MC-CDMA systems in the presence of frequency selective channels.
*Physical Communication*, 2009.
in press, available on-line: http://www.sciencedirect.com/science/article/B8JDV-4WK441J-1/2/cffb2fa8

[153] A. Mantravadi, V. V. Veeravalli, and H. Viswanathan.
Spectral efficiency of MIMO multiaccess systems with single-user decoding.
*IEEE J. Select. Areas Commun.*, 21(3):382–394, April 2003.

[154] L. Cottatellucci and R. R. Müller.
CDMA systems with correlated spatial diversity: A generalized resource pooling result.
*IEEE Trans. Inform. Theory*, 53(3):1116–1136, March 2007.

[155] H. Li and H. V. Poor.
Asymptotic analysis of outage region in CDMA MIMO systems.
*IEEE Trans. Inform. Theory*, 55(3):1206–1217, March 2009.

[156] F. D. Neeser and J. L. Massey.
Proper complex random processes with applications to information theory.
*IEEE Trans. Inform. Theory*, 39(4):1293–1302, July 1993.

[157] L. H. Ozarow, S. Shamai, and A. D. Wyner.
Information theoretic considerations for cellular mobile radio.
*IEEE Trans. Veh. Technol.*, 43(2):359–378, May 1994.

[158] E. Biglieri, J. Proakis, and S. Shamai.
Fading channels: Information-theoretic and communications aspects.
*IEEE Trans. Inform. Theory*, 44(6):2619–2692, October 1998.

[159] A. Nordio and G. Taricco.
Linear receivers for the multiple-input multiple-output multiple-access channel.
*IEEE Trans. Commun.*, 54(8):1446–1456, August 2006.

[160] K. Takeuchi.
*Statistical Mechanical Informatics on MIMO DS-CDMA Systems: Design of Spreading Schemes and Performance of Multiuser Decoding*.

PhD thesis, Kyoto University, Kyoto, Japan, 2009.

[161] A. P. Dawid.
Some matrix-variate distribution theory: Notational considerations and a Bayesian application.
*Biometrika*, 68(1):265–274, 1981.

[162] C. N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela.
Capacity scaling in MIMO wireless systems under correlated fading.
*IEEE Trans. Inform. Theory*, 48(3):637–650, March 2002.

[163] J. P. Kermoal, L. Schumacher, K. Pedersen, P. E. Mogensen, and F. Frederiksen.
A stochastic MIMO radio channel model with experimental validation.
*IEEE J. Select. Areas Commun.*, 20(6):1211–1226, August 2002.

[164] D. Chizhik, J. Ling, P. W. Wolnianski, R. A. Valenzuela, N. Costa, and K. Huber.
Multiple-input multiple output measurements and modeling in Manhattan.
*IEEE J. Select. Areas Commun.*, 21(3):321–331, April 2003.

[165] G. Caire, G. Taricco, and E. Biglieri.
Bit-interleaved coded modulation.
*IEEE Trans. Inform. Theory*, 44:927–946, May 1998.

[166] A. Guillén i Fàbregas, A. Martinez, and G. Caire.
Bit-interleaved coded modulation.
*Foundations and Trends in Communications and Information Theory*, 5(1-2):1–153, 2008.

[167] A. Martinez, A. Guillén i Fàbregas, G. Caire, and F. M. J. Willems.
Bit-interleaved coded modulation revisited: A mismatched decoding perspective.
*IEEE Trans. Inform. Theory*, 55(6):2756–2765, June 2009.

[168] A. Martinez, A. Guillén i Fàbregas, and G. Caire.
Error probability analysis of bit-interleaved coded modulation.
*IEEE Trans. Inform. Theory*, 52(1):262–271, January 2006.

[169] E. Zehavi.
8-PSK trellis codes for a Rayleigh channel.

*IEEE Trans. Commun.*, 40(5):873–884, May 1992.

[170] R. G. Gallager.
*Information Theory and Reliable Communication*.
John Wiley and Sons, New York, 1968.

[171] W. He and C. N. Georghiades.
Computing the capacity of a MIMO fading channel under PSK signaling.
*IEEE Trans. Inform. Theory*, 51(5):1794–1803, May 2005.

[172] K. Takeuchi, R. R. Müller, M. Vehkaperä, and T. Tanaka.
Practical signaling with vanishing pilot-energy for large noncoherent block-fading MIMO channels.
In *Proc. IEEE Int. Symp. Inform. Theory*, pages 759–763, Seoul, Korea, 2009.

[173] H. El Gamal and A. R. Hammons, Jr.
Analyzing the turbo decoder using the Gaussian approximation.
*IEEE Trans. Inform. Theory*, 47(2):671–686, February 2001.

[174] S. ten Brink.
Convergence behavior of iteratively decoded parallel concatenated codes.
*IEEE Trans. Commun.*, 49(10):1727–1737, 2001.

[175] D. Sherrington and S. Kirkpatrick.
Solvable model of a spin-glass.
*Phys. Rev. Lett.*, 35(26):1792–1796, December 1975.

[176] S. Kirkpatrick and D. Sherrington.
Infinite-ranged models of spin-glasses.
*Phys. Rev. B*, 17(11):4384–4403, June 1978.

[177] G. Parisi.
A sequence of approximated solutions to the S-K model for spin glasses.
*J. Phys. A: Math. Gen.*, 13(4):L115–L121, 1980.

[178] R. de Miguel.
*On the analysis and design of wireless communication systems using tools from statistical physics*.
PhD thesis, Norwegian University of Science and Technology, Trondheim, Norway, 2009.

[179] R. S. Ellis.
*Entropy, Large Deviations, and Statistical Mechanics*.
Springer-Verlag, New York, 1985.

[180] A. Dembo and O. Zeitounio.
*Large Deviations Techniques and Applications*.
Springer-Verlag, New York, 1998.

[181] M. Talagrand.
The Parisi formula.
*Annals of Math*, 163(1):221–263, 2006.

[182] K. B. Petersen and M. S. Pedersen.
The matrix cookbook, October 2008.
http://www2.imm.dtu.dk/pubdb/p.php?3274 [Version 20081110].

[183] K. Takeuchi, M. Vehkaperä, T. Tanaka, and R. R. Müller.
Replica analysis of general multiuser detection in MIMO DS-CDMA channels with imperfect CSI.
In *Proc. IEEE Int. Symp. Inform. Theory*, pages 514–518, Toronto, ON, Canada, July 6–11 2008.

[184] V. A. Aalo.
Bit-error rate of binary digital modulation schemes in generalized gamma fading channels.
*IEEE Trans. Commun.*, 43(8):2360–2369, August 1995.

[185] R. K. Mallik.
On multivariate Rayleigh and exponential distributions.
*IEEE Trans. Inform. Theory*, 49(6):1499–1515, 2003.

[186] D. A. Harville.
*Matrix Algebra From a Statistician's Perspective*.
Springer-Verlag, New York, 1997.

[187] Y.C. Eldar and A.M. Chan.
On the asymptotic performance of the decorrelator.
*IEEE Trans. Inform. Theory*, 49(9):2309–2313, September 2003.

[188] U. Fincke and M. Pohst.

Improved methods for calculating vectors of short length in a lattice, including a complexity analysis.
*Math. Comput.*, 44(5):463–471, May 1985.

[189] E. Viterbo and J. Boutros.
A universal lattice code decoder for fading channels.
*IEEE Trans. Inform. Theory*, 45(5):1639–1642, July 1999.

[190] B. Hochwald and S. Brink.
Achieving near-capacity on a multiple-antenna channel.
*IEEE Trans. Commun.*, 51(3):389–399, March 2003.

[191] M. O. Damen, H. El Gamal, and G. Caire.
On maximum-likelihood detection and the search for the closest lattice point.
*IEEE Trans. Inform. Theory*, 49(10):2389–2402, October 2003.

[192] A. D. Murugan, H. El Gamal, M. O. Damen, and G. Caire.
A unified framework for tree search decoding: rediscovering the sequential decoder.
*IEEE Trans. Inform. Theory*, 52(3):933–953, March 2006.

[193] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv.
Optimal decoding of linear codes for minimizing symbol error rate.
*IEEE Trans. Inform. Theory*, 20(2):284–287, March 1974.

[194] J. Hagenauer, E. Offer, and L. Papke.
Iterative decoding of binary block and convolutional codes.
*IEEE Trans. Inform. Theory*, 42(2):429–445, March 1996.

[195] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara.
A soft-input soft-output maximum APP module for iterative decoding of concatenated codes.
*IEEE Commun. Lett.*, 1(1):22–24, January 1997.

[196] G. Colavolpe, G. Ferrari, and R. Raheli.
Extrinsic information in iterative decoding: A unified view.
*IEEE Trans. Commun.*, 49(12):2088–2094, December 2001.

[197] H. V. Poor and S. Verdú.
Probability of error in MMSE multiuser detection.
*IEEE Trans. Inform. Theory*, 43(3):858–871, May 1997.

[198] K. Larsen.
Short convolutional codes with maximal free distance for rates $1/2$, $1/3$, and $1/4$.
*IEEE Trans. Inform. Theory*, 19(3):371–372, May 1973.

[199] K. Takeuchi, M. Vehkaperä, T. Tanaka, and R. R. Müller.
Large-system analysis of joint channel and data estimation for MIMO DS-CDMA systems.
arXiv:1002.4470v1 [cs.IT], Feb. 2010.

[200] K. Takeuchi, M. Vehkaperä, T. Tanaka, and R. R. Müller.
Asymptotic performance bounds of joint channel estimation and multiuser detection in frequency-selective fading DS-CDMA channels.
In *Proc. IEEE Int. Symp. Inform. Theory and its Applications*, Auckland, New Zealand, December 7–10 2008.

[201] R. A. Horn and C. R. Johnson.
*Matrix Analysis*.
Cambridge University Press, Cambridge, 1985.

[202] A. Hjørungnes and D. Gesbert.
Complex-valued matrix differentiation: Techniques and key results.
*IEEE Trans. Signal Processing*, 55(6):2740–2746, June 2007.

[203] K. Kreutz-Delgado.
The complex gradient operator and the $\mathbb{CR}$-calculus.
arXiv:0906.4835v1 [math.OC], June 2009.