EURASIP Journal on Applied Signal Processing

Mimo Communications and Signal Processing

Guest Editors: Sergio Barbarossa, Constantinos Papadias, H. Vincent Poor, and Xiaowen Wang

Mimo Communications and Signal Processing

EURASIP Journal on Applied Signal Processing

Mimo Communications and Signal Processing

Guest Editors: Sergio Barbarossa, Constantinos Papadias, H. Vincent Poor, and Xiaowen Wang

Contents

EURASIP Journal on Applied Signal Processing 2004:5, 587–590

Editorial

Sergio Barbarossa
Department of Information and Communication, University of Rome "La Sapienza," 00184 Rome, Italy
Email: sergio@infocom.ing.uniroma1.it

Constantinos Papadias
Bell Labs, Lucent Technologies, 791 Holmdel-Keyport Road, Holmdel, NJ 07733, USA
Email: papadias@bell-labs.com

H. Vincent Poor
Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA
Email: poor@princeton.edu

Xiaowen Wang
Agere Systems, Allentown, PA 18109, USA
Email: xiaowenw@agere.com

The topic of multiple-input multiple-output (MIMO) systems is one that has attracted a significant amount of attention in the research community over the past decade or so. MIMO systems refer to wireless systems that are equipped with multiple antenna elements on either side of a communication link. Propelled by the startling discovery in the mid 1990's that the capacity of MIMO systems grows roughly proportionally with the minimum number of antenna elements on each side of the wireless link, the field has undergone an explosive growth in both the academic and the industrial communities that has led to many further important advances. These advances have brought about not only the definition of new subareas of focused research, but also, equally importantly, a reconsideration of older techniques and a cross-fertilization of ideas from several other overlapping fields.

One of the research areas that has both affected strongly MIMO systems and has been equally affected by them is that of signal processing, as many of the developed/demonstrated MIMO transceiver architectures are based on advanced signal processing techniques. On the transmitter side, one can view most space-time coding/spatial multiplexing techniques as solving a problem of space-time signal design. On the receiver side, various flavors of multiuser detectors, space-time decoders, and related techniques for MIMO channel estimation and tracking (e.g., including blind/semiblind processing) are also typically derived in a signal processing framework.

More recent research on MIMO systems has started to focus on new areas of interest. At the link level, such areas are the handling of cochannel (e.g., in-cell and out-of-cell) interference; the development of precoding techniques at the transmitter to preempt adverse channel effects; and the design and use of efficient receiver-to-transmitter feedback mechanisms to improve the link throughput. In parallel, many studies have focused on the application of such MIMO techniques to specific transmission formats (dictated by different air interfaces) such as CDMA, OFDM, and so forth. Moving up the protocol stack, the interaction of MIMO techniques with MAC layer procedures such as adaptive retransmission and scheduling is an area that has started producing important know-how, especially regarding the suitability of MIMO techniques in high-speed data systems. Moving beyond wireless links, architecting an entire wireless network that uses MIMO connections poses a number of important questions, both at a fundamental level (e.g., MIMO network capacity) and at a practical level (e.g., MIMO network design).

With all of the above in mind, this special issue aims at giving a well-rounded snapshot of recent advances that cover most of these topics, with a special emphasis on signal processing methodologies as a design tool. As progress in the field is both rapidly emerging and voluminous, it has clearly not been our intent to provide an exhaustive coverage of all MIMO topics but rather a good selection of recent studies that are indicative of the progress in the field.

The papers included in this special issue address a broad range of issues arising in the development and application of MIMO techniques. These contributions range from general space-time coding and processing techniques and analytical methodologies to specific implementation issues arising in particular wireless standards and environments and to fundamentals of wireless MIMO networks. Among other topics, they touch upon the areas of transmitter and receiver design, blind and training-based techniques, link-level and system-level studies, open- and closed-loop systems, physical layer and higher layer issues, and wireless LAN and cellular applications.

The specific contributions of the papers in this issue are summarized in the following paragraphs.

Invited paper

In their invited paper, Jafar, Foschini, and Goldsmith present an in-depth analysis of the so-called "PhantomNet" wireless network concept. In such a network, the best possible service is provided to new users joining the network without affecting existing users. The problem is addressed in its full generality, that is, assuming multiple cells, users, and antennas, and results are obtained for both uplink and downlink communication. Optimality is sought in terms of the multiuser capacity region. This leads to a high degree of generality of the presented results and solutions. Furthermore, despite the inherent differences between the two directions of communication (and the resulting differences between the corresponding solutions), the authors demonstrate a remarkable symmetry between the uplink and downlink problems.

Channel estimation and multiuser detection in MIMO systems

In the first paper of this section, J. Du and Y. Li study the problem of channel estimation for D-BLAST OFDM systems. The authors propose a layerwise channel estimation algorithm that takes advantage of the D-BLAST structure. Further performance improvements are realized by introducing a subspace tracking scheme.

In the next paper, Buzzi, Grossi, and Lops study the problem of blind multiuser detection in asynchronous DS-CDMA systems equipped with multiple antennas. Several novel blind schemes are proposed and their performance is evaluated, showing their multiple access interference suppression capability, despite the absence of channel state information.

Another blind detection scheme that is specifically tailored to space-time differentially encoded systems is presented in the paper by Zhang and Ilow. Their proposed receiver algorithm is based on constant modulus characteristics of signaling and it is suitable for a rich multipath environment. The scheme requires no channel estimation and can work with small numbers of signal samples.

In their paper, Y. Du and Chan examine a technique for speeding up the search for an optimal multiuser detection solution using a genetic algorithm. The authors first study the objective function of the genetic algorithms. Then they propose two detectors to generate the seed chromosome of the initial population. Their results show that the proposed scheme not only reduces the computational complexity of finding the detector, but also improves performance.

MIMO systems, space-time coding, and beamforming

In the first paper of this section, C. Li and Xiaodong Wang compare the performance of three well-known MIMO techniques: BLAST, space-time block coding (STBC), and linear precoding/coding (i.e., beamforming) in the context of WCDMA. The authors study the signal-to-noise properties analytically, and the bit error rate performance via simulations. They also consider a subspace method for implementing the linear precoding method (which requires channel knowledge at the transmitter). The authors evaluate the trade-offs between BLAST and STBC in terms of data-rate and diversity in this situation (see also the following paper in this section) and demonstrate that subspace-based beamforming can be effectively realized in WCDMA systems.

In the next paper, Mecklenbräuker and Rupp consider a new STBC scheme that extends the well-known Alamouti codes to the situation in which the number of transmit antennas is an arbitrary power of two. Further solutions for arbitrary even numbers of transmit antennas are also presented, which offer improved orthogonalization properties while preserving high diversity. The authors also consider schemes that trade off the properties of Alamouti and BLAST-type systems (see also Li's and Wang's paper above) to achieve a continuous trade-off between quality of service and data rate. The appropriate trade-off can be selected using only the number of transmit antennas. Implications of these techniques for UMTS are also discussed.

MIMO systems and interference

In his paper, Blum studies the problem of maximum system mutual information in MIMO systems that employ antenna selection in the presence of interference. This leads to optimal signaling covariance matrices for the interesting case of limited channel feedback required for antenna selection.

The paper by Song and Blostein studies the effect of colored space-time interference on MIMO systems, emphasizing the problems of channel estimation, data detection, and interference correlation estimation. The focus is on the case of one dominant interferer and the quantification of its impact on the performance of a generalized BLAST ordered data detection algorithm. The authors show that exploiting the interference's spatio-temporal nature can result in important gains.

MIMO techniques in current/emerging air interfaces

In the first paper of this section, J. Liu and J. Li study some practical issues arising in the application of MIMO OFDM

to high-rate wireless LAN systems. The authors propose signaling and corresponding synchronization, channel estimation, and detection schemes that are backward compatible with the existing 802.11a standard. They also propose the use of a BLAST-type data transmission scheme and a simple LS-based soft detector to reduce the complexity of the receiver.

In the next paper, Hansen, Affes, and Mermelstein revisit the problem of multiuser detection in CDMA networks. The authors apply an interference subspace rejection technique to the downlink of networks in which the spreading factors or modulation used by the interferer may not be known. The schemes proposed in the paper require no prior knowledge of these factors. A new code allocation scheme is also proposed to reduce the complexity of the proposed interference cancellation schemes.

The paper by González-López, Míguez, and Castedo presents a maximum likelihood channel estimation scheme that is suitable for turbo equalization in a space-time coded system. The authors apply their scheme to GSM-based transmission in a subway tunnel. Their experiment shows a significant reduction in the required training sequence length.

In the final paper of this section, Leus, Petré, and Moonen propose novel transmit diversity and corresponding space-time chip equalization techniques for DS-CDMA systems. Their proposed scheme is shown to achieve both maximal antenna diversity and maximal multipath diversity.

Resource allocation and feedback in multiple antenna systems

In the first paper of this section, Han, Farrokhi, and K. J. Ray Liu revisit the problem of jointly optimizing power control and beamforming to minimize the cochannel interference. The authors optimize the bit error rate directly in calculating the power and beamforming vector. Both the power control and beamforming algorithms are updated iteratively and are shown to converge.

In their paper, Chung, Lozano, Huang, Sutivong, and Cioffi study closed-loop MIMO systems. In order to achieve the closed-loop capacity, the authors propose to use a low rate feedback channel to provide rate and power information to the transmitter. Two joint rate and power allocation schemes are proposed and studied by the authors. Their results show that the performance loss due to the quantization of power is marginal, and that the MIMO system demonstrates an average rate close to capacity with the low-rate feedback channel and strong coding scheme.

Higher layer issues in MIMO systems

In the final paper of the special issue, Zheng, Lozano, and Haleem propose an ARQ scheme based on the BLAST system. The authors suggest the use of separate ARQ for each layer of the BLAST transmission. This multiple ARQ structure not only improves the throughput, but also facilitates the interference cancellation.

We believe that the included papers present an excellent sampling of state-of-the-art research in the field of MIMO communications and signal processing. We would like to thank all of the authors for their timely contributions and we anticipate that these papers will make this special issue a useful reference that will act as a catalyst for further exciting research in the field of MIMO systems.

Sergio Barbarossa
Constantinos Papadias
H. Vincent Poor
Xiaowen Wang

Sergio Barbarossa graduated in 1984 and received his Ph.D. degree in 1989 from the University of Rome "La Sapienza," Italy. From 1984 to 1986, he was a Radar System Engineer at Selenia. In 1988, he was at the Environmental Research Institute of Michigan (ERIM), Ann Arbor, USA. From 1989 to 1991, he was with the University of Perugia and in 1991, he joined the University of Rome "La Sapienza," where he is now a Full Professor. Since 1997, he is a Member of the IEEE Signal Processing for Communications Technical Committee. From 1998 to 2000, he served as an Associate Editor for the IEEE Transactions on Signal Processing. He coauthored a paper that received the 2000 IEEE Best Paper Award in the Signal Processing for Communications area. He has been the General Chairman of the SPAWC 2003 (Rome, 2003). He has held visiting positions at the University of Virginia in 1995 and 1997, University of Minnesota in 1999, and Polytechnic University of Catalunya, Spain, in 2002. He is the author of a research monograph on Multiple Antenna Systems (Artech House, 2004). He is the scientific responsible, for his University, of two IST European projects on space-time coding and multihop networks. His current research interests lie in the area of self-organizing networks, random graphs, and distributed space-time coding.

Constantinos Papadias was born in Athens, Greece, in 1969. He received his Diploma of electrical engineering from the National Technical University of Athens (NTUA) in 1991 and the Ph.D. degree in signal processing (highest honors) from the Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France, in 1995. From 1992 to 1995, he was Teaching and Research Assistant at the Mobile Communications Department, Eurécom, France. From 1995 to 1997, he was a Postdoctoral Researcher at Stanford University's Smart Antennas Research Group. In November 1997, he joined the Wireless Research Laboratory of Bell Labs, Lucent Technologies, Holmdel, NJ, as member of technical staff. He is now Technical Manager in Global Wireless Systems Research Department, Bell Lab's, overseeing several research projects, with an emphasis on space-time and MIMO systems. He has authored several papers, patents, and standards contributions on these topics and he recently received the IEEE Signal Processing Society's 2003 Young Author Best Paper Award. He is a Member of the Signal Processing for Communications Technical Committee of the IEEE Signal Processing Society and Associate Editor for the IEEE Transactions on Signal Processing. Dr. Papadias is a Senior Member of IEEE and a Member of the Technical Chamber of Greece.

H. Vincent Poor received the Ph.D. degree in EECS from Princeton University in 1977. From 1977 until 1990, he was at the University of Illinois at Urbana-Champaign. Since 1990, he has been at Princeton University, where he is the George Van Ness Lothrop Professor in Engineering. Dr. Poor's research interests are in the areas of wireless networks, advanced signal processing, and related fields. He is the author of more than 500 publications in these areas, including the recent book *Wireless Communication Systems: Advanced Techniques for Signal Reception* (Prentice-Hall, Upper Saddle River, NJ, 2004). Dr. Poor is a Member of the U. S. National Academy of Engineering, and is a Fellow of the IEEE, the Institute of Mathematical Statistics, and other organizations. In 1990, he served as President of the IEEE Information Theory Society, and in 1991–1992, he was a Member of the IEEE Board of Directors. Among his recent honors are the IEEE Graduate Teaching Award in 2001, the Joint Paper Award of the IEEE Communications and Information Theory Societies in 2001, the NSF Director's Award for Distinguished Teaching Scholars in 2002, and a Guggenheim Fellowship in 2002–2003.

Xiaowen Wang received her B.S. degree from the Department of Electronics Engineering, Tsinghua University, Beijing, China in 1993, and the M.S. and Ph.D. degrees from the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, in 1999 and 2000, respectively. From 1993 to 1996, Dr. Wang was a Teaching Assistant at Tsinghua University, Beijing, China. From 1996 to 2000, she was a Research Assistant at the University of Maryland, College Park, MD. Since 2000, she has been with the Wireless Systems Research Department, Agere Systems (formerly Bell Labs, Lucent Technologies, Microelectronics). Her research interests include adaptive digital signal processing, wireless communications, and networking. Dr. Wang was ranked the first among the class of Department of Electronics Engineering for her B.S. degree from Tsinghua University in 1993, and was the recipient of the Graduate School Fellowship from University of Maryland.

EURASIP Journal on Applied Signal Processing 2004:5, 591–604

PhantomNet: Exploring Optimal Multicellular Multiple Antenna Systems

Syed A. Jafar
Electrical Engineering and Computer Science, University of California, Irvine, Irvine, CA 92697-2625, USA
Email: syed@uci.edu

Gerard J. Foschini
Bell Laboratories, Lucent Technologies, 791 Holmdel-Keyport Road, Holmdel, NJ 07733, USA
Email: gjf@lucent.com

Andrea J. Goldsmith
Wireless Systems Laboratory, Stanford University, Stanford, CA 94305-9505, USA
Email: andrea@ee.stanford.edu

Received 20 December 2002; Revised 11 August 2003

We present a network framework for evaluating the theoretical performance limits of wireless data communication. We address the problem of providing the best possible service to new users joining the system without affecting existing users. Since, interference-wise, new users are required to be invisible to existing users, the network is dubbed PhantomNet. The novelty is the generality obtained in this context. Namely, we can deal with multiple users, multiple antennas, and multiple cells on both the uplink and the downlink. The solution for the uplink is effectively the same as for a single cell system since all the base stations (BSs) simply amount to one composite BS with centralized processing. The optimum strategy, following directly from known results, is successive decoding (SD), where the new user is decoded before the existing users so that the new users' signal can be subtracted out to meet its invisibility requirement. Only the BS needs to modify its decoding scheme in the handling of new users, since existing users continue to transmit their data exactly as they did before the new arrivals. The downlink, even with the BSs operating as one composite BS, is more problematic. With multiple antennas at each BS site, the optimal coding scheme and the capacity region for this channel are unsolved problems. SD and dirty paper (DP) are two schemes previously reported to achieve capacity in special cases. For PhantomNet, we show that DP coding at the BS is equal to or better than SD. The new user is encoded before the existing users so that the interference caused by his signal to existing users is known to the transmitter. Thus the BS modifies its encoding scheme to accommodate new users so that existing users continue to operate as before: they achieve the same rates as before and they decode their signal in precisely the same way as before. The solutions for the uplink and the downlink are particularly interesting in the way they exhibit a remarkable simplicity and an unmistakable, near-perfect, up-down symmetry.

Keywords and phrases: channel capacity, dirty paper coding, duality, broadcast channel, successive decoding, multiple-input multiple-output systems.

1. INTRODUCTION

The rapid growth of cellular networks and the anticipation of ever increasing demand for higher data rates have expanded the scope of wireless research from single user, and single cell, and single antenna systems to multiuser multicellular systems employing multiple antennas. A traditional way of handling the multiantenna, multiuser, and multicellular system has been to reduce it to a single antenna, single user, and single cell system by orthogonally splitting the channel among the users in time/frequency/code/space, employing the base station antennas for sectoring/beamforming, and treating cochannel interference from other cells as noise. Moreover, since early wireless networks have been designed primarily for voice traffic, rate adaptation was not considered. This constrained approach may be simpler, but quite often it leads to suboptimal strategies. In order to estimate the absolute performance limits of these multidimensional systems, we need to explicitly account for the presence of multiple users, multiple antennas, and multiple cells on both the uplink and the downlink.

In this paper, where wireless data communication is

highlighted, the focus is on finding the best transmit strategy. Due to the presence of a multiplicity of contending users, the best transmit strategy is not as straightforward as for a single-user system. Assigning limited communication resources to effect the best transmit strategy is particularly relevant for handling delay tolerant data traffic since helping some users typically amounts to slowing others. The best strategy, of course, depends on the priorities assigned to each user. Given the prioritization, say, for example, first-come-first-served (FCFS), we find here the optimum communication means under different criteria.

Although we will proceed with the FCFS prioritization in our presentation, our results hold for other means of prioritizing such as last-come-first-served, random ordering, or any scheme that predetermines an ordering among users.

We consider both the uplink and the downlink of a multiuser multicellular system using multiple antennas at both ends. We consider a system that evolves in time with new users entering the system and old users leaving the system. Using FCFS, our objective is to provide the best service possible to the new users as they enter the system, without penalizing the users already in the system. Thus each user in the system has a higher priority than the users that come after him. Subsequent users are served under the requirement that the previous ones are not affected: interference-wise, new users must be invisible to exiting users. Since for both the uplink and the downlink only earlier entrants interfere while later entrants are invisible, the network is dubbed PhantomNet. The strategies that affect this invisibility will be seen to be successive decoding (SD) for the uplink (a form of multiuser detection) and dirty paper (DP) coding for the downlink. In our network context, these strategies are particularly interesting both because of their simplicity as well as the unmistakable symmetry evident between uplink-downlink operation. Just how resources like base stations, bandwidth, spatial modes, and power are used is not preordained. Rather, under the FCFS regime, the network can self-organize the deployment of these communication resources.

The FCFS model assigns lower priority to new users. However, as previous users complete their transmission, the user moves up on the priority scale. So users that stay in the system longer tend to experience a better average service. In other words, shorter messages experience a lower average rate, while longer messages experience a higher average rate. It is therefore reasonable to expect that the FCFS scheduling algorithm would make the time required to transmit to different users' messages more equal.[1]

[1]If one chooses instead a last-come-first-served model, short messages would see higher average rates, and long messages would see lower average rates. Thus last-come-first-served scheduling would make the time required to transmit different users' messages more disparate. The average number of simultaneously active users would reflect the average interference seen by the users. Overall, the choice of the scheduling algorithm for a system will depend on such criteria.

Our scope here is limited to the presentation of theoretical findings. These findings provide a tractable framework in which performance of multicellular, multiuser, and multiantenna wireless networks can be numerically evaluated through simulation. Information theoretic optimization is at the core of our approach. Simulation results with DP coding presented in [1] complement this work.

2. SYSTEM MODEL

Although we are ultimately interested in a multicellular system, for simplicity, we start with a single base station. Multiple base stations will be addressed in Section 7.

2.1. Uplink

The uplink is characterized by the following equation:

$$Y = \sum_{i=1}^{K} H_i X_i + N, \tag{1}$$

where Y is the received vector at the base station, K is the number of users currently active in the system, H_i is the flat-fading matrix channel of user i, and N is the additive white Gaussian noise (AWGN) vector at the base station.

Without loss of generality, we assume that the users are indexed by the order in which they arrive. So user 1 is the first user in the system, while user K is the last user to join the system. The users are subject to transmit power constraints given by

$$\text{trace}\left[E\left[X_i X_i^{\dagger}\right]\right] \leq P_i, \quad 1 \leq i \leq K. \tag{2}$$

Note that there is no data coordination between users, so the X_i are independent.

2.2. Downlink

Finding the optimal transmit strategy for the downlink with multiple antennas is a hard problem. This is because the multiple antenna downlink channel is a nondegraded broadcast channel and its capacity region is a long standing unsolved problem in information theory [2]. The optimal coding strategy for the multiple antenna downlink is therefore unknown. The special cases of the AWGN broadcast channel where the optimal coding strategy is known include the degraded broadcast channel (single transmit antenna at the BS), and the recently solved sum rate capacity of multiple user vector broadcast channel with multiple transmit antennas at the BS and at each of the mobiles [3, 4, 5, 6, 7]. While SD achieves capacity in the first case, DP coding based on the results of [8] achieves capacity in the latter. DP coding can also be shown to achieve capacity for the degraded AWGN broadcast channel. Note that for all these cases where the capacity is known, it is achieved with SD or DP coding and with Gaussian codebooks. For this reason, in this paper, we will restrict our downlink transmit strategies to these

two coding schemes and we will assume that Gaussian codebooks are used. These assumptions may not be restrictive at all in case the conjectures about the optimality of Gaussian codebooks on the downlink can be established [9, 10]. Thus, our downlink model is given by the following equation:

$$Y_i = H_i \sum_{j=1}^{K} X_j + N_i, \tag{3}$$

where Y_i, X_i, H_i, and N_i are the output vector, the input vector, the channel matrix, and the AWGN vector for user i. For both SD and DP coding strategies, the input vectors corresponding to different users are independent. As in the uplink model described earlier, the downlink model also assumes that the users are indexed by the order in which they arrive. Further, the power in each user's input vector is given by

$$\text{trace}\left[E\left[X_i X_i^{\dagger}\right]\right] \leq P_i, \quad 1 \leq i \leq K. \tag{4}$$

We would also like to point out that a "ranked known interference" scheme based on the results of [3] was used in [11] to minimize the delay in a multiuser multicellular system with multiple antennas at the base station and a single receive antenna at each mobile. While the scheme itself is suboptimal and limited in scope to a single receive antenna at each mobile, it is another example of a simple way to perform resource allocation on the downlink. The results of [11] are interesting and complement this work.

Unlike the uplink where users have individual power constraints, on the downlink, it is possible to redistribute transmit powers across users without changing the total transmitted power from the base station. Thus the downlink is typically characterized by a sum power constraint.

For both the uplink and the downlink, the channel is assumed to experience slow and flat fading. Note that, with a sufficiently refined partition of the frequency band, a frequency-selective fading channel can be viewed as a number of parallel spectrally disjoint noninterfering essentially flat subchannels. It follows that, for any desired accuracy, the resulting channel matrix is equivalent to a block-diagonal flat-fading channel matrix. Hence the flat channel analysis presented here extends to frequency-selective fading in a straightforward manner. We assume that the channel matrices are perfectly known to the BS. The users are assumed to know their own channel and the spatial covariance structure of the sum of the noise and the relevant interference seen at the receiver.

Lastly, since the notion of *substreams* comes up in later sections, we elaborate what we mean by it. Note that a user's input vector X_i may further be composed of several *independent* vectors $X_{i1}, X_{i2}, \ldots$. This amounts to splitting the total rate for that user among several *substreams*. For a single user, it can be shown that rate splitting does not decrease capacity. For a single-antenna multiple access AWGN channel, rate splitting allows all points in the capacity region to be achieved without time-sharing [12]. For our purpose, splitting a users' power into substreams allows the substreams from different users to be interleaved in any manner with respect to the encoding/decoding order.

3. PROBLEM DEFINITION

Based on the FCFS model, our primary objective is to accommodate new users only to the extent that the users that are already active in the system are not affected. While this constitutes the general idea, to be precise, we need to distinguish between the following two cases.

Existing users are unaffected (preserving rates)

This would mean that the existing users continue to have the same rates as before. However, this leaves open the possibility that the existing users may adjust their transmit strategy on the uplink or their receive strategy on the downlink in some way to accommodate the new user. For example, on the downlink, it is conceivable that if superposition coding was used, then the existing users may need to decode and subtract out the new users signal before detecting their own signal. If this allows the existing users to achieve the same rates as before, we say that the existing users are not affected, or the rates are preserved.

Existing users are strictly unaffected (making the accommodation of new users invisible)

We could be more strict in our problem statement. We could demand that the new users be accommodated in such a way that not only do the existing users continue to achieve the same rates as before but also they are completely oblivious to the presence of new users. That is, the existing users' transmitters/receivers on the uplink/downlink continue to process the input data stream/received signal exactly as before to generate the transmitted signal/output data stream. Thus the only changes needed to accommodate the new user are made at the base stations. To distinguish this case from the previous one, we say that the existing users are *strictly* unaffected, or the new users are invisible.

Within each of the cases mentioned above, there are several, more or less equally significant, problems that one can pose. We list these problems in Sections 3.1 and 3.2 for the uplink and the downlink, respectively. We will see later that all the uplink problems really amount to the same problem—basically the same solution procedure covers all of the uplink variations. Among the downlink problems, we will encounter some substantive differences.

3.1. Uplink

On the uplink, the user's transmit power is the limiting factor. So, for the uplink, the first set of problems UP1a and UP1b (uplink problems 1a and 1b) that we wish to solve are as follows.

UP1a (preserving rates). Allocate the maximum possible rate to user K (new user) with transmit power P_K such

that the existing users' rates are not affected. Note that this allows the existing users to modify their transmit strategy to accommodate the new user so long as their rates are unaffected.

UP1b (making the new user invisible). Allocate the maximum possible rate to user K (new user) with transmit power P_K such that the existing users are *strictly* unaffected. Note that now, we require that the new user be invisible to the existing users, that is, the existing users must not modify their transmit strategy or their rates. Thus, the existing users are, in effect, oblivious to the presence of the new user.

We also briefly address the alternate problem where users have certain rate requirements and wish to achieve those rates with the minimum possible transmit power as follows.

UP2a (preserving powers). Determine the minimum possible transmit power for a new user K with rate requirement R_K such that the existing users' transmit powers are not affected.

UP2b (making the new user invisible). Determine the minimum possible transmit power for a new user K with rate requirement R_K such that the existing users are strictly unaffected.

3.2. Downlink

On the downlink, each base station distributes the total transmit power among the users it serves. Thus, unlike the uplink where each user has an individual power constraint, the downlink is characterized by a sum power constraint instead. The coding schemes we consider for the downlink are SD and DP. A brief description of these schemes is presented later. In particular, we wish to determine the following.

DP1. Is DP or SD a better scheme for the downlink in general?

For FCFS scheduling, the corresponding problems on the downlink would be as follows.

DP2a (preserving rates). Determine the maximum possible rate for user K subject to a total transmit power $P_1 + P_2 + \cdots + P_K$ such that existing users' rates are not affected.

DP2b (making the new user invisible). Determine the maximum possible rate for a user K subject to a total transmit power $P_1 + P_2 + \cdots + P_K$ such that existing users are strictly not affected.

Note that in problems DP2a and DP2b, the BS adds a power P_K to the total power to accommodate a new user (user K) into the system. The powers $P_1, P_2, \ldots, P_K$ determine how the rates are allocated to the users and need not be the actual transmitted powers in each user's input signal.

Note that as the channel changes, the users' rates/powers may change. So for each channel realization, we solve the FCFS scheduling problems listed above. The assumption that the channel varies slowly is important in this respect.

4. MIMO CAPACITY REVIEW

Before proceeding with the solutions to the problem defined in Section 3, we briefly visit the MIMO capacity expression. Consider the MIMO channel

$$Y = HX + \sum_{i=1}^{I} H_i X_i + N. \tag{5}$$

Here, X is the desired signal and $X_1, X_2, \ldots, X_I$ represent I independent interference signals. All input signals are assumed to be Gaussian with input covariance matrices $Q, Q_1^\star, Q_2^\star, \ldots, Q_I^\star$, respectively. Recall that the input covariance matrices identify the optimal spatial eigenmodes and the optimal power allocation across those eigenmodes. The input covariance matrices of the interfering signals $Q_i^\star$ are already fixed. We are interested in the optimal input covariance matrix $Q^\star$ for the desired signal X subject to total power constraint $\text{trace}(Q) \le P$. The H matrices represent the channels. The noise is assumed to be AWGN with covariance matrix normalized to identity. Note that this could apply to either the downlink or the uplink.

Since the interference is independent of the signal, the capacity of this channel is

$$\begin{aligned}
C &= \max_Q I(X;Y) \\
&= \max_Q h(Y) - h(Y \mid X) \\
&= \max_Q h\left(HX + \sum_{i=1}^{I} H_i X_i + N\right) - h\left(HX + \sum_{i=1}^{I} H_i X_i + N \mid X\right) \\
&= \max_Q h\left(HX + \sum_{i=1}^{I} H_i X_i + N\right) - h\left(\sum_{i=1}^{I} H_i X_i + N\right) \\
&= \max_Q \log \left| I + HQH^\dagger + \sum_{i=1}^{I} H_i Q_i^\star H_i^\dagger \right| \\
&\quad - \log \left| I + \sum_{i=1}^{I} H_i Q_i^\star H_i^\dagger \right| \\
&= \max_Q \log \left| I + \left(I + \sum_{i=1}^{I} H_i Q_i^\star H_i^\dagger\right)^{-1} HQH^\dagger \right|.
\end{aligned} \tag{6}$$

Thus the capacity of this channel can be expressed as $C = \log|I + (I + \sum_{i=1}^{I} H_i Q_i^\star H_i^\dagger)^{-1} HQ^\star H^\dagger|$. The optimal $Q^\star$ is determined as follows.

Since $\log|I + AB| = \log|I + BA|$, we can also express the capacity as

$$\begin{aligned}
C = \max_Q \log \Bigg| I + \left(I + \sum_{i=1}^{I} H_i Q_i^\star H_i^\dagger\right)^{-1/2} \\
\times HQ^\star H^\dagger \left(I + \sum_{i=1}^{I} H_i Q_i^\star H_i^\dagger\right)^{-1/2\dagger} \Bigg|
\end{aligned} \tag{7}$$

$$= \max_Q \log \left| I + \tilde{H} Q \tilde{H}^\dagger \right|, \tag{8}$$

where

$$\tilde{H} = \left(I + \sum_{i=1}^{I} H_i Q_i^\star H_i^\dagger\right)^{-1/2} H. \tag{9}$$

But (8) is the familiar MIMO capacity expression for a single user with channel $\tilde{H}$ in the presence of AWGN and without interference. The optimal input covariance matrix Q is obtained by the well-known waterfilling algorithm over the eigenmodes of $\tilde{H}$ [13].

Thus, in summary, the capacity for the channel (5) is given by

$$C = \log\left|I + \left(I + \sum_{i=1}^{I} H_i Q_i^\star H_i^\dagger\right)^{-1} H Q^\star H^\dagger\right|, \tag{10}$$

where $Q^\star$ is the optimal input covariance matrix obtained by waterfilling over the *effective* channel (9). Similar expressions appear quite frequently in later sections. To avoid repetition, instances of the same expressions presented later may be less descriptive. We advise the reader to refer back to this section and the references for details.

5. UPLINK SOLUTION

The uplink presents a relatively simple problem since the capacity region and the optimal coding strategy are known even with multiple antennas at the BS and the mobiles [14]. The desired solution is easily seen to be the well-recognized points on the capacity region corresponding to SD of users in a particular order. However, for the sake of completeness, and to strike a parallel with the downlink solutions presented later, we provide the solution and a self-contained proof as follows.

The solution to the first uplink problem UP1a (preserving rates) is given by the following theorem.

Theorem 1. *The optimal set of rates $R_i^\star$ on the uplink is*

$$R_i^\star = \log\left|I + \left(I + \sum_{j=1}^{i-1} H_j Q_j^\star H_j^\dagger\right)^{-1} H_i Q_i^\star H_i^\dagger\right|, \tag{11}$$

where $Q_i^\star$ is the optimal input covariance matrix obtained by waterfilling over the eigenmodes of the effective channel matrix $(I + \sum_{j=1}^{i-1} H_j Q_j^\star H_j^\dagger)^{-1/2} H_i$ subject to the power constraint $\mathrm{trace}(Q_i) = P_i$.

In other words, an optimal strategy for the uplink is to use SD (multiuser detection with successive interference cancellation) at the base station in the inverse order of the user's indices. The new user gets decoded first and his signal is subtracted out so that the existing users do not see him as interference. The highest rate that the new user can support without affecting existing users is simply given by the single-user waterfilling solution treating the existing users' signal as colored Gaussian noise.

Proof. We start with user 1. Ignoring the rest of the users, the highest rate he can support with power P_1 is

$$R_1^\star = \max_{p_1(\cdot)} I(X_1; H_1 X_1 + N), \tag{12}$$

where the maximization is over all distributions $p_1(X_1)$ that satisfy the power constraint (2). The optimal $p_1^\star(\cdot)$ is the well known zero-mean vector Gaussian distribution with covariance matrix $Q_1^\star$ determined by waterfilling over the eigenmodes of H_1. Let $X_1^\star \sim p_1^\star$. Note that the users' channels H_i are known and therefore H_1 is not a random variable in (12).

Now for the user 2, ignoring all but the user 1, from the multiple access capacity region, we have

$$R_1 + R_2 \le \max_{p_1(\cdot), p_2(\cdot)} I(X_1, X_2; H_1 X_1 + H_2 X_2 + N). \tag{13}$$

But R_1 and p_1 are already determined by the user 1. So we have

$$R_2^\star = \max_{p_2(\cdot)} I(X_1^\star, X_2; H_1 X_1^\star + H_2 X_2 + N) - R_1^\star, \tag{14}$$

$$\begin{aligned} R_2^\star &= \max_{p_2(\cdot)} I(X_1^\star, X_2; H_1 X_1^\star + H_2 X_2 + N) \\ &\quad - I(X_1^\star; H_1 X_1^\star + N), \end{aligned} \tag{15}$$

$$\begin{aligned} R_2^\star &= \max_{p_2(\cdot)} I(X_2; H_1 X_1^\star + H_2 X_2 + N) \\ &\quad + I(X_1^\star; H_1 X_1^\star + H_2 X_2 + N | X_2) \\ &\quad - I(X_1^\star; H_1 X_1^\star + N), \end{aligned} \tag{16}$$

$$\begin{aligned} R_2^\star &= \max_{p_2(\cdot)} I(X_2; H_1 X_1^\star + H_2 X_2 + N) \\ &\quad + I(X_1^\star; H_1 X_1^\star + N) - I(X_1^\star; H_1 X_1^\star + N), \end{aligned} \tag{17}$$

$$R_2^\star = \max_{p_2(\cdot)} I(X_2; H_1 X_1^\star + H_2 X_2 + N), \tag{18}$$

where (16) follows from the chain rule of mutual information and (17) follows from the independence of $X_1^\star$ and X_2. Note that this corresponds to decoding user 2 while treating user 1 as noise. Thus, at the base station, user 2 is decoded first and his signal is subtracted to obtain a clean channel for user 1. The optimal input distribution for user 2 is the waterfill distribution over the eigenmodes of $(I + H_1 Q_1^\star H_1^\dagger)^{-1/2} H_2$.

Proceeding in this fashion, we obtain the result of Theorem 1. □

It is interesting to note the simplicity of the solution. Note that the SD scheme requires only the BS to make some changes in the way it decodes the received signal. Specifically, the BS needs to decode the new user and subtract his signal before proceeding to decode the existing users' signals. However, the existing users themselves do not need to do anything different because of the new user. Thus the new user is completely invisible to existing users. Thus, we conclude that on the uplink, an optimal strategy that leaves the existing users' rates unaffected also leaves the existing users unaffected. In particular an optimal solution to UP1a (preserving rates) is also the optimal solution to UP1b (making the new user invisible).

The second pair of uplink problems UP2a (preserving powers, while using minimum additional power to meet a new user's rate) and UP2b (making the new user invisible, while meeting his rate with minimum additional power) are also very similar to UP1a and UP1b. Clearly for the user 1, the required transmit power is the one that achieves a capacity equal to his required rate R_1 with optimal waterfilling over his channel. In order for user 1's transmit power to be unaffected by user 2, the BS must decode user 2 before user 1. This also ensures that user 1 is not affected by user 2. Therefore, user 2 must see user 1 as noise. The required transmit power for user 2 is the one that achieves a capacity equal to his required rate R_2 with optimal waterfilling over his channel in the presence of colored noise due to the interference from user 1's signal. Thus, except that we know the rates and we need to solve for the transmit powers, the solution is the same as given by Theorem 1. Again UP2a and UP2b have the same solution.

6. DOWNLINK

6.1. Successive decoding and dirty paper

We begin this section with a brief summary of the key features of the SD and DP schemes. The details can be found in references.

SD is the well-known strategy, where several substreams are encoded directly on the channel input alphabet and *independent* of each other. Figure 1 shows an SD encoder. If a user has access to all codebooks, then he can decode any substream that is encoded at a rate lower than the capacity of his channel for that substream's input covariance matrix and treat other simultaneously transmitted codewords as noise. This allows him to reconstruct the transmitted codeword for the decoded substream and subtract its effect from the received signal, thus obtaining a cleaner channel for detecting other substreams.

With this strategy, a user may need to decode several codewords carrying other users' data and subtract their effect before he achieves a channel good enough to decode the codeword carrying his own data. Notice from Figure 1 that each encoder operates independent of all the other encoders.

Now, without loss of generality, we can assume that the substreams are encoded in some order, one after the other. This means that while choosing the codeword $\mathcal{C}_i^n$ for the ith substream, the transmitter has precise, noncausal information about the interference caused by all the $i-1$ substreams that have already been encoded. This brings us into the realm of DP coding. Figure 2 shows a DP encoder. Notice that unlike the SD scheme illustrated in Figure 1, where each encoder operates independent of the rest, in the DP scheme, there is a definite order such that the output of each encoder depends not only on the input substream data but also on the outputs of the encoders before it. This is possible because the encoders are collocated at the base station which allows them to cooperate perfectly.

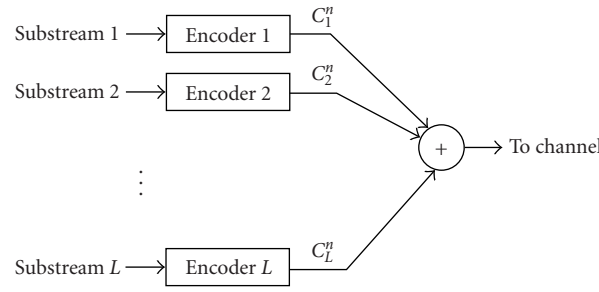


Figure 1: Encoding of L substreams in a successive decoding scheme.

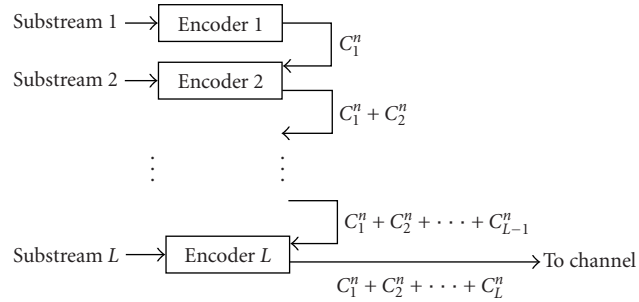


Figure 2: Encoding of L substreams in a dirty paper scheme.

The most powerful aspect of the DP scheme comes from the interesting work of Costa [8]. This paper presented the following result.

Costa's dirty paper result

Consider the scalar channel

$$Y_i = X_i + S_i + N_i, \tag{19}$$

where at each instant $i \in \mathbb{Z}^+$, Y_i is the output symbol, N_i is AWGN with power P_N, X_i is the input symbol constrained so that $E[X_i^2] \leq P_X$, and S_i is the interference symbol generated according to a Gaussian distribution. Now suppose the entire realization of the interference sequence $S_1, S_2, \ldots$ is known to the transmitter noncausally, that is, before the beginning of the transmission. This information is not available at the receiver. Then the capacity of the channel is given by

$$C = \log\left(1 + \frac{P_X}{P_N}\right), \tag{20}$$

irrespective of the power in the interference signal. In other words, if the interference is known to the transmitter beforehand, the capacity is the same as if the interference was not present. The capacity-achieving input distribution is $X \sim \mathcal{N}(0, P_X)$. Further, the channel input X and the interference S are independent.

Costa's result assumed a Gaussian distribution for the interference. The coding scheme described in [8] requires a

knowledge of the distribution of the interference for designing the codebooks. Thus, if the statistics of the interference changed from one codeword to another, the receiver would have to be informed and it would have to switch to a different codebook. Thus, with Costa's scheme, even though the capacity of a channel with interference known only to the transmitter would be the same as without it, the receiver would have to be informed about any change in the interference statistics so it can use the correct codebook.

Recent work by Erez et al. [15] showed that lattice strategies can be used to extend the Costa's result to *arbitrarily* varying interference. Their scheme is able to handle arbitrarily varying interference by communicating modulo a fundamental lattice cell and using dithering techniques. It is this lattice strategy that we imply by the term DP coding in this paper. For a detailed exposition of the scheme and the required background, see [15, 16, 17, 18].

Although Costa's work in [8] and the recent work of Erez et al. in [15] assume a scalar channel, the extension to the complex matrix channel is straightforward. A MIMO system with the channel matrix H known to both the transmitter and the receiver can be transformed into several parallel noninterfering *scalar* channels by a singular value decomposition [19] of the channel. Thus, it is easily verified that Costa's result carries through to the MIMO system with arbitrary interference and we have the following.

Extension to complex MIMO systems with arbitrarily varying interference

Consider the MIMO channel

$$Y_i = HX_i + S_i + N_i, \tag{21}$$

where H is the channel matrix known to both the transmitter and the receiver and at each instant $i \in Z^+$, Y_i is the output vector, N_i is AWGN vector with covariance matrix Q_N, X_i is the input vector constrained so that $Q_X = \text{trace}(E[X_i X_i^\dagger]) \leq P_X$, and S_i is an arbitrarily varying interference vector. All symbols are complex. Now suppose the entire realization of the interference sequence $S_1, S_2, \ldots$ is known to the transmitter non-causally. Then the capacity of the channel is given by

$$C = \max_{Q_X:\text{trace}(Q_X)\leq P_X} \log \frac{|HQ_XH^\dagger + Q_N|}{|Q_N|}, \tag{22}$$

irrespective of the power in the interference signal. In other words, if the interference is known to the transmitter beforehand, the capacity is the same as if the interference was not present. It is worth mentioning that this does assume that both the transmitter and receiver have access to a common source of randomness to allow the dithering operation. The capacity-achieving input distribution is $X \sim \mathcal{N}(\mathbf{0}, Q_X)$. Further, the channel input X and the interference S are independent.

Unlike Costa's scheme, the DP scheme works for arbitrarily varying interference. Therefore, no knowledge of interference statistics is required at the receiver. Thus, even if the interference statistics change from one codeword to another, the receiver continues to operate exactly the same way. This property in particular is crucial for our FCFS scheduling problem.

An important feature of the DP scheme is that the capacity-achieving codes are not the channel input symbols $\mathcal{C}_i^n$ but the functions used to map the data and the transmitter side information to the channel input alphabet. Since the coding is not performed on the channel input alphabet itself, even if one decodes the data carried by a substream, it is not possible to subtract the effect of the transmitted symbols of the substream and obtain a cleaner channel. For example, refer to Figure 2. Decoding the ith substream does not allow a user to reconstruct the transmitted symbols $\mathcal{C}_i^n$ and therefore the user cannot subtract out $\mathcal{C}_i^n$ to obtain a cleaner channel.

In Figure 2, before encoding substream i, the transmitter knows the interference from substreams $1, 2, \ldots, i-1$. Thus the capacity achieved by substream i is the same as if substreams $1, 2, \ldots, i-1$ were not present. The interference from substreams $i+1, i+2, \ldots, L$ is not known and so it must be treated as noise.

To highlight the distinction between SD and DP, consider the following example of a broadcast system with two encoded substreams: substream 1 and substream 2. With SD, especially on a nondegraded broadcast channel, it is possible that one user can decode and cancel substream 2 before decoding substream 1, and at the same time another user with a different channel can decode and cancel substream 1 before decoding substream 2. Thus the decoding order may vary from user to user. On the other hand, with DP, there is a *fixed encoding order* such that the substreams encoded later achieve the same capacity as if the substreams encoded before them were not present. Moreover, the substreams encoded earlier can achieve a capacity no higher than that achievable by treating all substreams encoded after them as noise. In a nutshell, in SD, the encoding order is irrelevant and the optimal decoding order may vary from one user to another. In DP, there is no notion of decoding order. Instead, there is only one encoding order, where each substream has a unique position relative to every other substream. For each receiver, this unique order decides which substreams have to be treated as noise and which substreams do not impact the capacity of its own substream.

6.2. Solution to DP1 (DP versus SD)

The first problem we address on the downlink is to determine whether SD or DP is a better scheme in general. Before stating the solution, we see why it is not trivial. Consider two substreams intended for two different users. With DP, one of the users (the one encoded second) can achieve the same capacity as if the other user was not present. However, the other user (who was encoded first) must treat this user as noise and his capacity is reduced. With SD on the other hand, depending on the users' channels and the input covariance matrices, several situations are possible. It could be that the channels are such that each user can decode the other user's substream and subtract it before decoding his own substream. This seems to be better than DP. However, it

could also happen that the channels are such that neither user can decode the other user's substream. In that case, SD would be worse than DP. Since it is the downlink, one can also optimize the transmit power across users while keeping the same total transmit power. Further, the rate regions may not be convex. In such a case, we can make the rate region convex by including rate vectors achievable with time-sharing. With all these possibilities, the question as to whether SD or DP is the better strategy on the downlink does not seem to have an obvious answer.

With the following theorem, we show that DP is the better downlink strategy in general.

Theorem 2. *Subject to a sum power constraint, the set of rate vectors achievable with SD and time-sharing is also achievable with DP and time-sharing.*

In other words, the convex hull of the achievable rate region with SD is completely contained within the convex hull of the achievable rate region with DP.

Proof. We prove this by showing that the boundary of the achievable rate region with SD and time division is contained within the boundary of the achievable rate region with DP and time-sharing. Note that in either scheme, the points in the interior can always be attained by throwing away some codewords.

The boundary points of the rate region are obtained by maximizing

$$\sum_{i=1}^{K} \mu_i R_i \tag{23}$$

for all $\vec{\mu}$ such that $\vec{\mu} \geq \vec{0}$ and $\sum_{i=1}^{K} \mu_i = 1$.

Let $\mathcal{R}^{\mathrm{SD}}$ and $\mathcal{R}^{\mathrm{DP}}$ denote the sets of rate vectors achievable with SD and DP, respectively. Note that in order to prove the result of Theorem 2, it suffices to prove that for all $\vec{\mu}$,

$$\max_{\vec{R} \in \mathcal{R}^{\mathrm{DP}}} \sum_{i=1}^{K} \mu_i R_i \geq \max_{\vec{R} \in \mathcal{R}^{\mathrm{SD}}} \sum_{i=1}^{K} \mu_i R_i. \tag{24}$$

In order to prove (24), we assume without loss of generality that the users' priorities are arranged as $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_K$. We start with the SD scheme and show that DP can achieve at least the same value of $\vec{\mu} \cdot \vec{R}$. Let $\vec{R}^{\mathrm{SD}}$ be the rate vector that maximizes $\vec{\mu} \cdot \vec{R}$ with SD. Without loss of generality, we can assume that $\vec{R}^{\mathrm{SD}}$ does not use time-sharing. This is because simple linear programming tells us that a rate vector corresponding to time-sharing between several different rate vectors is a convex combination of those rate vectors and therefore cannot achieve a higher value of $\vec{\mu} \cdot \vec{R}^{\mathrm{SD}}$ than the best of those rate vectors.

Let the total number of substreams being transmitted be L. Further, and again without loss of generality, we label the substreams from 1 to L such that if $i < j$ and substream i carries data for user $u(i)$ and substream j carries data for user $u(j)$, then $\mu_{u(i)} \geq \mu_{u(j)}$. That is, the substreams are arranged in decreasing order of the priority of the user whose data they are carrying. For multiple substreams carrying the same user's data, we label them in the order in which they are decoded by that user.

Now note that no user can decode a substream carrying data for a user with a lower priority. This is easily proved by contradiction as follows. Suppose that user A can decode a substream that carries user B's data at a rate r. Now if user A has a higher priority than user B, that is, if $\mu_A > \mu_B$, then we can increase $\vec{\mu} \cdot \vec{R}^{\mathrm{SD}}$ by simply having the substream carry user A's data instead of user B's data at the same rate, r so that,

$$\vec{\mu} \cdot \vec{R}(\mathrm{new}) = \vec{\mu} \cdot \vec{R}^{\mathrm{SD}} - \mu_B r + \mu_A r > \vec{\mu} \cdot \vec{R}^{\mathrm{SD}}. \tag{25}$$

But this is a contradiction since we assumed that the rate vector $\vec{R}^{\mathrm{SD}}$ maximizes $\vec{\mu} \cdot \vec{R}$ over all rate vectors $\vec{R}$ achievable with SD and without time-sharing.

In light of this observation, it is clear that while decoding substream l, the intended user must treat substreams $l+1$ to L as noise. The substreams 1 to $l-1$ may or may not be treated as noise depending upon whether it is possible to decode and subtract those substreams or not. So with SD, the rate achieved on the lth substream is no greater (could be smaller) than r_l, where r_l is the achievable rate when the substreams $l+1$ to L are treated as noise while substreams 1 to $l-1$ are not present. Next, we show that DP can achieve r_l on each of these substreams.

Suppose we use DP to encode the L substreams in the order in which they are labeled. Then the lth substream sees substreams $l+1$ to L as noise since these substreams are encoded *after* substream l and therefore the interference caused by them is not known. However, since substreams 1 to $l-1$ have already been encoded, they present known interference to substream l and therefore do not affect the data rate that substream l is capable of supporting. Thus DP allows substream l a rate r_l that is at least as large as the maximum allowed rate for that substream in the optimum SD rate vector that maximizes $\vec{\mu} \cdot \vec{R}$. This proves (24) and completes the proof of Theorem 2. □

We can also easily extend this theorem to show that the achievable rate region of the pure DP scheme includes the achievable rate region of not only the pure SD scheme but also any hybrid scheme where some users use SD while others use DP. Lastly, we need time-sharing for this result because the achievable rate region for SD and DP without time-sharing may not be convex.

6.3. Downlink solutions for DP2a (preserving rates) and DP2b (making the accommodation of new users invisible)

In DP2a, we are only requiring rate conservation in dealing with the Kth user. This leaves open the possibility that, in meeting the earlier rates, if the earlier users are handled in a different way than before, we can actually achieve a strictly

greater rate for the Kth user. Indeed, in some instances, a greater rate is possible. This DP2a problem is exceptional in that we encounter the most difficult of the optimization problems in this paper and a solution is only presented for a special case. In the general case, *based on the conjecture in [9]*, a solution can, in theory, be obtained by solving a number of convex programming problems to obtain the achievable rate region with DP coding [20]. However, the complexity of this is exponential in the number of users.

In problem DP2b, we insist that earlier users be treated exactly as before. Later users must be invisible (phantoms) to earlier ones. It turns out that, with this added constraint, we can obtain a complete solution. Moreover, as we will see in Section 7, a solution is possible for the full multiple base station setup.

6.3.1. Solution to DP2a (preserving rates)

Next, we address the problem of assigning the maximum rate to new user K subject to total power $P_1 + P_2 + \cdots + P_K$ such that the existing users' *rates* are not affected. So we wish to allocate the maximum possible rates to each user such that

(i) user 1 gets $R_1^\star$, the maximum rate possible with power P_1 as if no other user was present,

(ii) user 2 gets $R_2^\star$, the maximum rate possible with *total* power $P_1 + P_2$ such that user 1 still gets $R_1^\star$ and as if users $3, \ldots, K$ were not present,

(iii) user K gets $R_K^\star$, the maximum rate possible with total power $P_1 + P_2 + \cdots + P_K$ such that users 1 through $K - 1$ still get rates $R_1^\star$ through $R_{K-1}^\star$.

While the overall optimization seems hard for the general multiple antenna broadcast system, limiting the number of transmit antennas at the base station to one does lead to a simple solution. A single transmit antenna at the base station makes the channel degraded and the optimality of Gaussian inputs is established from Bergman's proof in [21]. Note that although Bergman's proof is for scalar broadcast channels, that is, broadcast channels with a single transmit antenna at the base station and a single receive antenna at each user, the vector broadcast channel with a single antenna at the base station and multiple receive antennas at each user is easily seen to be equivalent to the scalar broadcast channel [22]. Thus, in this case, the capacity region is well known and we do not need the conjecture of [9]. Next, we present this solution to gain some insight.

With a single transmit antenna at the base station, the downlink is a degraded broadcast channel. Even with multiple receive antennas, each user can perform spatial matched filtering to yield a *scalar* AWGN channel for himself [22]. For this channel, the broadcast capacity is well known and either SD or DP can be used to achieve any point in the capacity region. In particular, all the rate points can be achieved with SD/DP with the *same* encoding/decoding order [23]. The user with the weakest channel is decoded/encoded first so that he sees everyone else as noise. The decoding/encoding proceeds in the order of the users' channel strengths so that weaker users who cannot decode the stronger users are forced to treat their signal as noise while the stronger users can decode the weaker users' data, and are therefore unaffected by the presence of weaker users. Thus, in this case, the encoding/decoding order is decided by the users' channels and not by the order of users' arrivals or their relative priorities.

For each channel state, we calculate the optimal rates and powers in an iterative fashion as follows. We start with only user 1 in the system with total power P_1 and find $R_1^\star$. Then we incrementally add users to the system, in the order $2, 3, \ldots, K$, each time finding the optimal rates for the set of users in the system with total power given by the sums of the powers of those users. The ith user is added as follows.

(1) Arrange the users in the order of their channel strengths.

(2) The users with a stronger channel than user i are not affected. That is, they continue to use the same power and rates as before.

(3) The users with a weaker channel than user i have to treat user i as noise. So the additional power P_i available to the system is distributed among user i and the weaker users so that the weaker users can sustain the same rates as before.

The optimal distribution of the additional power among the new user and the weaker users requires only a one-dimensional optimization and is easily obtained. Proceeding in this fashion, after the Kth user has been added, we obtain the optimal rate and power allocation for all the users in the system. Note that this is the optimal allocation because the rate vector obtained in this fashion lies on the boundary of the capacity region.

While this solution does not affect the existing users' rates, it does affect the existing users in that they may have to decode the new user before decoding their own signals if SD is used. If DP is used, then the existing users may have to see the new user as spatially colored noise. They are still able to achieve the same rates as before because they have a higher power. Thus, the solution does not allow the existing users to continue operating as before.

Next, we present a solution that gives the new user K the maximum rate possible with total transmit power $P_1 + P_2 + \cdots + P_K$ without affecting existing users.

6.3.2. Solution to DP2b (making the accommodation of new users invisible)

Theorem 3. *The optimal set of rates $R_i^\star$ on the downlink such that existing users are oblivious to the presence of the new users is given by*

$$R_i^\star = \log \left| I + \left(I + \sum_{j=1}^{i-1} H_i Q_j^\star H_i^\dagger \right)^{-1} H_i Q_i^\star H_i^\dagger \right|, \qquad (26)$$

where $Q_i^\star$ is the optimal input covariance matrix obtained by waterfilling over the eigenmodes of the effective channel matrix $(I + \sum_{j=1}^{i-1} H_i Q_j^\star H_i^\dagger)^{-1/2} H_i$ subject to the power constraint $\text{trace}(Q_i) = P_i$.

In other words, an optimal strategy for the downlink that does not allow new users to affect existing users is to use DP encoding at the base station in the inverse order of the user's indices. The new user gets encoded first so his signal is a known interference and the existing users' rates do not get affected. The highest rate that the new user can support without affecting existing users is simply given by the single-user waterfilling solution treating the existing users' signal as colored Gaussian noise. A simple example to illustrate the optimal downlink scheme is presented after the proof.

Proof. DP's ability to handle arbitrarily varying interference makes it the obvious choice in this case. Using SD would require existing users to decode the new user, thus acknowledging the new user's presence. However, since DP is able to handle arbitrary interference, it does not matter if the interference known to the ith user's encoder comes from users $i, i+1, \ldots, K-1$ or from users $i, i+1, \ldots, K$. The rate and decoding strategy for user i depend only on the interference from users $1, 2, \ldots, i-1$ that came before him and whose signals must be treated as noise for user i.

Note that time-sharing and rate-splitting are not required. This is easily seen as follows. With only user 1 in the system, time-sharing between different rates at different powers would decrease his overall rate since capacity is strictly concave in transmit power (Jensen's inequality). Rate splitting is not needed either. Thus user 1 does not use time-sharing when he is the only user in the system. Since user 1 is oblivious to the presence of new users, the BS cannot use time-sharing or split user 1's data into substreams and rearrange the encoding order of these substreams when new users appear. The same logic applies to all users.

Thus, no time-sharing or rate-splitting is required and the optimal DP vector is the one where users are encoded in the inverse order of their indices. □

To better illustrate the downlink strategy, we present a detailed example for a system with 3 users. The base station follows the following sequence of steps *in this order*.

(1) Determine the rate $R_1^\star$ and the input covariance matrix $Q_1^\star$ for user 1 according to equation (26). Note that these are simply the single-user capacity of user 1's channel and the waterfilling distribution that achieves that capacity when no other user is present.

(2) Determine the rate $R_2^\star$ and the input covariance matrix $Q_2^\star$ for user 2 according to equation (26). These are the single-user capacity and the waterfilling distribution that achieves that capacity for user 2's channel treating the interference from user 1 at the output of user 2's channel as colored Gaussian noise.

(3) Determine the rate $R_3^\star$ and the input covariance matrix $Q_3^\star$ for user 3 according to equation (26). These are the single-user capacity for user 3's channel and the waterfilling distribution that achieves that capacity treating the interference from users 1 and 2 as colored Gaussian noise.

(4) Encode user 3's data. That is, generate $\mathcal{C}_3^n$.

(5) Using the knowledge of the interference caused by $\mathcal{C}_3^n$ at the output of user 2's channel, encode user 2's data. That is, generate $\mathcal{C}_2^n$. Thus, user 3 presents known interference to user 2 and does not affect user 2's capacity.

(6) Using the knowledge of the interference caused by $\mathcal{C}_3^n + \mathcal{C}_3^n$ at the output of user 1's channel, encode user 1's data. That is, generate $\mathcal{C}_1^n$. Thus, users 2 and 3 present known interference to user 1 and do not affect user 1's capacity.

Note that in order to determine the users' optimal rates and input distributions, we need to proceed in the order $1, 2, \ldots, K$. However, after that the actual codes are generated in the order $K, K-1, \ldots, 1$.

The solution for the downlink is interesting for its simplicity and also for its striking symmetry with the uplink solution.

7. MULTIPLE BASE STATIONS

In this section, we incorporate multiple base stations to model a multicell environment. We assume that all the base stations are connected through a high-speed reliable network. It allows perfect coordination and information exchange between base stations. *Cooperation between base stations has also been considered previously for the uplink by Wyner in [24] and for the downlink by Shamai and Zaidel in [25].*

7.1. Uplink

On the uplink, the received signal at the bth base station is characterized by the following equation:

$$Y^{[b]} = \sum_{i=1}^{K} H_i^{[b]} X_i + N^{[b]}, \tag{27}$$

where $Y^{[b]}$ is the received vector at the bth base station, K is the number of users currently active in the system, $H_i^{[b]}$ is the flat-fading $B_b \times U_i$ matrix channel between user i and base station b, B_b and U_i are the numbers of antennas at the bth base station and the ith user, respectively, and N^b is the AWGN vector at the bth base station.

However, since we allow perfect coordination and information exchange between base stations, note that we can treat all the base stations together as one big base station with all the antennas. The equivalent description of the received signal at this base station is given by (1).

$$Y = \sum_{i=1}^{K} H_i X_i + N. \tag{28}$$

Here Y, H_i, and N are obtained by stacking up on top of each other the corresponding $Y^{[b]}$, $H_i^{[b]}$, and $N^{[b]}$ for all the base stations. But this brings us back to the single-cell model. Thus, for the uplink, the optimal solutions for the single cell simply carry through to the multicell environment.

7.2. Downlink

We extend the downlink solution to DP2b (existing users oblivious to the presence of new users) with multiple cells.

The downlink with B base stations is described as

$$Y_i = \sum_{b=1}^{B} H_i^{[b]} \sum_{j=1}^{K} X_j^{[b]} + N_i, \quad 1 \le i \le K,\ 1 \le b \le B, \quad (29)$$

where Y_i is the output vector, $X_i^{[b]}$ and $H_i[b]$ are the input vector and the channel matrix from base station b, and N_i is the AWGN vector for user i. Further, the additional power for each new user is limited per base station so that

$$\text{trace}\left[E\left[X_i^{[b]} X_i^{[b]\dagger}\right]\right] \le P_i^{[b]}, \quad 1 \le i \le K,\ 1 \le b \le B. \quad (30)$$

Note that a system where each user is assigned to only one base station is included as a special case by setting the appropriate power constraints to zero.

Again, since we allow perfect coordination between base stations, we can represent the B base stations as one big base station. Defining

$$H_i = \left[H_i^{[1]} \ \ H_i^{[2]} \ \ \cdots \ \ H_i^{[B]}\right], \quad 1 \le i \le K, \quad (31)$$

and X_i as the vector obtained by stacking all the $X_i^{[b]}$ into a single column, we obtain an equivalent representation for the downlink as (3). Now this looks similar to the single-cell downlink model we had earlier. However, note that the components of the input vector X_i come from different base stations. There is a different input power constraint on each base station. Thus, the solution presented earlier does not apply in the exact same form. *However, a natural extension of the single-cell downlink solution to multiple base stations is obtained as follows.*

Although rate splitting is not necessary, recall that it does not reduce capacity. We explain the multicell extension of the single-cell downlink solution in terms of rate splitting for clarity. Specifically, we split each user's rate into B substreams. The idea is to perform the waterfill in B stages. At each stage, we waterfill until a base station meets its power constraint. Then we null out the antenna gains from that base station so that no more power is allocated to it and proceed with the waterfill. This gives us B layers or B substreams that can be encoded using DP. Consider the ith user. As shown in Theorem 3, this user sees the interference from users $1, 2, \ldots, i-1$ as colored noise and is unaffected by the interference from users $i+1, i+2, \ldots, K$. Therefore, the maximum rate he can achieve is given by

$$R_i^{\star} = \max_{Q_i} \log \left| I + \left(I + \sum_{j=1}^{i-1} H_j Q_j^{\star} H_j^{\dagger}\right)^{-1} H_i Q_i H_i^{\dagger} \right|, \quad (32)$$

where the maximization is over all input covariance matrices that satisfy the power constraints per base station. We split the user's rate into B substreams to be encoded in the order $B, B-1, \ldots, 1$ using DP encoding. So the Bth substream sees all the other substreams as noise, while the first substream's rate is unaffected by substreams $B, B-1, \ldots, 2$. Let the rates on these substreams be $R_i^{[b]\star}$, and the corresponding input covariance matrices be $Q_i^{[b]\star}$. Then we have

$$\begin{aligned} R_i^{\star} &= R_i^{[1]\star} + R_i^{[2]\star} + \cdots + R_i^{[B]\star}, \\ Q_i^{\star} &= Q_i^{[1]\star} + Q_i^{[2]\star} + \cdots + Q_i^{[B]\star}. \end{aligned} \quad (33)$$

The optimal $Q_i^{\star}$ is obtained as follows.

(1) Perform a singular value decomposition of the effective composite channel $(I + \sum_{j=1}^{i-1} H_j Q_j^{\star} H_j^{\dagger})^{-1/2} H_i$ as

$$\left(I + \sum_{j=1}^{i-1} H_j Q_j^{\star} H_j^{\dagger}\right)^{-1/2} H_i = F_i \Lambda_i M_i. \quad (34)$$

Start water-pouring over the eigenmodes of this channel. Continue adding power until one of the base stations meets its power constraint for the ith user $P_i^{[b]}$. Without loss of generality, we assume base station 1 runs out of power for user i. This corresponds to the first rate split, that is, call this the first substream for user i. The input covariance matrix obtained in this way is $Q_i^{[1]\star}$. Among the B substreams corresponding to user i, this substream will be encoded last, so it is unaffected by the interference from the remaining $B-1$ substreams. The rate on this substream is

$$R_i^{[1]\star} = \log \left| I + \left(I + \sum_{j=1}^{i-1} H_j Q_j^{\star} H_j^{\dagger}\right)^{-1} H_i Q_i^{[1]\star} H_i^{\dagger} \right|. \quad (35)$$

(2) Since base station 1 already used up its power for user i, we null out the contribution from $H_i^{[1]}$ to the compound channel matrix by setting it to zero. Define a new composite channel

$$H_i^{[-1]} = \left[\mathbf{0} \ \ H_i^{[2]} \ \ H_i^{[3]} \ \ \cdots \ \ H_i^{[B]}\right]. \quad (36)$$

Again, perform a singular value decomposition on the new composite effective channel

$$\begin{aligned} &\left(I + \sum_{j=1}^{i-1} H_j Q_j^{\star} H_j^{\dagger} + H_i Q_i^{[1]\star} H_i^{\dagger}\right)^{-1/2} H_i^{[-1]} \\ &\quad = F_i^{[-1]} \Lambda_i^{[-1]} M_i^{[-1]}. \end{aligned} \quad (37)$$

Note that this treats the interference from the first substream as noise. Again, start water-pouring over the eigenmodes of this new channel until another base station meets its power constraint. Without loss of generality, we call this base station 2. This gives us the input covariance matrix $Q_i^{[2]\star}$ on the second substream. The rate for the second substream is

$$\begin{aligned} R_i^{[2]\star} = \log \Bigg| &I + \left(I + \sum_{j=1}^{i-1} H_j Q_j^{\star} H_j^{\dagger} + H_i Q_i^{[1]\star} H_i^{\dagger}\right)^{-1} \\ &\times H_i^{[-1]} Q_i^{[2]\star} H_i^{[-1]\dagger} \Bigg|. \end{aligned} \quad (38)$$

Proceeding in this fashion, we obtain the input covariance matrices on all the substreams and the corresponding rates as well. Combining the substreams, we get the overall rate and input covariance matrix for each user from equations (33).

Thus we find that multiple base stations only affect the downlink solution to the extent that the waterfilling algorithm needs some modification in order to accommodate the different power constraints per base station. Otherwise, the solution does not change. In particular, we still use DP coding, and the ordering of users is the same as before. Also note that while we used rate splitting to derive the optimal input covariance matrix, it is not necessary to split the rates into substreams. The same overall input covariance matrix can be used without rate-splitting to achieve the same capacity.

8. CONCLUSIONS AND DISCUSSION

We addressed the problem of providing best possible rates to new users as they enter a wireless data network, without penalizing the existing users. We have dubbed the network a PhantomNet. This is because of the design theme, that when a user enters, all subsequent entrants must, to him, be phantoms, that is, interference-wise, they must be invisible. For both the uplink and the downlink, only earlier entrants can interfere with an entering user. PhantomNet operation involves treating all bases as a single composite base, so that the actual bases simply serve as multiple antenna sites which are networked, say with fibers, to and from a single central processor.

For the uplink we found that, to achieve the phantom requirement, we could make a straightforward application of the well-established SD strategy where the new user is *decoded* before the existing users. For the downlink, achieving the invisibility requirement is more problematic. The optimal downlink strategy is to use DP coding, where the new user is *encoded* before the existing users. This makes use of the fact that the bases have knowledge of all signals that are to be transmitted. This enables simultaneous communication to the users despite arbitrarily varying interference by signalling modulo a fundamental lattice cell and using dithering techniques.

The striking feature of the uplink and the downlink strategies is their simplicity, and even more than that, their similarity. In both cases, the new users are forced to see the existing users as noise while the existing users are not affected by the presence of the users who joined the system after them. That is, they can continue to operate exactly as before. The only changes need to be made at the BS. For the uplink, the base station is the decoder and thus the solution hinges on the optimal decoding order, whereas for the downlink, the base station is the encoder and the solution is based on the encoding order. Note that as users leave the system, the same structure is maintained. As a user exits, it does not affect the rates of the users who joined the system before him. It does help the users who joined the system after him since they no longer have to face interference from his signal.

With multiple cells, we found that the uplink was effectively the same as a single-cell system since all the base stations are treated as one composite base station. Thus the single-cell strategy extends to multiple cells without loss of optimality. In contrast to the uplink, while the downlink is also viewed as a single virtual base station, there is a refinement since each of the actual base stations has a separate total power constraint. Consequently, the multiple cell downlink solution is different in that the distinct total transmit power constraints require a multistage waterfilling solution in determining the optimal input covariance matrix for each user. At each stage, waterfilling is performed until each base station meets its total power constraint. Those base stations that have already met their power constraints are not considered in the successive waterfilling stages.

While we drew heavily on published results, the novelty of our finding is the generality achieved in our setting: multiple base stations and multiple users with multiple antennas accommodated at both the transmit and receive sites. We also proved a general result that extends beyond our framework. We showed that the achievable rate region with SD and time-sharing is contained within the achievable rate region with DP coding and time-sharing.

We stress that PhantomNet uses information theoretic means for self-organizing the allocation of communication resources. There is allowance of extreme flexibility in allocating resources to a user. For example, which bases, which antennas at the bases, (which sectors) and which frequency bands are made available to a user need not be imposed over the network area. Instead, resource allocations can be left to develop, dynamically as needed (on the fly), in a fine-grained manner as expressed by the information theory formulas that we have presented. Dynamic choices would emerge as users come and go. Whenever and wherever and to what extent such amorphous allocations result in a superior network compared to imprinting a rigid regular structure from the outset is a topic for future study. Through constraints, one is free to impose structure when it looks advisable. A simulation testbed could be used to study PhantomNet operation to learn which beneficial features should first be moved into practice. Such a testbed could also be used to quantify the value of more antennas, more sectorization, and so forth.

ACKNOWLEDGMENT

It is a pleasure to thank Angel Lozano, Harish Viswanathan, Howard C. Huang, and Sriram Vishwanath for helpful discussions.

REFERENCES

[1] H. Viswanathan, S. Venkatesan, and H. Huang, "Downlink capacity evaluation of cellular networks with known interference cancellation," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 802–811, 2003.

[2] T. M. Cover, "Comments on broadcast channels," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2524–2530, 1998.

[3] G. Caire and S. Shamai, "On achievable rates in a multi-antenna Gaussian broadcast channel," in *Proc. IEEE International Symposium on Information Theory (ISIT '01)*, Washington, DC, USA, June 2001.

[4] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.

[5] S. Vishwanath, N. Jindal, and A. Goldsmith, "On the capacity of multiple input multiple output broadcast channels," in *Proc. IEEE International Conference on Communications (ICC '02)*, pp. 1444–1450, New York, NY, USA, April 2002.

[6] W. Yu and J. Cioffi, "Trellis precoding for the broadcast channel," in *Proc. IEEE Global Telecommunication Conference (Globecom '01)*, pp. 1338–1344, San Antonio, Tex, USA, November 2001.

[7] P. Viswanath and D. N. Tse, "Sum capacity of the multiple antenna Gaussian broadcast channel and uplink-downlink duality," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, 2003.

[8] M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[9] S. Vishwanath, G. Kramer, S. Shamai, S. Jafar, and A. Goldsmith, "Capacity bounds for Gaussian vector broadcast channels," in *Proc. of DIMACS Workshop on Signal Processing for Wireless Transmission*, Rutgers University, Piscataway, NJ, USA, October 2002.

[10] P. Viswanath and D. Tse, "On the capacity of the multiple antenna broadcast channel," in *Proc. DIMACS Workshop on Signal Processing for Wireless Transmission*, Rutgers University, Piscataway, NJ, USA, October 2002.

[11] H. Viswanathan and K. Kumaran, "Rate scheduling in multiple antenna downlink wireless systems," Technical Memorandum 10009626-010720-01TM, Bell Laboratories, 2001.

[12] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple access channel," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 364–375, 1995.

[13] F. R. Farrokhi, G. J. Foschini, A. Lozano, and R. A. Valenzuela, "Link-optimal space-time processing with multiple transmit and receive antennas," *IEEE Communications Letters*, vol. 5, no. 3, pp. 85–87, 2001.

[14] W. Yu, W. Rhee, S. Boyd, and J. Cioffi, "Iterative water-filling for vector multiple access channels," in *Proc. IEEE International Symposium on Information Theory (ISIT '01)*, p. 332, Washington, DC, USA, June 2001.

[15] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice-strategies for cancelling known interference," in *Proc. International Symposium on Information Theory and Its Applications (ISITA '00)*, Honolulu, Hawaii, USA, November 2000.

[16] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Transactions on Information Theory*, vol. 42, no. 4, pp. 1152–1159, 1996.

[17] R. Urbanke and B. Rimoldi, "Lattice codes can achieve capacity on the AWGN channel," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 273–278, 1998.

[18] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423–1443, 2001.

[19] J. Kim and J. Cioffi, "Spatial multiuser access with antenna diversity using singular value decomposition," in *Proc. IEEE International Conference on Communications (ICC '00)*, pp. 1253–1257, New Orleans, La, USA, June 2000.

[20] S. Vishwanath, N. Jindal, and A. Goldsmith, "On the capacity of multiple input multiple output broadcast channels," in *Proc. IEEE International Conference on Communications (ICC '02)*, New York, NY, USA, April 2002.

[21] P. P. Bergmans, "A simple converse for braodcast channels with addtiive white Gaussian noise," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 279–280, 1974.

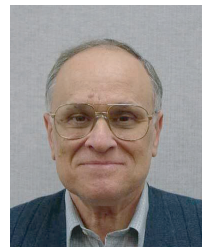
[22] S. Jafar and A. Goldsmith, "On the capacity of vector Gaussian MAC and BC with feedback," in *Proc. of 41st Annual Allerton Conference on Communication, Control and Computing*, Monticello, Ill, USA, October 2003.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1991.

[24] A. D. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Transactions on Information Theory*, vol. 40, no. 6, pp. 1713–1727, 1994.

[25] S. Shamai and B. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *Proc. of 53rd Vehicular Technology Conference*, vol. 3, pp. 1745–1749, Rhodes, Greece, Spring 2001.

Syed A. Jafar received his B. Tech. degree in electrical engineering from the Indian Institute of Technology (IIT), Delhi, India, in 1997, his M.S. degree in electrical engineering from California Institute of Technology (Caltech), Pasadena, USA, in 1999, and his Ph.D. degree in electrical engineering from Stanford University in 2003. He was a Senior Engineer at Qualcomm Inc., San Diego, in 2003. He did a summer internship in the Wireless Communications Group, Lucent Bell Laboratories, Holmdel, NJ, in 2001, and has two pending patents resulting from that work. He was an Engineer in the satellite networks division, Hughes Software Systems, India, from 1997 to 1998. In January 2004, he joined the Department of Electrical Engineering and Computer Science at the University of California, Irvine, where he is currently an Assistant Professor. His research interests include spread-spectrum systems, multiple antenna systems, and multiuser information theory.



Gerard J. Foschini received his B.S.E.E. degree from New Jersey Institute of Technology (NJTT) in 1961, his M.E.E. degree from New York University (NYU) in 1963, and his Ph.D. degree in mathematics from Stevens Institute of Technology in 1967. Mr. Foschini has been at Bell Laboratories for 41 years. He is a Bell Labs Fellow and he holds the position of Distinguished Member of Staff and Distinguished Inventor. He has conducted data communications research on many kinds of systems, most recently wireless communications and optical communications systems. Gerard has done extensive research on point-to-point systems as well as on networks. He has authored one hundred papers and holds seven patents. He has won the IEEE Eric E. Sumner Award, the Thomas Alva Edison Patent Award, and the Bell Labs President's Gold Award. He has taught at both Princeton and Rutgers Universities.

Andrea J. Goldsmith received her B.S., M.S., and Ph.D. degrees in electrical engineering from U.C. Berkeley in 1986, 1991, and 1994, respectively. She was an Assistant Professor in the Department of Electrical Engineering at Caltech from 1994 to 1999. In 1999, she joined the Electrical Engineering Department at Stanford University, where she is currently an Associate Professor. Her industry experience includes affiliation with Maxim Technologies from 1986 to 1990, where she worked on packet radio and satellite communication systems, and with AT&T Bell Laboratories from 1991 to 1992, where she worked on microcell modeling and channel estimation. Her research includes works in capacity of wireless channels and networks, wireless information and communication theory, multiantenna systems, joint source and channel coding, cross-layer wireless network design, communications for distributed control, and adaptive resource allocation for cellular systems and ad-hoc wireless networks. Dr. Goldsmith is a Terman Faculty Fellow at Stanford and a recipient of the Alfred P. Sloan Fellowship, the National Academy of Engineering Gilbreth Lectureship, a National Science Foundation Career Development Award, the Office of Naval Research Young Investigator Award, a National Semiconductor Faculty Development Award, an Okawa Foundation Award, and the David Griep Memorial Prize from U.C. Berkeley. She was an Editor for the IEEE Transactions on Communications from 1995 to 2002, and has been an Editor for the IEEE Wireless Communications Magazine since 1995. She is also an elected member of Stanford's faculty senate and the board of governors for the IEEE Information Theory Society.

EURASIP Journal on Applied Signal Processing 2004:5, 605–612

D-BLAST OFDM with Channel Estimation

Jianxuan Du
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250, USA
Email: jxdu@ece.gatech.edu

Ye (Geoffrey) Li
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250, USA
Email: liye@ece.gatech.edu

Received 28 January 2003; Revised 26 September 2003

Multiple-input and multiple-output (MIMO) systems formed by multiple transmit and receive antennas can improve performance and increase capacity of wireless communication systems. Diagonal Bell Laboratories Layered Space-Time (D-BLAST) structure offers a low-complexity solution for realizing the attractive capacity of MIMO systems. However, for broadband wireless communications, channel is frequency-selective and orthogonal frequency division multiplexing (OFDM) has to be used with MIMO techniques to reduce system complexity. In this paper, we investigate D-BLAST for MIMO-OFDM systems. We develop a layerwise channel estimation algorithm which is robust to channel variation by exploiting the characteristic of the D-BLAST structure. Further improvement is made by subspace tracking to considerably reduce the error floor. Simulation results show that the layerwise estimators require 1 dB less signal-to-noise ratio (SNR) than the traditional blockwise estimator for a word error rate (WER) of 10^{-2} when Doppler frequency is 40 Hz. Among the layerwise estimators, the subspace-tracking estimator provides a 0.8 dB gain for 10^{-2} WER with 200 Hz Doppler frequency compared with the DFT-based estimator.

Keywords and phrases: MIMO, OFDM, channel estimation.

1. INTRODUCTION

Multiple-input and multiple-output (MIMO) systems formed by multiple transmit and receive antennas are under intense research recently for its attractive potential to offer great capacity increase. Space-time coding, proposed in [1], performs channel coding across the space and time to exploit the spatial diversity offered by MIMO systems to increase system capacity. However, the decoding complexity of the space-time codes is exponentially increased with the number of transmit antennas, which makes it hard to implement real-time decoding as the number of antennas grows. To reduce the complexity of space-time based MIMO systems, diagonal Bell Laboratories layered space-time (D-BLAST) architecture has been proposed in [2]. Rather than try to optimize channel coding scheme, in D-BLAST architecture, the input data stream is divided into several substreams. Each substream is encoded independently using one-dimensional coding and the association of output stream with transmit antennas is periodically cycled to explore spatial diversity.

Orthogonal frequency division multiplexing (OFDM) systems have the desirable immunity to intersymbol interference (ISI) caused by delay spread of wireless channels. Therefore, the combination of D-BLAST with OFDM is an attractive technique for high-speed transmission over frequency-selective fading wireless channels. As in [3], when combining D-BLAST structure with OFDM, we implement the space-time structure in space-frequency domain to avoid decoding delay. To decode each layer, channel parameters are used to cancel interference from detected signals and suppress interference from undetected signals to make the desired signal as "clean" as possible. Therefore, estimation of channel parameters is a prerequisite for realizing D-BLAST structure and to a great extent determines system performance. In this paper, we investigate D-BLAST OFDM systems and address the channel estimation problem.

DFT-based least-square (LS) channel estimation for MIMO-OFDM systems and simplified estimation algorithm using parallel interference cancellation have been addressed in [4, 5], respectively. For D-BLAST OFDM, we propose a layerwise LS channel estimator that exploits the characteristics of the system structure by updating channel parameters after each layer is detected so that later layers in the same OFDM block can be detected with more accurate channel state information.

In spite of low complexity of DFT-based channel estimators, there is leakage when the multipaths are not exactly sample spaced [6], which induces an error floor for channel estimation. To reduce the error floor of DFT-based algorithm and increase estimation accuracy, more taps have

to be used. Consequently, the estimation problem becomes ill-conditioned and noise may be enhanced. To improve the channel estimation accuracy for D-BLAST OFDM, we use optimum training sequences in [5, 7] not only for initial channel estimation but also for tracking channel autocorrelation matrix and then its dominant eigenvectors. The resultant eigenvectors are then used to form a transform which requires fewer taps to be estimated and reduces the error floor. The low-rank adaptive filter 1 (LORAF 1) in [8] is used for subspace tracking. For both proposed estimators, further refinement can be achieved by a robust filter [9] to exploit time-domain correlation.

The rest of this paper is organized as follows. In Section 2, we introduce D-BLAST OFDM systems. Then, in Section 3, we derive a layerwise LS channel estimator and analyze the mean square error (MSE) performance. Next, in Section 4, we propose an improved channel estimator based on subspace tracking. In Section 5, we evaluate the performance of a D-BLAST OFDM system with different channel estimation algorithms by computer simulation and major results of the paper are summarized in Section 6.

2. D-BLAST OFDM SYSTEM

Before introducing the channel estimation algorithm, we briefly describe D-BLAST for MIMO-OFDM in this section.

The complex baseband representation of a delay spread channel can be expressed as [10]

$$h(t,\tau) = \sum_{l} \alpha_l(t)\Delta(\tau - \tau_l), \tag{1}$$

where $\alpha_l(t)$'s are wide-sense stationary narrowband complex Gaussian processes and are assumed to be independent among different paths. The channel may vary from block to block but stays the same within each OFDM block, which means that the effect of intercarrier interference (ICI) is not considered. Moreover, we assume the same normalized time-domain correlation function for all paths, that is,

$$E\{\alpha_l(t+\Delta t)\alpha_m^*(t)\} = \begin{cases} \sigma_l^2 r_t(\Delta t), & l = m, \\ 0, & l \neq m. \end{cases} \tag{2}$$

Without loss of generality, we assume the total average power of the channel impulse response to be unity, that is,

$$\sum_{l} \sigma_l^2 = 1. \tag{3}$$

For a MIMO-OFDM system with N_t transmit and N_r receive antennas ($N_r \geq N_t$), the received signal at the kth subcarrier of the nth block from the jth receive antenna can be expressed as

$$x_j[n,k] = \sum_{i=1}^{N_t} H_{ij}[n,k]b_i[n,k] + w_j[n,k], \tag{4}$$

for $j = 1,\ldots,N_r$ and $k = 0,\ldots,K-1$, where K is the total number of subcarriers of OFDM, $b_i[n,k]$ is the symbol transmitted from the ith transmit antenna at the kth subcarrier of the nth block, $H_{ij}[n,k]$ is the channel's frequency response at the kth subcarrier of the nth block corresponding to the ith transmit and the jth receive antenna, and $w_j[n,k]$ is additive (complex) Gaussian noise that is assumed to be independent and identically distributed (i.i.d.) with zero-mean and variance ρ.

Equation (4) can also be written in matrix form as

$$\mathbf{x}[n,k] = \mathbf{H}[n,k]\mathbf{b}[n,k] + \mathbf{w}[n,k], \tag{5}$$

where

$$\mathbf{x}[n,k] = \begin{pmatrix} x_1[n,k] \\ \vdots \\ x_{N_r}[n,k] \end{pmatrix},$$

$$\mathbf{H}[n,k] = \begin{pmatrix} H_{11}[n,k] & H_{21}[n,k] & \cdots & H_{N_t1}[n,k] \\ H_{12}[n,k] & H_{22}[n,k] & \cdots & H_{N_t2}[n,k] \\ \vdots & \cdots & \ddots & \vdots \\ H_{1N_r}[n,k] & \cdots & \cdots & H_{N_tN_r}[n,k] \end{pmatrix}, \tag{6}$$

$$\mathbf{b}[n,k] = \begin{pmatrix} b_1[n,k] \\ \vdots \\ b_{N_t}[n,k] \end{pmatrix},$$

$$\mathbf{w}[n,k] = \begin{pmatrix} w_1[n,k] \\ \vdots \\ w_{N_r}[n,k] \end{pmatrix}.$$

D-BLAST is an effective MIMO technique [2] that has been originally developed for a single-carrier system with flat fading channel. In this paper, we will use this technique for a MIMO-OFDM system, which can be shown in Figure 1. From the figure, the set of all subcarriers in an OFDM block is divided into N_t subsets, each with $L = K/N_t$ subcarriers. Each layer, composed of N_t such subsets associated with different transmit antennas, is encoded and decoded independently. Note that each layer still has K subcarriers, but different subcarriers may be associated with different transmit antennas. Layers starting at block n are denoted as $L_p[n]$, $p = 1,2,\ldots,N_t$. With some abuse of notations, k is both the subcarrier index and the symbol index for each layer. Given the structure of D-BLAST OFDM, the received signal at each receive antenna is the superposition of the desired signal, the signals already detected in the previous layers, and those undetected.

The signal detection of D-BLAST MIMO-OFDM is also very similar to the original D-BLAST. Assume that layer $L_p[n]$, $p = 1,2,\ldots,N_t$, is to be detected. From (4) we have

$$\begin{aligned} x_j[g_p(n,k),k] = {} & H_{f_p(k),j}[g_p(n,k),k]b_{f_p(k)}[g_p(n,k),k] \\ & + \sum_{i=1}^{f_p(k)-1} H_{ij}[g_p(n,k),k]b_i[g_p(n,k),k] \\ & + \sum_{i=f_p(k)+1}^{N_t} H_{ij}[g_p(n,k),k]b_i[g_p(n,k),k] \\ & + w_j[g_p(n,k),k], \end{aligned} \tag{7}$$

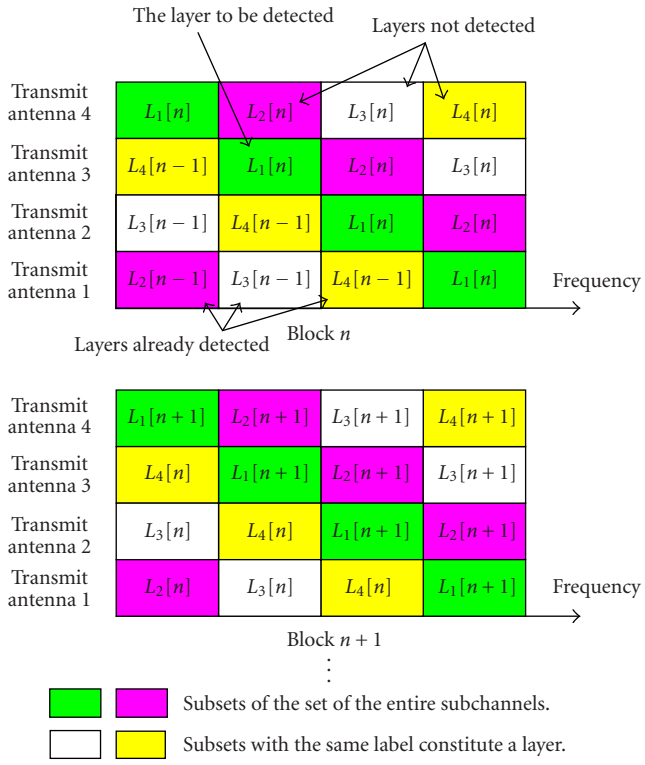


Figure 1: D-BLAST MIMO-OFDM structure.

for $j = 1, 2, \ldots, N_r$ and $k = 0, \ldots, K-1$, where $f_p(k)$ and $g_p(n,k)$ are associations of the kth symbol of layer $L_p[n]$ with transmit antenna and OFDM block, respectively, that is, the kth symbol of layer $L_p[n]$ is sent from the $f_p(k)$th transmit antenna via the kth subcarrier of the $g_p(n,k)$th OFDM block. Note that, in general, a layer spans two consecutive OFDM blocks, thus $g_p(n,k)$ is either n or $n+1$. Equation (7) can be written in matrix notation as

$$\begin{aligned}\mathbf{x}[g_p(n,k),k] &= \mathbf{H}_{f_p(k)}[g_p(n,k),k]\,b_{f_p(k)}[g_p(n,k),k] \\ &\quad + \sum_{i=1}^{f_p(k)-1} \mathbf{H}_i[g_p(n,k),k]\,b_i[g_p(n,k),k] \\ &\quad + \sum_{i=f_p(k)+1}^{N_t} \mathbf{H}_i[g_p(n,k),k]\,b_i[g_p(n,k),k] \\ &\quad + \mathbf{w}[g_p(n,k),k],\end{aligned} \tag{8}$$

where $\mathbf{H}_i[n,k]$ is the ith column of $\mathbf{H}[n,k]$. Signals from antennas 1 to $f_p(k)-1$ have been detected and those from antennas $f_p(k)+1$ to N_t are yet to be detected.

First, interference cancellation is carried out by subtracting detected signals from the received signal:

$$\begin{aligned}\tilde{\mathbf{x}}_p[g_p(n,k),k] &= \mathbf{x}[g_p(n,k),k] \\ &\quad - \sum_{i=1}^{f_p(k)-1} \mathbf{H}_i[g_p(n,k),k]\,\hat{b}_i[g_p(n,k),k],\end{aligned} \tag{9}$$

where $\hat{b}_i[n,k]$'s are detected symbols. Then interference from undetected signals is suppressed by linear combination that yields the maximum signal-to-interference-plus-noise ratio (SINR). Let

$$\tilde{\mathbf{H}}_p[n,k] \triangleq (\mathbf{H}_{f_p(k)+1}[n,k], \mathbf{H}_{f_p(k)+2}[n,k], \ldots, \mathbf{H}_{N_t}[n,k]), \tag{10}$$

then from [11], we have the following weighting vector:

$$\mathbf{v}_p[n,k] = \left(\tilde{\mathbf{H}}_p[n,k]\tilde{\mathbf{H}}_p^H[n,k] + \rho\mathbf{I}\right)^{-1}\mathbf{H}_{f_p(k)}[n,k]. \tag{11}$$

Thus, if we assume Gaussian distribution for the residual interference plus noise, the maximum likelihood decoding of layer $L_p[n]$ is to find $\{\hat{b}_{f_p(k)}[g_p(n,k),k]\}$ that minimizes

$$\begin{aligned}&M(\{\hat{b}_{f_p(k)}[g_p(n,k),k];\ k = 0,1,\ldots,K-1\}) \\ &= \sum_{k=0}^{K-1} \frac{1}{\mathbf{v}_p^H[m,k](\tilde{\mathbf{H}}_p[m,k]\tilde{\mathbf{H}}_p^H[m,k] + \rho\mathbf{I})\mathbf{v}_p[m,k]} \\ &\quad \cdot \left|\mathbf{v}_p^H[m,k](\tilde{\mathbf{x}}_p[m,k] - \mathbf{H}_{f_p(k)}[m,k]\hat{b}_{f_p(k)}[m,k])\right|^2\Big|_{m=g_p(n,k)}\end{aligned} \tag{12}$$

which can be solved by standard Viterbi algorithm when convolutional codes are used. From the above discussions, channel information is crucial for the signal detection of D-BLAST MIMO-OFDM. Therefore, we focus on channel estimation in the paper.

3. LAYERWISE CHANNEL ESTIMATION

In this section, we develop a layerwise LS channel estimation algorithm and analyze its performance.

3.1. Layerwise least-square channel estimation

Due to layerwise detection in D-BLAST, usually only partial knowledge of the symbols transmitted from all transmit antennas at one OFDM block is available after decoding of each layer. To exploit the characteristics of D-BLAST structure, channel estimation is carried out each time a layer is detected. Since channel responses are independent among different transmit-receive antenna pairs, we consider the channel estimation for one particular receive antenna and omit the receive antenna subscript j in (4) to get

$$x[n,k] = \sum_{i=1}^{N_t} H_i[n,k]\,b_i[n,k] + w[n,k]. \tag{13}$$

After detection of layer $L_p[n]$, we estimate the channel responses at the nth block. Since only part of all the subcarriers of the current OFDM block have signals from all transmit antennas detected, we replace the received signals at subcarriers not fully detected with those of the previous OFDM block to form a complete received signal vector, due to the fact that

$H_i[n,k] \approx H_i[n-1,k]$. Define

$$z^{(p)}[n,k] = \begin{cases} x[n-1,k], & k \in \Sigma^{(p)}, \\ x[n,k], & \text{else}, \end{cases}$$
$$\hat{d}_i^{(p)}[n,k] \triangleq \begin{cases} \hat{b}_i[n-1,k], & k \in \Sigma^{(p)}, \\ \hat{b}_i[n,k], & \text{else}, \end{cases} \tag{14}$$

where $\Sigma^{(p)} = \{k; g_p(n,k) = n \text{ and } f_p(k) < N_t\}$ is the set of subcarriers with signals not fully detected.

It is observed that with some leakage [6], channel frequency response can be approximated as

$$H_i[n,k] = \sum_{l=0}^{\chi-1} h_i[n,l]\, W_K^{kl}, \tag{15}$$

where $W_K = e^{-j(2\pi/K)}$, $\chi \geq \lceil t_d/t_s \rceil$, t_d is the maximum delay spread and t_s is the sampling interval which is equal to $1/K\Delta f$ with Δf being the tone spacing.

Let

$$\mathbf{z}^{(p)}[n] = \left(z^{(p)}[n,0], \ldots, z^{(p)}[n,K-1]\right)^T, \tag{16}$$

$$\hat{\mathbf{D}}_i^{(p)}[n] = \operatorname{diag}\{\hat{d}_i^{(p)}[n,0], \ldots, \hat{d}_i^{(p)}[n,K-1]\}, \tag{17}$$

$$\mathbf{U} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & W_K & \cdots & W_K^{\chi-1} \\ \vdots & \cdots & \ddots & \vdots \\ 1 & W_K^{(K-1)} & \cdots & W_K^{(\chi-1)(K-1)} \end{pmatrix}, \tag{18}$$

$$\hat{\mathbf{h}}^{(p)}[n] \triangleq \left(\hat{\mathbf{h}}_1^{(p)^T}[n], \hat{\mathbf{h}}_2^{(p)^T}[n], \ldots, \hat{\mathbf{h}}_{N_t}^{(p)^T}[n]\right)^T, \tag{19}$$

$$\hat{\mathbf{h}}_i^{(p)}[n] \triangleq \left(\hat{h}_i^{(p)}[n,0], \ldots, \hat{h}_i^{(p)}[n,\chi-1]\right)^T. \tag{20}$$

The LS channel estimation is to minimize the following cost function [4]:

$$C\left(\{\hat{\mathbf{h}}_i^{(p)};\ i = 1,2,\ldots,N_t-1\}\right) = \left\| \mathbf{z}^{(p)}[n] - \sum_{i=1}^{N_t} \hat{\mathbf{D}}_i^{(p)}[n]\mathbf{U}\hat{\mathbf{h}}_i^{(p)}[n] \right\|^2. \tag{21}$$

Then

$$\hat{\mathbf{h}}^{(p)}[n] = \left(\mathbf{T}^{(p)^H}[n]\mathbf{T}^{(p)}[n]\right)^{-1}\mathbf{T}^{(p)^H}[n]\mathbf{z}^{(p)}[n], \tag{22}$$

where

$$\mathbf{T}^{(p)}[n] = \left(\hat{\mathbf{D}}_1^{(p)}[n]\mathbf{U}, \hat{\mathbf{D}}_2^{(p)}[n]\mathbf{U}, \ldots, \hat{\mathbf{D}}_{N_t}^{(p)}[n]\mathbf{U}\right). \tag{23}$$

The above estimate is further refined by applying a robust estimator for OFDM systems in [9], which makes full use of the correlation of channel parameters at different OFDM blocks.

3.2. Performance analysis

Here, we briefly analyze the performance of the above channel estimator for D-BLAST OFDM. Let

$$\Delta\mathbf{h} \triangleq \mathbf{h}[n] - \mathbf{h}[n-1],$$
$$\hat{s}_i^{(p)}[n,k] \triangleq \begin{cases} \hat{b}_i[n-1,k], & k \in \Sigma^{(p)}, \\ 0, & \text{else}, \end{cases}$$
$$\hat{\mathbf{S}}_i^{(p)}[n] = \operatorname{diag}\{\hat{s}_i^{(p)}[n,0], \ldots, \hat{s}_i^{(p)}[n,K-1]\}, \tag{24}$$
$$\mathbf{G}^{(p)}[n] \triangleq \left(\mathbf{T}^{(p)^H}[n]\mathbf{T}^{(p)}[n]\right)^{-1}\mathbf{T}^{(p)^H}[n]\hat{\mathbf{S}}_i^{(p)}[n].$$

The MSE of the channel estimator is

$$\begin{aligned} \mathrm{MSE}^{(p)}[n] &\triangleq \frac{1}{N_t\chi} E\left\{ \left\| \hat{\mathbf{h}}^{(p)}[n] - \mathbf{h}[n] \right\|^2 \right\} \\ &= \frac{1}{N_t\chi} \mathbf{Tr}\Big\{ \rho\left(\mathbf{T}^{(p)^H}[n]\mathbf{T}^{(p)}[n]\right)^{-1} \\ &\qquad + \mathbf{G}^{(p)}[n] E\left(\Delta\mathbf{h}[n]\Delta\mathbf{h}^H[n]\right)\mathbf{G}^{(p)^H}[n] \Big\}, \end{aligned} \tag{25}$$

where $E\{\cdot\}$ denotes expected value of a random variable. Clearly the first term in the above equation results from noise and the second term is due to channel variation.

From the above discussion, the MSE of the channel estimate depends on the inverse of $\mathbf{T}^{(p)^H}[n]\mathbf{T}^{(p)}[n]$, which relates to the condition number of $\mathbf{T}^{(p)^H}[n]\mathbf{T}^{(p)}[n]$. It can be proved in the appendix using the bordering theorem for Hermitian matrices [12] that condition number of $\mathbf{T}^{(p)^H}[n]\mathbf{T}^{(p)}[n]$ increases with χ. It implies that the channel estimation becomes more ill-conditioned as the number of parameters to be estimated increases. Thus we should choose the number of parameters as small as possible while preserving energy of the channel response, which is the reason for tracking the optimum transform matrix $\mathbf{U}$ in Section 4.

4. SUBSPACE TRACKING

The major problem of decision-directed channel estimation is the randomness of the symbol sequences during data transmission mode. For example, when the symbol sequences from any two of the transmit antennas are the same or very close, it is impossible or very hard to distinguish channel responses corresponding to different transmit antennas. The greater the number of transmit antennas, the more likely the channel is unidentifiable, or the more ill-conditioned channel identification is. Furthermore, to reduce the leakage of decision-directed DFT-based channel estimation in MIMO-OFDM systems, the number of taps representing channel frequency response has to be increased, which will make channel identification more ill-conditioned at the same time, as shown in Section 3. Moreover, increasing the number of taps makes the inverse operation of matrices in (22) more complicated. Hence, it is essential for low-complexity and high-performance channel estimator to reduce the number of parameters to be estimated while preserving most of the energy of channel frequency responses

during the data transmission mode. Therefore, we will develop subspace tracking approaches to estimate channel parameters. And since the subspace only depends on channel autocorrelation matrix, which is time-invariant or changing very slowly, we apply subspace tracking only to training blocks and use the derived transform matrix instead of $\mathbf{U}$ defined in (18) for channel estimation during data transmission mode.

Let the $K \times K$ channel autocorrelation matrix $(\mathbf{R}_f)_{k_1,k_2} = E\{H[k_1]H^*[k_2]\}$ have singular value decomposition as follows:

$$\mathbf{R}_f = \mathbf{U}_f \Lambda \mathbf{U}_f^H, \tag{26}$$

where $\mathbf{U}_f$ is a $K \times K$ unitary matrix and $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_K\}$, $\lambda_1 \geq \lambda_2 \geq \cdots > 0$. From [13], optimum rank-χ estimator is to select eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_\chi$ corresponding to the χ biggest eigenvalues. Then the optimum rank-χ transform matrix is

$$\mathbf{U}_{\text{opt}} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_\chi). \tag{27}$$

Therefore, channel autocorrelation matrix is needed here for the optimum low-rank channel estimation.

To obtain the channel autocorrelation matrix, first we have to separate channel responses $H_i[n,k]$'s for different i's. This can be done by appropriately designing the training block. In [5, 7], optimal training sequences have been proposed to maximally separate frequency responses of different transmit antennas while preserving most of the energy of each channel response. The training sequences are

$$b_i[n,k] = b_1[n,k]\, W_K^{-K_0(i-1)k}, \tag{28}$$

for $i = 1, 2, \ldots, N_t$, where $K_0 = \lfloor K/N_t \rfloor \geq \lceil t_d/t_s \rceil$ is the number of taps used to represent the channel response as a DFT transform. During the training period, we choose K_0 taps in approximating the channel response according to (15). Since the leakage introduced by the DFT-based approximation is decreased as K_0 increases and the well-designed training sequences provide maximum separability, we can set K_0 to be big enough such that the leakage is negligible while introducing little aliasing between different channel responses [5]. The procedure to separate channel responses can be described in Algorithm 1.

The dimension χ of the subspace can be either determined by minimum description length (MDL) criterion [14] that is not accurate for low signal-to-noise ratio (SNR) or slow channel variation, or by the approach in [13, 15] which argues that the essential dimension of a random signal is about the product of the bandwidth and time interval of the signal plus one. We just choose the latter approach for its simplicity and effectiveness; therefore, $\chi = \lceil t_d/t_s \rceil$. Subspace tracking approach can be summarized, which is in Algorithm 2 modified LORAF 1 in [8].

It should be noted that the robust channel estimator depends only on the subspace spanned by the dominant eigenvectors rather than the particular eigenvectors. Let

$$\hat{\mathbf{U}}_i[n] = \mathbf{U}_{\text{opt}} \mathbf{Q}_i[n], \tag{29}$$

where $\mathbf{Q}_i[n]$ is a $\chi \times \chi$ unitary matrix which accounts for the change of dominant eigenvectors without changing the subspace. Substituting (29) into (25), it can be easily verified that the MSE of the channel estimator is invariant to rotation of the dominant eigenvectors, which can also be seen in [9]. Therefore, it is the dominant subspace spanned by channel frequency responses that affects the performance of the subspace tracking-based channel estimator.

(a) During each training block, $\eta[n,k] = x[n,k] \cdot b_1^*[n,k]$.
(b) Perform IFFT on $(\eta[n,0], \eta[n,1], \ldots, \eta[n,K-1])$ to get $(\zeta[n,0], \zeta[n,1], \ldots, \zeta[n,K-1])$.
(c) For the channel response of transmit antenna i, circularly left shift $(\zeta[n,0], \zeta[n,1], \ldots, \zeta[n,K-1])$ by $(i-1)K_0$ to get $(\zeta'[n,0], \zeta'[n,1], \ldots, \zeta'[n,K-1])$.
Let $\zeta''[n,k] \triangleq \begin{cases} \zeta'[n,k], & k \in [0, K_0 - 1], \\ 0, & \text{else.} \end{cases}$
(d) A channel estimate $\hat{H}_i[n,k]$ is obtained by performing FFT on $(\zeta''[n,0], \zeta''[n,1], \ldots, \zeta''[n,K-1])$.

Algorithm 1: Channel separation using the optimum training sequences.

Initialization:
$(\mathbf{U}_i[0])_{k,l} = W_K^{kl}/\sqrt{K}$, $0 \leq k \leq K-1$, $0 \leq l \leq \chi - 1$;
$\Phi[0] = \mathbf{I}$, $0 \leq \alpha \leq 1$;
During each training block:
input $\mathbf{v}_i[n] = (\hat{H}_i[n,0], \hat{H}_i[n,1], \ldots, \hat{H}_i[n,K-1])^T$,
$\mathbf{c}_i[n] = \mathbf{U}_i^H[n-1]\mathbf{v}_i[n]$,
$\mathbf{A}_i[n] = \alpha \mathbf{A}_i[n-1]\Phi_i[n-1] + (1-\alpha)\mathbf{v}_i[n]\mathbf{c}_i^H[n]$,
$\mathbf{A}_i[n] = \mathbf{U}_i[n]\mathbf{R}_i[n]$ QR decomposition,
$\Phi_i[n] = \mathbf{U}_i^H[n-1]\mathbf{U}_i[n]$,
Low-rank channel approximation:
$\mathbf{v}_i[n] = \mathbf{U}_i[n]\mathbf{h}_i[n]$

Algorithm 2: Subspace tracking for channel estimation.

5. SIMULATION RESULTS

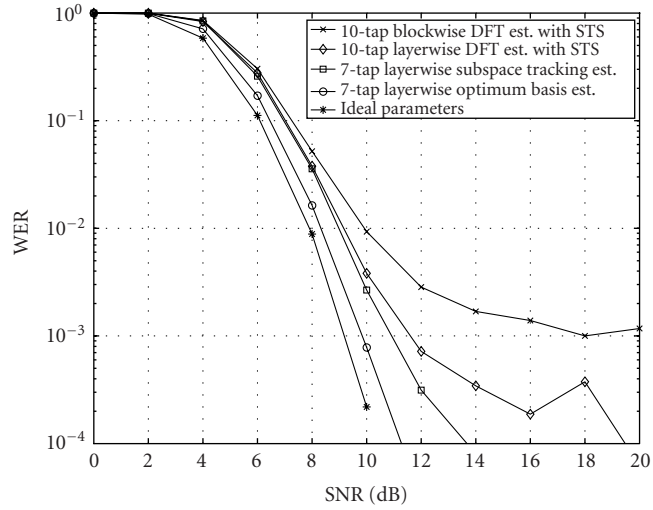
In this section, we evaluate the performance of different decision-directed channel estimation algorithms for D-BLAST OFDM by computer simulation. Typical-urban (TU) channel with Doppler frequency $f_d = 40$ and 200 Hz is used in our simulation. Performance of the proposed 7-tap layerwise subspace tracking estimator is simulated. As a comparison, performances of systems with ideal channel parameters, 7-tap layerwise estimator with optimum transform, as defined in (27), and 10-tap layerwise DFT-based estimator with significant tap selection (STS) [4] are evaluated. The performance of the traditional 10-tap blockwise DFT-based channel estimator is also given, where channel estimation is carried out once per OFDM block and the estimated channel parameters are used for the detection of the next OFDM block.

Four transmit antennas and four receive antennas are employed to form four D-BLAST layers. Channel parameters corresponding to different transmit and receive antenna pairs are assumed to be independent but have the same statistics. The system bandwidth of 1.25 MHz is divided into 256 subchannels: 2 subchannels on each side are used as guard tones, and the rest of the subchannels are used for data transmission. The symbol duration is 204.8 μs and another 20.2 μs is added as cyclic prefix (CP), resulting in a total block duration of 225 μs. A 16-state binary-to-4-ary convolutional codes of rate 1/2 with the octal generators being (26, 37) [16] is used to encode the information bits in each layer. Four tail bits are used for trellis termination, leaving 248 information bits per layer. The encoder output is interleaved before sending to a transmit antenna at a particular subcarrier.

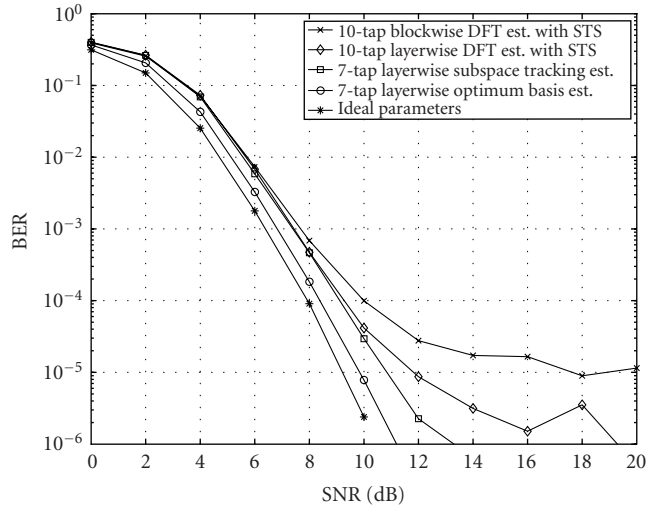
In each independent simulation, 2000 OFDM blocks of data are transmitted with 1 training block sent every 10 blocks. The performance averaged over independent simulations is evaluated. For channel estimator with subspace tracking, the first 50 blocks use 10-tap DFT-based estimator with STS. The estimated channel parameters are used for initial subspace acquisition so that initial training overhead can be saved at the expense of negligible performance loss for continuous data transmission. Then channel estimation is switched to the estimator with subspace tracking and the subspace is updated at each training block. The forgetting factor α is chosen to be 0.995.

Figures 2a and 2b compare the word error rate (WER) and bit-error-rate (BER) performance of different channel estimation algorithms when Doppler frequency is 40 Hz. Of all estimators, the blockwise DFT-based channel estimator has the worst performance since it uses channel state information at the previous OFDM block for detection and thus is most susceptible to channel variation. The blockwise DFT-based estimator requires about 1 dB more SNR than the layerwise DFT-based estimator for a WER of 10^{-2} and its WER curve levels off quickly at high SNR's since its performance is bounded by channel variation. Among layerwise estimators, the subspace tracking estimator, requires 0.7 dB less SNR than the DFT-based estimator for 10^{-3} WER. Figure 2c shows how MSE evolves as the layerwise channel estimation progresses. From the figure, we can see that for all layerwise channel estimation methods, the most significant MSE improvement is seen after detection of the first layer of the current OFDM block, which is 0.7 dB at SNR = 16 dB, compared with about 0.16 dB per layer improvement for layers detected later with the proposed subspace tracking channel estimator.

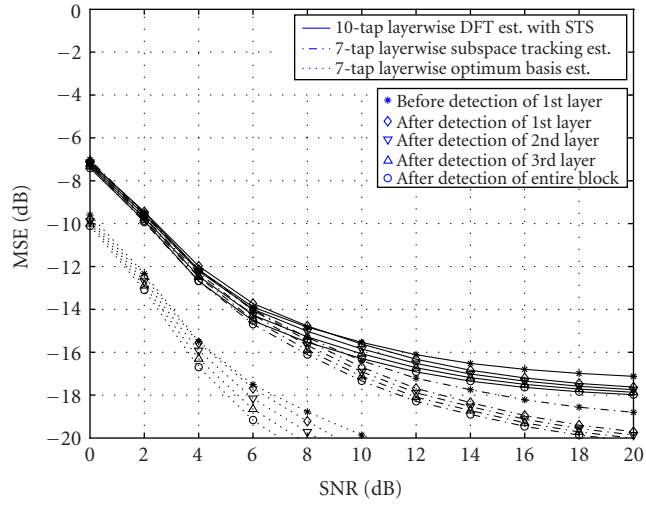
For $f_d = 200$ Hz, from Figure 3 we see that the performance difference between the blockwise channel estimator and layerwise estimators is even bigger now that the system performance is dominated by fast variation of channel parameters. The SNR gain for using layerwise subspace tracking estimator is about 0.8 dB for 10^{-2} WER compared with layerwise DFT-based estimator. It is clear that as the channel variation rate increases, the MSE performance improvement with layerwise channel estimation becomes more significant, with the successive MSE improvements being 3.4 dB, 1.2 dB,



(a)

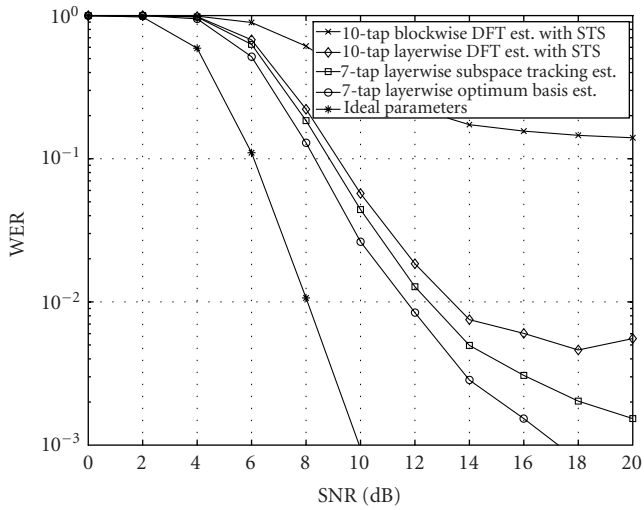


(b)

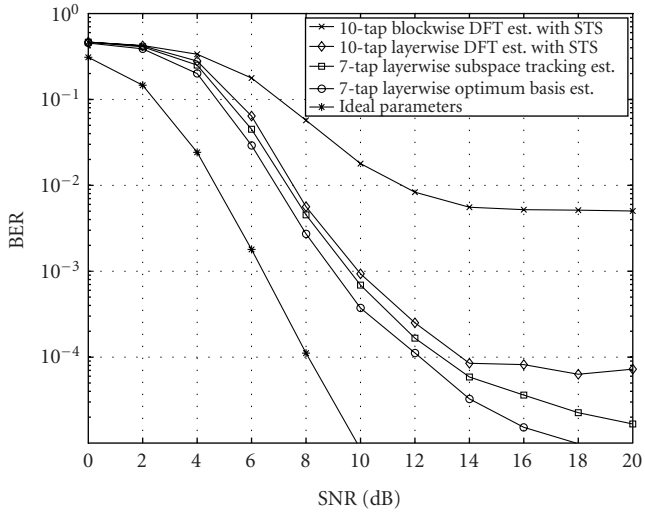


(c)

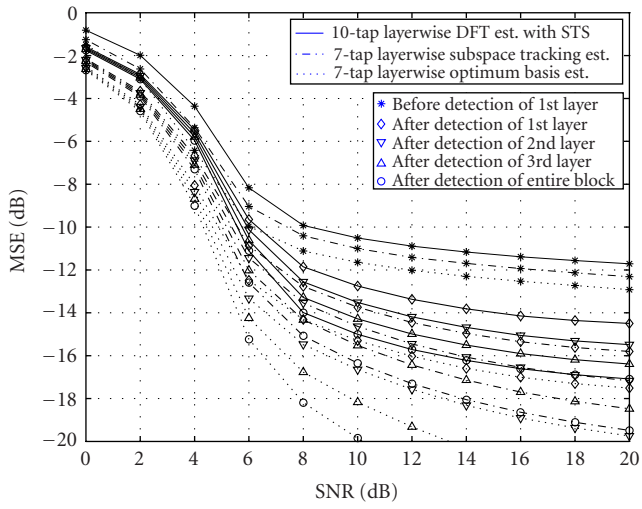
Figure 2: (a) WER, (b) BER, and (c) MSE of D-BLAST systems for channels with TU delay profile and $f_d = 40$ Hz.



(a)



(b)



(c)

Figure 3: (a) WER, (b) BER, and (c) MSE of D-BLAST systems for channels with TU delay profile and $f_d = 200$ Hz.

1.2 dB, and 1 dB at SNR = 16 dB, as observed in Figure 3c. For both $f_d = 40$ and 200 Hz, the subspace tracking estimator can effectively reduce the error floor thus provide better performance than that of the DFT-based estimator.

6. CONCLUSION

MIMO-OFDM is a promising technology that embraces advantages of both MIMO system and OFDM, that is, immunity to delay spread as well as huge transmission capacity. In this paper, we apply the D-BLAST structure to MIMO-OFDM systems and develop a channel estimator that updates the estimated channel parameters in a layerwise fashion. Since we update channel estimation using detected signals to improve detection of the rest of the signals in the current OFDM block, the system is more robust to fast fading channels when compared with the traditional blockwise estimator. To further reduce the channel estimation error, we use the training blocks not only for channel estimation, but also for tracking of the dominant subspace spanned by the channel frequency response to reduce the number of parameters to be estimated during data transmission mode. Thus, additional performance improvement is obtained by using subspace tracking for the layerwise estimator, which is about 0.8 dB for 10^{-2} WER with $f_d = 200$ Hz.

APPENDIX

Let $\mathbf{U}' = (\mathbf{U}, \mathbf{u}_{\chi+1})$, and from $\mathbf{U}'$, we define $\mathbf{T}^{(p)'}[n]$ as in (23) by substituting $\mathbf{U}'$ for $\mathbf{U}$. We will show that

$$\operatorname{cond}\left(\mathbf{T}^{(p)'^H}[n]\mathbf{T}^{(p)'}[n]\right) \geq \operatorname{cond}\left(\mathbf{T}^{(p)^H}[n]\mathbf{T}^{(p)}[n]\right), \quad \text{(A.1)}$$

where $\operatorname{cond}(\cdot)$ means condition number of a matrix.

Proof. Let the eigenvalues of $\mathbf{T}^{(p)^H}[n]\mathbf{T}^{(p)}[n]$ be $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_{\chi N_t} > 0$. From (23) and by exchanging columns which does not change the eigenvalues, we have

$$\mathbf{T}^{(p)'}[n] = \left(\mathbf{T}^{(p)}[n], \mathbf{Y}^{(p)}[n]\right), \quad \text{(A.2)}$$

where

$$\mathbf{Y}^{(p)}[n] = \left(\hat{\mathbf{D}}_1^{(p)}[n]\mathbf{u}_{\chi+1}, \ldots, \hat{\mathbf{D}}_{N_t}^{(p)}[n]\mathbf{u}_{\chi+1}\right),$$
$$\mathbf{T}^{(p)'^H}[n]\mathbf{T}^{(p)'}[n] = \begin{pmatrix} \mathbf{T}^{(p)^H}[n]\mathbf{T}^{(p)}[n] & \mathbf{T}^{(p)^H}[n]\mathbf{Y}^{(p)}[n] \\ \mathbf{Y}^{(p)^H}[n]\mathbf{T}^{(p)}[n] & \mathbf{Y}^{(p)^H}[n]\mathbf{Y}^{(p)}[n] \end{pmatrix}. \quad \text{(A.3)}$$

Let the eigenvalues of $\mathbf{T}^{(p)'^H}[n]\mathbf{T}^{(p)'}[n]$ be $\gamma'_1 \geq \gamma'_2 \geq \cdots \geq \gamma'_{(\chi+1)N_t} > 0$. By the bordering theorem for Hermitian matrices [12], we have

$$\gamma'_1 \geq \gamma_1 \geq \gamma_{\chi N_t} \geq \gamma'_{(\chi+1)N_t} > 0, \quad \text{(A.4)}$$

thus

$$\operatorname{cond}\left(\mathbf{T}^{(p)'^H}[n]\mathbf{T}^{(p)'}[n]\right) = \frac{\gamma'_1}{\gamma'_{(\chi+1)N_t}} \geq \frac{\gamma_1}{\gamma_{\chi N_t}} = \operatorname{cond}\left(\mathbf{T}^{(p)^H}[n]\mathbf{T}^{(p)}[n]\right). \quad \text{(A.5)}$$

□

ACKNOWLEDGMENT

This work was jointly supported by the National Science Foundation (NSF) under Grant CCR-0121565 and Nortel networks.

REFERENCES

[1] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.

[2] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[3] J. X. Du and Y. G. Li, "Channel estimation for D-BLAST OFDM systems," in *Proc. IEEE Global Telecommunications Conference*, vol. 1, pp. 335–339, Taipei, Taiwan, November 2002.

[4] Y. Li, N. Seshadri, and S. Ariyavisitakul, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 3, pp. 461–471, 1999.

[5] Y. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, pp. 67–75, 2002.

[6] J. J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM systems," in *IEEE 45th Vehicular Technology Conference (VTC '95)*, vol. 2, pp. 815–819, Chicago, Ill, USA, July 1995.

[7] I. Barhumi, G. Leus, and M. Moonen, "Optimal training sequences for channel estimation in MIMO OFDM systems in mobile wireless channels," in *Proc. IEEE International Zurich Seminar on Broadband Communications, Access, Transmission, Networking*, pp. 44/1–44/6, Zurich, Switzerland, February 2002.

[8] P. Strobach, "Low-rank adaptive filters," *IEEE Trans. Signal Processing*, vol. 44, no. 12, pp. 2932–2947, 1996.

[9] Y. Li, L. J. Cimini, and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Transactions on Communications*, vol. 46, no. 7, pp. 902–915, 1998.

[10] R. Steele, *Mobile Radio Communications*, IEEE Press, New York, NY, USA, 1992.

[11] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*, John Wiley & Sons, New York, NY, USA, 1980.

[12] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.

[13] O. Edfors, M. Sandell, J. J. van de Beek, S. K. Wilson, and P. O. Borjesson, "OFDM channel estimation by singular value decomposition," *IEEE Transactions on Communications*, vol. 46, no. 7, pp. 931–939, 1998.

[14] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.

[15] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, fourier analysis and uncertainty-III: the dimension of the space of essentially time- and band-limited signals," *Bell System Technical Journal*, vol. 41, pp. 1295–1336, 1962.

[16] W. E. Ryan and S. G. Wilson, "Two classes of convolutional codes over GF(q) for q-ary orthogonal signaling," *IEEE Transactions on Communications*, vol. 39, no. 1, pp. 30–40, 1991.

Jianxuan Du obtained his B.S. and M.S. degrees in electrical engineering in 1998 and 2001, respectively, from Xi'an Jiaotong University, China. Since 2001, he has been pursuing the Ph.D. degree in electrical and computer engineering at Georgia Institute of Technology, Ga. He is currently a Research Assistant in Information Transmission and Processing Laboratory at Georgia Institute of Technology. His research interests include signal processing for wireless communications, channel estimation, and MIMO-OFDM systems.

Ye (Geoffrey) Li received his B.S.E. and M.S.E. degrees in 1983 and 1986, respectively, from the Department of Wireless Engineering, Nanjing Institute of Technology, Nanjing, China, and his Ph.D. degree in 1994 from the Department of Electrical Engineering, Auburn University, Alabama. After spending several years at AT&T Labs - Research, he joined Georgia Tech as an Associate Professor in 2000. His general research interests include statistical signal processing and wireless communications. In these areas, he has contributed over 100 papers published in referred journals and presented in various international conferences. He also has over 10 USA patents granted or pending. He once served as a Guest Editor for two special issues on Signal Processing for Wireless Communications for the IEEE J-SAC. He is currently serving as an Editor for Wireless Communication Theory for the IEEE Transactions on Communications and an Editorial Board Member of EURASIP Journal on Applied Signal Processing. He organized and chaired many international conferences, including Vice-Chair of IEEE 2003 International Conference on Communications.

EURASIP Journal on Applied Signal Processing 2004:5, 613–628

Timing-Free Blind Multiuser Detection for Multicarrier DS/CDMA Systems with Multiple Antennae

Stefano Buzzi
DAEIMI, Università degli Studi di Cassino, Via Di Biasio 43, 03043 Cassino (FR), Italy
Email: buzzi@unicas.it

Emanuele Grossi
DAEIMI, Università degli Studi di Cassino, Via Di Biasio 43, 03043 Cassino (FR), Italy
Email: e.grossi@unicas.it

Marco Lops
DAEIMI, Università degli Studi di Cassino, Via Di Biasio 43, 03043 Cassino (FR), Italy
Email: lops@unicas.it

Received 30 December 2002; Revised 30 July 2003

The problem of blind multiuser detection for an asynchronous multicarrier DS-CDMA system employing multiple transmit and receive antennae over a Rayleigh fading channel is considered in this paper. The solutions that we develop require prior knowledge of the spreading code of the user to be decoded only, while no further information either on the user to be decoded or on the other active users is required. Several combining rules for the observables at the output of each receive antenna are proposed and assessed, and the implications of the different options are studied in depth in terms of both detection performance and computational complexity. A closed form expression is also derived for the conditional error probability and a lower bound for the near-far resistance is provided. Results confirm that the proposed blind receivers can cope with both multiple access interference suppression and channel estimation at the price of a limited performance loss as compared to the ideal linear receivers which assume perfect channel state information.

Keywords and phrases: MC CDMA, multiple antennae, MIMO systems, channel estimation, timing-free detection, near-far resistance.

1. INTRODUCTION

Multicarrier code division multiple access (MC-CDMA) has been conceived as a transmission format which retains the potentials of direct sequence CDMA (DS-CDMA)—and in particular its resistance to multipath effects induced by the radio channel as the communication rate grows larger and larger [1]—while relaxing some very demanding requirements posed by its competitor. In particular, the efficacy of DS-CDMA on wireless channels is mainly due to the recombination of multiple rays so as to increase the average signal-to-noise ratio, but this inevitably poses the problem of a tight synchronization so as to avoid heavy mismatch losses in the replicas-retrieving process. MC-CDMA, instead, by partitioning the available bandwidth in many subbands, no larger than the channel coherence bandwidth, and allocating in each subband independently modulated digital signals, achieves two advantages, that is, (a) the propagation channel in each subband is frequency-flat, and (b) the symbol duration for the data signals occupying the frequency subbands grows linearly with the number of subbands, thus implying that the need for fast electronics and high-performance synchronization schemes is less stringent. The combination of the MC concept with the CDMA technology has led to the birth of three main access schemes, that is, multitone CDMA [2, 3], MC CDMA [4, 5, 6], and MC DS-CDMA [7, 8, 9, 10].

On the other hand, both MC-CDMA and DS-CDMA are expected to support, in future wireless networks, extremely high data rates, which may be in contradiction with their inherent spectral inefficiency. A viable mean to cope with this problem is to resort to multiple transmit and receive antennae. Indeed, recent results from information theory have shown that the capacity of a multiantenna wireless communication system in a rich scattering environment grows with a

law approximately linear in the minimum between the number of transmit and receive antennae [11]. Roughly speaking, multiple transmit antennae generate a spatial diversity which can be successfully exploited at the receiver end to improve performance, especially if space-time coding techniques are employed at the transmitter [12]. Motivated by these considerations, many studies have been recently published for either single-user or multiuser multiantenna systems [13, 14].

All of these studies, though, assume either perfect channel state information (CSI) or error-free estimation thereof. The problem of evaluating the cost of such an information has been only recently considered [15] and the main results are as follows: (a) the training and the data transmission phase should be carefully designed in order to ensure reliable transmission in a multiantenna system on wireless channel; (b) in the large signal-to-noise ratio regime, the length of the training phase should be in the order of the number of transmit antennae; (c) in the region of low signal-to-noise ratios, about half the transmission time should be devoted to training, and, moreover, the capacity of trained systems is far from the optimal one. It is also worth pointing out that in a CDMA multiaccess network, the signal-to-interference-plus-noise ratio is expected to be quite low, at least as far as the network load increases, whereby the task of reducing—if not nullifying—the training phase is more and more stringent.

Motivated by these results, the present paper deals with the problem of blind multiantenna systems employing an MC DS-CDMA modulation format.[1] Since the prior uncertainty as to the CSI results in a complete lack of knowledge of the spatial signatures of both the user of interest and of the other users, while knowledge of the spreading code of all of the active users can be reasonably assumed only at the "base station" of an isolated cell, we consider the more general scenario where the receiver cannot avail itself of any prior information beyond the spreading code of the user of interest, and is thus faced with *asynchronous* cochannel interference (whether from the same cell or from nearby cells); thus differential encoding-decoding is assumed, as a result of the lack of a phase reference. For the sake of simplicity, we also consider uncoded transmission, even though the results can be extended to account for space-time block coding. The main contributions of this paper can be summarized as follows.

(1) We develop a signal model for an MC DS-CDMA system operating over a fading dispersive channel and employing multiple transmit and receive antennae that resembles the signal model developed in [16, 17, 18] with reference to a single-antenna DS-CDMA system operating in the same conditions.

(2) Based on the above analogy, we extend the subspace techniques developed in [16, 19] to the multiantenna MC DS-CDMA system and, moreover, we propose several combining schemes to integrate the statistics observed on each receive antenna branch. It should be noted that the resulting receivers are blind and timing-free, that is, they do not require any information beyond the spreading code of the user to be detected. Interestingly, not even the propagation delay and initial transmitter timing offset for the user of interest is required.

(3) As a by-product of the previous derivations, we also introduce a subspace-based technique which enables blind channel estimation up to a complex scaling factor.

(4) We also provide a thorough performance analysis of the proposed receivers; in particular, we derive closed-form formulas for the conditional error probability and for the near-far resistance, given the channel impulse response realization. It is worth noticing that the methodology outlined here is quite general and can be used to express the performance of any linear receiver in differentially encoded systems.

The rest of the paper is organized as follows. Section 2 outlines the system model, while Section 3 is devoted to the development of the detection structures. In Section 4, the statistical analysis of the receiver is provided, while Section 5 is devoted to the discussion of the numerical results. Finally, concluding remarks are given in Section 6.

Notation

In the following, $\overline{(\cdot)}$, $(\cdot)^T$, and $(\cdot)^H$ denote conjugate, transpose, and conjugate transpose, respectively; $\mathcal{M}_{m\times n}(\mathbb{C})$ is the set of all the $m \times n$-dimensional matrices with complex-valued entries. $E[\cdot]$ denotes statistical expectation; $\Re(\cdot)$ and $\Im(\cdot)$ denote real part and coefficient of the imaginary part, respectively; column-vectors and matrices are indicated through boldface lowercase and uppercase letters, respectively. The term $\mathrm{Im}(\mathbf{A})$ is the image of $\mathbf{A}$, that is, its column span, while $\mathrm{Ker}(\mathbf{A})$ is the null space of $\mathbf{A}$, that is, the orthogonal complement of $\mathrm{Im}(\mathbf{A})$; $\dim(S)$ is the dimensionality of the subspace S; the symbols $\langle\cdot,\cdot\rangle$, $\otimes$, and $\odot$ denote the canonical scalar product, the Kronecker product, and the Schur (i.e., component-wise) matrix product, respectively; $\mathbf{I}_n$ denote the identity matrix of order n; $\mathbf{O}_{m,n}$ and $\mathbf{0}_m$ are the $m \times n$-dimensional matrix and m-dimensional vector with null entries, respectively, and $\mathrm{diag}(\mathbf{a})$ is a diagonal matrix containing the elements of the vector $\mathbf{a}$ on its diagonal; $\mathbf{A}^+$ is the *Moore-Penrose* generalized inverse of $\mathbf{A}$. $\mathrm{supp}\{f\}$ is the support of the function f, that is, the set of its arguments for which f is not zero and $u_T(\tau)$ is the unit height rectangular waveform of support $(0, T)$. $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ denotes the distribution of a Gaussian vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$ while $Q(\cdot)$ is the area under the leading tail of standard Gaussian pdf; finally $Q_1(\cdot,\cdot)$ and $I_0(\cdot)$ are the Marcum function and the modified Bessel function of the first kind and order zero, respectively.

[1] The results presented here can be easily extended to the multitone CDMA and to the MC-CDMA techniques as well.

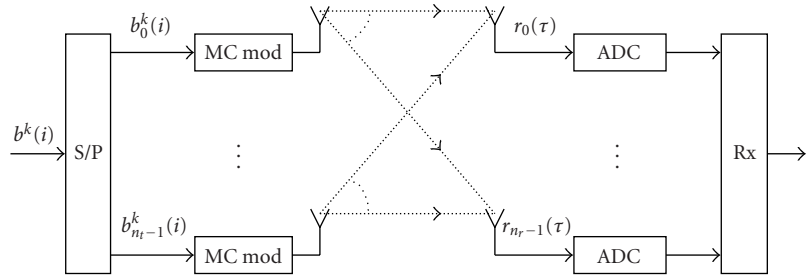


Figure 1: Scheme of a communication system with multiple transmit and receive antennae.

2. SYSTEM MODEL

The general scheme of an MC communication system equipped with multiple transmit and receive antennae is shown in Figure 1. A block of n_t symbols is converted from serial to parallel and each symbol feeds a (spatially) separate antenna. Thus, the n_t symbols are transmitted in parallel, achieving an n_t-fold increase in the data rate, and received on n_r spatially separated receive antennae, providing an n_rth-order receive diversity to combat fading.

The complex envelope of the signal received on the rth antenna can be formally written as

$$\varrho_r(\tau) = \sum_{k=0}^{K-1} \sum_{l=0}^{P-1} A^k \sum_{t=0}^{n_t-1} b_t^k(l)\beta_t^k(\tau - \tau^k - lT_b) * h_{t,r}^k(\tau) + w_r(\tau), \tag{1}$$

where

(1) K is the number of active users;
(2) P is the length of the transmitted frame;
(3) A^k is the amplitude of the signal transmitted by the kth user;
(4) $b_t^k(l)$ is the symbol transmitted by the tth antenna of the kth user at the lth bit interval;
(5) $\beta_t^k(\tau)$ is the signature assigned to the tth transmitter of the kth user;
(6) T_b is the bit duration;
(7) τ^k is the kth user's overall delay, that is, the sum of the kth user transmission delay and of the propagation time through the channel;
(8) $h_{t,r}^k(\tau)$ is the channel impulse response from the tth transmit of the k-user to the rth receive;
(9) $w_r(\tau)$ is the additive white Gaussian noise on the rth receive antenna, independent for different antennae, with power spectral density $2\mathcal{N}_0$.

On the other hand, the signatures in (1) are

$$\beta_t^k(\tau) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} c_t^k(nM+m)\psi_{tx}(\tau - mT_c)e^{2\pi i f_n \tau}, \tag{2}$$

where

(1) N is the number of subcarriers provided to each user;
(2) M is the spreading gain on each subcarrier (hence $PG = MN$ is the overall processing gain);
(3) $c_t^k(l)$, $l = 0, \ldots, MN-1$, is the spreading sequence assigned to the tth antenna of the kth user;
(4) $T_c = T_b/M$ is the chip duration;
(5) $\psi_{tx}(\tau)$ is a unit-energy chip waveform supported in $[0, \Delta_{tx}T_c]$, with bandwidth B_{sc}; Δ_{tx} is a suitable integer so that the signal energy content outside B_{sc} is negligible;
(6) f_n, $n = 0, \ldots, N-1$, are the frequencies assigned to the subcarriers.

Notice that, denoting by $\mathcal{E}_b^k$ the energy per bit of the kth user, we have $A^k = \sqrt{2\mathcal{E}_b^k/NM}$.

The number of subcarriers N employed in an MC system and their spacing Δf have to be properly chosen, based on the channel characteristics. Indeed, if B_{coher} is the coherence bandwidth of the channel, N should be chosen so as to ensure fading flatness in each subband and fading independence between adjacent subbands; thus, if $2W$ is the overall bandwidth assigned for transmission, N, B_{sc}, and Δf result from the following set of constraints:

(i) $B_{sc} \le B_{\text{coher}}$: fading flatness on the single subband;
(ii) $\Delta f \ge B_{\text{coher}}$: fading independence for different subband;
(iii) $(N-1)\Delta f + B_{sc} = 2W$: available bandwidth.

For given N, the processing gain on each subcarrier is fixed ($M = PG/N$), and the channel frequency response can be approximated as follows:

$$H_{t,r}^k(f)u_{2W}(f-W) \simeq \sum_{n=0}^{N-1} H_{t,r}^k(f_n)u_{\Delta f}\left(f - \left(f_n - \frac{\Delta f}{2}\right)\right) = \sum_{n=0}^{N-1} H_{t,r,n}^k u_{\Delta f}\left(f - \left(f_n - \frac{\Delta f}{2}\right)\right), \tag{3}$$

where $f_n = (n - (N-1)/2)\Delta f$. We assume a slowly fading channel, namely, whose coherence time exceeds the packet duration PT_b. As to $H_{t,r,n}^k$, it is modelled as a sequence of complex standard Gaussian random variables, independent

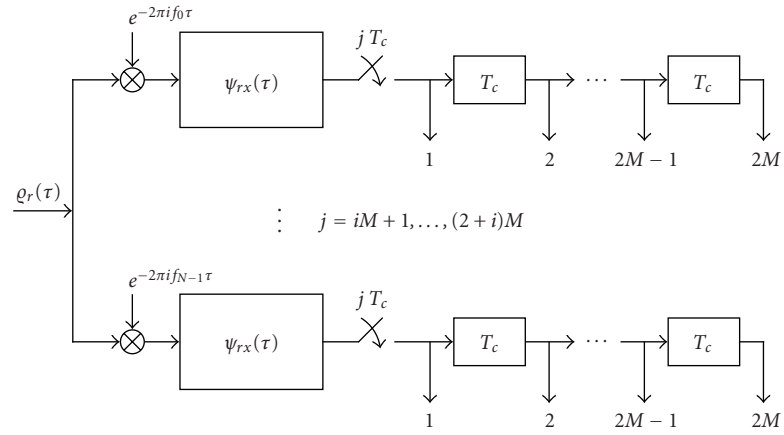


Figure 2: General A/D converter for an MC DS-CDMA system.

for all n; additionally, due to the spatial separation, they are also independent for different t, r, and k.

At the receiver side, the signal observed on each antenna is converted to discrete-time. According to the scheme in Figure 2, there are N branches (i.e., as many as the number of carriers) in the anolog-to-digital converter (ADC), each one consisting of a mixer and of a low-pass filter $\psi_{rx}(\tau)$, whose output is sampled every T_c seconds. Ideally, the filter $\psi_{rx}(\tau)$ should be strictly bandlimited, with bandwidth not smaller than B_{sc} and not larger than Δf; in practice, it is realized through a waveform with finite support $[0, \Delta_{rx} T_c]$ and bandwidth extending between B_{sc} and Δf. It is also required to have a Nyquist autocorrelation, that is, $r_{\psi_{rx}}(jT_c) = \int_{\mathbb{R}} \psi_{rx}(\tau)\overline{\psi}_{rx}(\tau - jT_c)d\tau = \delta(j)$: this implies that output noise samples are uncorrelated. At the nth branch, the output of the low-pass filter at the rth antenna is written as follows:

$$\begin{aligned} r_{r,n}(\tau) &= \left(\varrho_r(\tau)e^{-2\pi i f_n \tau}\right) * \psi_{rx}(\tau) \\ &= \sum_{k=0}^{K-1}\sum_{l=0}^{P-1} A^k \sum_{t=0}^{n_t-1} b_t^k(l) \\ &\quad \times \sum_{m=0}^{M-1} c_t^k(nM+m)\psi_{tx}(\tau - \tau^k - mT_c - lT_b) \\ &\quad * \left(h_{t,r}^k(\tau - \tau^k)e^{-2\pi i f_n \tau}\right) * \psi_{rx}(\tau) \\ &\quad + \left(w_r(\tau)e^{-2\pi i f_n \tau}\right) * \psi_{rx}(\tau) \\ &= \sum_{k=0}^{K-1}\sum_{l=0}^{P-1} A^k \sum_{t=0}^{n_t-1} b_t^k(l)s_{t,r,n}^k(\tau - lT_b) + w_{r,n}(\tau), \end{aligned} \tag{4}$$

where

$$\begin{aligned} s_{t,r,n}^k(\tau) &= \sum_{m=0}^{M-1} c_t^k(nM+m)g_{t,r,n}^k(\tau - mT_c), \\ g_{t,r,n}^k(\tau) &= A^k\psi_{tx}(\tau) * \left(h_{t,r}^k(\tau - \tau^k)e^{-2\pi i f_n \tau}\right) * \psi_{rx}(\tau) \\ &= H_{t,r,n}^k \varphi^k(\tau - \tau^k). \end{aligned} \tag{5}$$

In this equation, $\varphi^k(\tau) = A^k\psi_{tx}(\tau) * \psi_{rx}(\tau)$ and use has been made of the fact that the channel is flat on each subcarrier. It is worthwhile noticing that

(i) in (4), the only substream surviving filtering is the nth one as, due to the bandlimitedness of the transmitted chip waveform, there is no intercarrier interference;

(ii) all of the unknown parameters ($H_{t,r,n}^k$ and τ^k) due to propagation through the channels and users transmitting delay have been shoved in the *unknown* functions $g_{t,r,n}^k(\tau)$.

Notice that the prior uncertainty as to the delay parameter τ^k derives from the initial timing offset of the kth transmitter and from the propagation delay. However, while the latter contribution could be easily absorbed in the channel impulse response, the former should be explicitly accounted for in the context of an asynchronous network: this fact, coupled with the use of strictly bandlimited chip waveforms, poses some limitations on the maximum users number that will be discussed in greater detail later on in the paper.

Upon sampling at chip rate, the signal $r_{r,n}(\tau)$ is converted to the sequence

$$r_{r,n}(jT_c) = \sum_{k=0}^{K-1}\sum_{l=0}^{P-1} A^k \sum_{t=0}^{n_t-1} b_t^k(l)s_{t,r,n}^k(jT_c - lT_b) + w_{r,n}(jT_c). \tag{6}$$

As $\varphi^k(\tau)$ has a compact support in $[0, \Delta T_c]$, with $\Delta = \Delta_{tx} + \Delta_{rx}$, according to (5), we have

$$\begin{aligned} \text{supp}\,\{g_{t,r,n}^k(\tau)\} &= [\tau^k, \tau^k + \Delta T_c] \subset [0, T_b + 2T_c], \\ &\quad \text{with } g_{t,r,n}^k(0) = g_{t,r,n}^k(T_b + 2T_c) = 0, \\ \text{supp}\,\{s_{t,r,n}^k(\tau)\} &= [\tau^k, \tau^k + \Delta T_c + (M-1)T_c] \\ &\subset [0, 2T_b + T_c], \\ &\quad \text{with } s_{t,r,n}^k(0) = s_{t,r,n}^k(2T_b + T_c) = 0, \end{aligned} \tag{7}$$

where the inclusions stem from the assumption that $\tau^k + \Delta - 2T_c < T_b$. Thus, assuming that we are interested in

decoding the information symbols transmitted by the 0th antenna of the 0th user, as $s^k_{t,r,n}(jT_c - iT_b) \neq 0$ only for $j = iM+1,\ldots,(i+2)M$, $b^0_0(i)$ can be detected through the windowed observables $r_{r,n}(jT_c)$, for $j = iM+1,\ldots,(i+2)M$, that can be arranged in the vector

$$\mathbf{r}_{r,n}(i) = \left(r_{r,n}(iT_b + T_c)\cdots r_{r,n}((i+2)T_b)\right)^T \in \mathbb{C}^{2M}. \quad (8)$$

Stacking now the discrete-time version of $g^k_{t,r,n}(\tau)$ into the vector

$$\begin{aligned}\mathbf{g}^k_{t,r,n} &= \left(g^k_{t,r,n}(T_c)\cdots g^k_{t,r,n}(T_b+T_c)\right)^T \\ &= H^k_{t,r,n}\left(\varphi^k(T_c-\tau^k)\cdots\varphi^k(T_b+T_c-\tau^k)\right)^T \\ &= H^k_{t,r,n}\boldsymbol{\varphi}^k \in \mathbb{C}^{M+1},\end{aligned} \quad (9)$$

and defining the following matrices:

$$\begin{aligned}&\mathbf{C}^k_{t,n,0} \\ &= \begin{pmatrix} c^k_t(nM) & 0 & 0 \\ c^k_t(nM+1) & c^k_t(nM) & 0 \\ \vdots & \vdots & \vdots \\ c^k_t(nM+M-1) & c^k_t(nM+M-2) & \ddots \quad c^k_t(nM) \\ 0 & c^k_t(nM+M-1) & c^k_t(nM+1) \\ \vdots & \vdots & \vdots \\ 0 & 0 & c^k_t(nM+M-1) \end{pmatrix} \\ &\qquad \in \mathcal{M}_{2M\times M+1}(\mathbb{C}),\end{aligned}$$

$$\begin{aligned}\mathbf{C}^k_{t,n,-1} &= \begin{pmatrix} \mathbf{C}^k_{t,n,0L} \\ \mathbf{O}_{M,M+1} \end{pmatrix} \in \mathcal{M}_{2M\times M+1}(\mathbb{C}), \\ \mathbf{C}^k_{t,n,+1} &= \begin{pmatrix} \mathbf{O}_{M,M+1} \\ \mathbf{C}^k_{t,n,0H} \end{pmatrix} \in \mathcal{M}_{2M\times M+1}(\mathbb{C}),\end{aligned} \quad (10)$$

where $\mathbf{C}^k_{t,n,0H}$ and $\mathbf{C}^k_{t,n,0L} \in \mathcal{M}_{M\times M+1}(\mathbb{C})$ contain the M upper and M lower rows of the matrix $\mathbf{C}^k_{t,n,0}$, respectively, the discrete-time version $s^k_{t,r,n}(jT_c - lT_b)$, $l = i-1, i, i+1$, of the signatures $s^k_{t,r,n}(\tau - lT_b)$ are represented by the vectors

$$\begin{aligned}\mathbf{s}^k_{t,r,n,-1} &= \left(s^k_{t,r,n}(Tb+T_c)\cdots s^k_{t,r,n}(3T_b)\right)^T \\ &= \mathbf{C}^k_{t,n,-1}\mathbf{g}^k_{t,r,n} \in \mathbb{C}^{2M}, \\ \mathbf{s}^k_{t,r,n,0} &= \left(s^k_{t,r,n}(T_c)\cdots s^k_{t,r,n}(2T_b)\right)^T \\ &= \mathbf{C}^k_{t,n,0}\mathbf{g}^k_{t,r,n} \in \mathbb{C}^{2M}, \\ \mathbf{s}^k_{t,r,n,+1} &= \left(s^k_{t,r,n}(-T_B+T_c)\cdots s^k_{t,r,n}(T_b)\right)^T \\ &= \mathbf{C}^k_{t,n,+1}\mathbf{g}^k_{t,r,n} \in \mathbb{C}^{2M}.\end{aligned} \quad (11)$$

Thus, the discrete-time observable $\mathbf{r}_{r,n}(i)$ in (8) can be recast as

$$\mathbf{r}_{r,n}(i) = \sum_{k=0}^{K-1}\sum_{l=-1}^{1}\sum_{t=0}^{n_t-1} b^k_t(i+l)\mathbf{s}^k_{t,r,n,l} + \mathbf{w}_{r,n}(i), \quad (12)$$

where

$$\begin{aligned}\mathbf{w}_{r,n}(i) &= \left(w_{r,n}(iT_b+T_c)\cdots w_{r,n}((i+2)T_b)\right)^T \\ &\sim \mathcal{N}(\mathbf{0}_{2M}, 2\mathcal{N}_0\mathbf{I}_{2M}).\end{aligned} \quad (13)$$

Stacking up the vectors corresponding to the N subcarriers, we obtain the following discrete observable at the rth receive antenna:

$$\begin{aligned}\mathbf{r}_r(i) &= \begin{pmatrix} \mathbf{r}_{r,0}(i) \\ \vdots \\ \mathbf{r}_{r,N-1}(i) \end{pmatrix} \\ &= \sum_{k=0}^{K-1}\sum_{l=-1}^{1}\sum_{t=0}^{n_t-1} b^k_t(i+l)\mathbf{s}^k_{t,r,l} + \mathbf{w}_r(i) \in \mathbb{C}^{2MN},\end{aligned} \quad (14)$$

where we have let

$$\begin{aligned}\mathbf{s}^k_{t,r,l} &= \begin{pmatrix} \mathbf{s}^k_{t,r,0,l} \\ \vdots \\ \mathbf{s}^k_{t,r,N-1,l} \end{pmatrix} \\ &= \mathbf{C}^k_{t,l}\mathbf{g}^k_{t,r} \in \mathbb{C}^{2MN}, \\ \mathbf{C}^k_{t,l} &= \begin{pmatrix} \mathbf{C}^k_{t,0,l} & & \mathbf{O}_{M,M+1} \\ & \ddots & \\ \mathbf{O}_{M,M+1} & & \mathbf{C}^k_{t,N-1,l} \end{pmatrix} \in \mathcal{M}_{2MN\times(M+1)N}(\mathbb{C}), \\ \mathbf{g}^k_{t,r} &= \begin{pmatrix} \mathbf{g}^k_{t,r,0} \\ \vdots \\ \mathbf{g}^k_{t,r,N-1} \end{pmatrix} \\ &= \begin{pmatrix} H^k_{t,r,0} \\ \vdots \\ H^k_{t,r,N-1} \end{pmatrix} \otimes \boldsymbol{\varphi}^k \\ &= \mathbf{h}^k_{t,r} \otimes \boldsymbol{\varphi}^k \in \mathbb{C}^{(M+1)N}, \\ \mathbf{w}_r(i) &= \begin{pmatrix} \mathbf{w}_{r,0}(i) \\ \vdots \\ \mathbf{w}_{r,N-1}(i) \end{pmatrix} \in \mathbb{C}^{2MN}.\end{aligned} \quad (15)$$

Notice that in (14), $\mathbf{s}^k_{t,r,0}$ is the complete signature transmitted by the tth antenna of the k-user and received, after propagation, at the rth antenna (namely, it is a spatial signature related to the real one through the channel impulse response); $\mathbf{s}^k_{t,r,-1}$ and $\mathbf{s}^k_{t,r,+1}$ are parts of the signature related to the previous and successive transmitted symbol; the vectors $\mathbf{g}^k_{t,r}$ contain both the unknown channel coefficients (through the vectors $\mathbf{h}^k_{t,r} \sim \mathcal{N}(\mathbf{0}_N, \mathbf{I}_N)$) and the users timings (through the vectors $\boldsymbol{\varphi}^k$); finally, $\mathbf{w}_r(i) \sim \mathcal{N}(\mathbf{0}_{2MN}, 2\mathcal{N}_0\mathbf{I}_{2MN})$ accounts for the thermal noise.

The above model represents the extension to the MC DS-CDMA case with multiple antennae of a well-known

representation derived for single-antenna DS-CDMA systems operating over fading dispersive channels [16, 17, 18, 19]. In this scenario, in order to allow possible joint processing of the observables at all of the receive antennae, it is useful to define the vector $\mathbf{r}(i) = (\mathbf{r}_0(i) \cdots \mathbf{r}_{n_r-1}(i))^T$, which, upon defining quantities

$$\begin{aligned}
\mathbf{s}_{t,l}^k &= \begin{pmatrix} \mathbf{s}_{t,0,l}^k \\ \vdots \\ \mathbf{s}_{t,n_r-1,l}^k \end{pmatrix} = \mathbf{S}_{t,l}^k \mathbf{g}_t^k \in \mathbb{C}^{2MNn_r}, \\
\mathbf{S}_{t,l}^k &= \mathbf{I}_{n_r} \otimes \mathbf{C}_{t,l}^k \in \mathcal{M}_{2MNn_r \times (M+1)Nn_r}(\mathbb{C}), \\
\mathbf{g}_t^k &= \begin{pmatrix} \mathbf{g}_{t,0}^k \\ \vdots \\ \mathbf{g}_{t,n_r-1}^k \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{h}_{t,0}^k \\ \vdots \\ \mathbf{h}_{t,n_r-1}^k \end{pmatrix} \otimes \boldsymbol{\varphi}^k \\
&= \mathbf{h}_t^k \otimes \boldsymbol{\varphi}^k \in \mathbb{C}^{(M+1)Nn_r}, \\
\mathbf{w}(i) &= \begin{pmatrix} \mathbf{w}_0(i) \\ \vdots \\ \mathbf{w}_{n_r-1}(i) \end{pmatrix} \in \mathbb{C}^{2MNn_r},
\end{aligned} \tag{16}$$

can be also written as follows:

$$\begin{aligned}
\mathbf{r}(i) &= \begin{pmatrix} \mathbf{r}_0(i) \\ \vdots \\ \mathbf{r}_{n_r-1}(i) \end{pmatrix} \\
&= \sum_{k=0}^{K-1} \sum_{l=-1}^{1} \sum_{t=0}^{n_t-1} b_t^k(i+l) \mathbf{s}_{t,l}^k + \mathbf{w}(i) \\
&= \underbrace{b_0^0(i)\mathbf{s}_{0,0}^0}_{\text{useful signal}} + \underbrace{b_0^0(i-1)\mathbf{s}_{0,-1}^0 + b_0^0(i+1)\mathbf{s}_{0,+1}^0}_{\text{ISI}} \\
&\quad + \underbrace{\sum_{l=-1}^{1} \sum_{t=1}^{n_t-1} b_t^0(i+l)\mathbf{s}_{t,l}^0}_{\text{self-interference}} + \underbrace{\sum_{k=1}^{K-1} \sum_{l=-1}^{1} \sum_{t=0}^{n_t-1} b_t^k(i+l)\mathbf{s}_{t,l}^k}_{\text{MAI}} + \underbrace{\mathbf{w}(i)}_{\text{noise}} \\
&= b_0^0(i)\mathbf{s}_{0,0}^0 + \mathbf{z}(i) + \mathbf{w}(i) \\
&= \mathbf{q}(i) + \mathbf{w}(i) \in \mathbb{C}^{2MNn_r}.
\end{aligned} \tag{17}$$

In (17), $\mathbf{s}_{0,0}^0$ is the useful signature, $\mathbf{z}(i)$ represents the self-interference, multiuser interference (MAI), and intersymbol interference (ISI) contribution, and $\mathbf{w}(i) \sim \mathcal{N}(\mathbf{0}_{2MNn_r}, 2\mathcal{N}_0\mathbf{I}_{2MNn_r})$ is the thermal noise. Notice that the subscript "t" points out that each transmit antenna of a given user is assigned a different spreading sequence, a condition that will be shown to be necessary in blind uncoded systems. For future reference, notice that the covariance matrix of $\mathbf{r}(i)$ is equal to

$$\begin{aligned}
\mathbf{R}_{rr} &= E[\mathbf{r}(i)\mathbf{r}^H(i)] \\
&= \sum_{k=0}^{K-1} \sum_{t=0}^{n_t-1} [\mathbf{s}_{t,-1}^k \mathbf{s}_{t,-1}^{kH} + \mathbf{s}_t^k \mathbf{s}_t^{kH} + \mathbf{s}_{t,+1}^k \mathbf{s}_{t,+1}^{kH}] + 2\mathcal{N}_0\mathbf{I}_{2MNn_r} \\
&= \mathbf{R}_{qq} + 2\mathcal{N}_0\mathbf{I}_{2MNn_r}.
\end{aligned} \tag{18}$$

3. DETECTOR DESIGN

The detectors that are considered in this paper are linear, and thus uniquely specified by a suitable complex-valued vector $\mathbf{m}$.[2] As anticipated, differential coding/decoding is to be adopted to cope with the absence of a phase reference, whereby the desired information is contained in the quantity $d_0^0(i) = b_0^0(i)b_0^0(i-1)$. At the receiver side, the observables $\mathbf{r}_0(i), \ldots, \mathbf{r}_{n_r-1}(i)$ can be either processed separately and then combined or processed jointly through the vector in (17); we refer to the former case as *noncooperative* detection and to the latter case as *cooperative* detection.

Noncooperative detection

If we adopt a noncooperative scheme, the signals at the output of the n_r antennae are processed through as many detectors, whose outputs are expressed by $\vartheta_r(i) = \langle \mathbf{r}_r(i), \mathbf{m}_r \rangle$, $r = 1, \ldots, n_r - 1$. The vector $\boldsymbol{\vartheta}(i) = (\vartheta_0(i) \cdots \vartheta_{n_r-1}(i))^T$ is then forwarded to a combining block, which makes the decisions $\widehat{d}_0^0(i) = f(\boldsymbol{\vartheta}(i), \boldsymbol{\vartheta}(i-1))$. We consider three different scenarios.

(1) *Soft integration*. In this case, the decision rule assumes the form

$$\begin{aligned}
\widehat{d}_0^0(i) &= f(\boldsymbol{\vartheta}(i), \boldsymbol{\vartheta}(i-1)) = \operatorname{sgn}\left[\Re\left(\langle \boldsymbol{\vartheta}(i), \boldsymbol{\vartheta}(i-1) \rangle\right)\right] \\
&= \operatorname{sgn}\left[\Re\left(\sum_{r=1}^{n_t-1} \vartheta_r(i)\overline{\vartheta}_r(i-1)\right)\right],
\end{aligned} \tag{19}$$

that is, the decision takes place after the integration of the soft differential statistics $\vartheta_r(i)\overline{\vartheta}_r(i-1)$.

(2) *Hard integration* (with a randomized offset):

$$\begin{aligned}
\widehat{d}_0^0(i) &= f(\boldsymbol{\vartheta}(i), \boldsymbol{\vartheta}(i-1)) \\
&= \operatorname{sgn}\left[\sum_{r=1}^{n_t-1} \operatorname{sgn}\left[\Re\left(\vartheta_r(i)\overline{\vartheta}_r(i-1)\right)\right] + u\right], \\
&\qquad u \sim U\left(\left(-\frac{1}{2}, \frac{1}{2}\right)\right);
\end{aligned} \tag{20}$$

that is, the combination takes place after one-bit quantization of the soft differential statistics. Observe that, for n_r odd, the randomized offset has no effect and this decision amounts to a majority rule, which is optimal for hard-quantized statistics; on the other hand, for n_r

[2] From now on, we adopt the normalization $\|\mathbf{m}\| = 1$.

even, the possibility that $f(\vartheta(i), \vartheta(i-1)) = 0$ is ward off through the secondary threshold u.[3]

(3) *Maximal ratio combiner (MRC)*. According to (14), the vector $\vartheta(i)$ is expressed as follows:

$$\begin{aligned}\vartheta(i) &= \begin{pmatrix} \langle \mathbf{s}^0_{0,0,0}, \mathbf{m}_0 \rangle \\ \vdots \\ \langle \mathbf{s}^0_{0,n_r-1,0}, \mathbf{m}_{n_r-1} \rangle \end{pmatrix} b^0_0(i) + \begin{pmatrix} \langle \mathbf{z}_0(i), \mathbf{m}_0 \rangle \\ \vdots \\ \langle \mathbf{z}_{n_r-1}(i), \mathbf{m}_{n_r-1} \rangle \end{pmatrix} \\ &\quad + \begin{pmatrix} \langle \mathbf{w}_0(i), \mathbf{m}_0 \rangle \\ \vdots \\ \langle \mathbf{w}_{n_r-1}(i), \mathbf{m}_{n_r-1} \rangle \end{pmatrix} \\ &= \tilde{\mathbf{a}}\, b^0_0(i) + \tilde{\mathbf{z}} + \tilde{\mathbf{w}}. \end{aligned} \tag{21}$$

A possible detection strategy consists of weighting the n_r unquantized statistics of the vector $\vartheta(i)$ with the elements of the *gain vector* $\tilde{\mathbf{a}}$, thus realizing an MRC; afterwards, the uncertainty on the phase can be removed though differential detection. The detection rule is thus

$$\begin{aligned}\hat{d}^0_0(i) &= f(\vartheta(i), \vartheta(i-1)) \\ &= \operatorname{sgn}\left[\Re\left(\langle \vartheta(i), \tilde{\mathbf{a}} \rangle \overline{\langle \vartheta(i-1), \tilde{\mathbf{a}} \rangle}\right)\right]. \end{aligned} \tag{22}$$

Cooperative detection

In this scheme, the observables are first stacked in a unique vector and then jointly processed, obtaining $\vartheta(i) = \langle \mathbf{r}(i), \mathbf{m} \rangle$; a decision is finally made through

$$\hat{d}^0_0(i) = \operatorname{sgn}\left[\Re\left(\vartheta(i)\overline{\vartheta}(i-1)\right)\right]. \tag{23}$$

Obviously, the cooperative scheme is expected to achieve, at the price of some complexity increase, a substantial performance improvement with respect to the noncooperative detection schemes.

Notice also that (17) reduces to (14) for $n_r = 1$; as a consequence, the synthesis of the receiver can be carried out starting from the observables in (17) and then specify the results to the case $n_r = 1$. There are, of course, a number of different criteria to design $\mathbf{m}$. The first step is to generalize the subspace-based detector, introduced in [16, 21], to the new scenario and then move on to the newly proposed detector family that is referred to as "two-stage" receivers in what follows.

3.1. Subspace-based receiver

The correlation matrix $\mathbf{R}_{rr}$ of the received signal can be decomposed as

$$\mathbf{R}_{rr} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H = \mathbf{U}_s\boldsymbol{\Lambda}_s\mathbf{U}_s^H + \mathbf{U}_n\boldsymbol{\Lambda}_n\mathbf{U}_n^H, \tag{24}$$

where $\mathbf{U} = (\mathbf{U}_s\mathbf{U}_n)$, $\boldsymbol{\Lambda} = \operatorname{diag}(\boldsymbol{\Lambda}_s, \boldsymbol{\Lambda}_n)$; $\boldsymbol{\Lambda}_s = \operatorname{diag}(\lambda_1, \ldots, \lambda_{3Kn_t})$ contains the $3Kn_t$ largest eigenvalues of $\mathbf{R}_{rr}$ in descending order and $\mathbf{U}_s$ the corresponding orthonormal eigenvectors; $\operatorname{Im}(\mathbf{U}_s)$ and $\operatorname{Im}(\mathbf{U}_n)$ are the signal subspace and the noise subspace, respectively. Based on the above decomposition, the orthogonality between the noise subspace and the useful signal $\mathbf{s}^0_{0,0}$ can be exploited to obtain an estimate, $\hat{\mathbf{g}}^0_0$, say, of the vector $\mathbf{g}^0_0$. In particular, under the condition[4]

$$\dim\left(\operatorname{Im}(\mathbf{R}_{qq}) \cap \operatorname{Im}\left(\mathbf{S}^0_{0,0}\mathbf{S}^{0H}_{0,0}\right)\right) = 1, \tag{25}$$

$\mathbf{g}^0_0$ can be obtained as the unique, nontrivial solution of the equation

$$0 = \mathbf{U}_n^H\mathbf{s}^0_{0,0} = \mathbf{U}_n^H\mathbf{S}^0_{0,0}\mathbf{g}^0_0. \tag{26}$$

Since in practice the covariance matrix $\mathbf{R}_{rr}$ is not known, it has to be replaced by its sample estimate $\hat{\mathbf{R}}_{rr} = (1/Q)\sum_{i=0}^{Q-1}\mathbf{r}(i)\mathbf{r}^H(i)$, whose spectral decomposition is

$$\hat{\mathbf{R}}_{rr} = \hat{\mathbf{U}}_s\hat{\boldsymbol{\Lambda}}_s\hat{\mathbf{U}}_s^H + \hat{\mathbf{U}}_n\hat{\boldsymbol{\Lambda}}_n\hat{\mathbf{U}}_n^H. \tag{27}$$

Accordingly, $\hat{\mathbf{g}}^0_0$ solves the problem

$$\hat{\mathbf{g}}^0_0 = \arg\min_{\|\mathbf{x}\|=1} \left\|\hat{\mathbf{U}}_n^H\mathbf{S}^0_{0,0}\mathbf{x}\right\|^2, \tag{28}$$

that is, it is the eigenvector corresponding to the smallest eigenvalue of the matrix $\mathbf{S}^{0H}_{0,0}\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^H\mathbf{S}^0_{0,0}$.

The vector $\hat{\mathbf{g}}^0_0$ is then used to obtain the classical minimum mean square error (MMSE) and zero-forcing (ZF) receivers, that is,

$$\begin{aligned}\mathbf{m}_{\mathrm{MMSE}} &= \hat{\mathbf{R}}_{rr}^{-1}\mathbf{S}^0_{0,0}\hat{\mathbf{g}}^0_0, \\ \mathbf{m}_{\mathrm{ZF}} &= \hat{\mathbf{R}}^{+}_{qq}\mathbf{S}^0_{0,0}\hat{\mathbf{g}}^0_0, \end{aligned} \tag{29}$$

with

$$\begin{aligned}\hat{\mathbf{R}}_{qq} &= \hat{\mathbf{U}}_s(\hat{\boldsymbol{\Lambda}}_s - 2\widehat{\mathcal{N}_0}\mathbf{I})\hat{\mathbf{U}}_s^H, \\ 2\widehat{\mathcal{N}_0} &= \frac{1}{2MNn_r - 3Kn_t}\sum_{i=3Kn_t+1}^{2MNn_r}(\hat{\boldsymbol{\Lambda}}_n)_{ii}. \end{aligned} \tag{30}$$

3.2. Two-stage receiver

The subspace-based receivers exhibit a noticeable performance degradation as the users number grows large, since the dimensionality of the noise subspace decreases and the estimate of the vector $\mathbf{g}^0_0$ becomes worse and worse. A possible mean to cope with these overloaded scenarios is to resort to the "two-stage" receivers, introduced in [18, 19] with reference to single-antenna DS-CDMA networks. As a consequence, the mathematical proofs of the results in Sections 3.2.1 and 3.2.3 will be omitted so as to avoid any overlap with available literature.

[3] For further details on the optimality of randomized tests, see [20].

[4] Remember that $\operatorname{Im}(\mathbf{R}_{qq}) = \operatorname{Im}(\mathbf{U}_s) = \operatorname{Ker}(\mathbf{U}_n^H)$ and $\operatorname{Im}(\mathbf{S}^0_{0,0}\mathbf{S}^{0H}_{0,0}) = \operatorname{Im}(\mathbf{S}^0_{0,0})$.

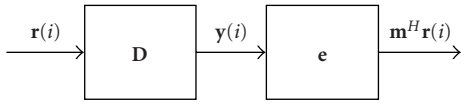


Figure 3: Two-stage linear receiver scheme.

Two-stage detectors owe their name to a functional split of their operation in a suppression block, represented by the matrix $\mathbf{D}$ of Figure 3, and a BER optimization block, represented by the vector $\mathbf{e}$ of the same figure. Obviously, the two stages may collapse into the single vector $\mathbf{m} = \mathbf{De}$.

3.2.1. Synthesis of the interference cancellation stage $\mathbf{D}$

The useful signature $\mathbf{s}_{0,0}^0$ lies in $\mathrm{Im}(\mathbf{S}_{0,0}^0)$, which, in turn, is a vector subspace of $\mathbb{C}^{(M+1)Nn_r}$. The first stage is thus a noninvertible transformation of the observables, that is,

$$\mathbf{y}(i) = \mathbf{D}^H \mathbf{r}(i), \tag{31}$$

where $\mathbf{D} \in \mathcal{M}_{2MNn_r \times (M+1)Nn_r}(\mathbb{C})$ solves one of the following two constrained minimization problems:

$$\begin{aligned} &E\left[\left|\left|\mathbf{D}^H \mathbf{r}(i)\right|\right|^2\right] = \min, \quad \det\left(\mathbf{D}^H \mathbf{S}_{0,0}^0\right) \neq 0; \\ &E\left[\left|\left|\mathbf{D}^H \mathbf{q}(i)\right|\right|^2\right] = \min, \quad \det\left(\mathbf{D}^H \mathbf{S}_{0,0}^0\right) \neq 0. \end{aligned} \tag{32}$$

The former cost function is the classical one for minimum mean output energy (MOE), while the latter involves the minimization of the noise-free observables; in both cases, the constraint ensures that the signal of interest always survives after the noninvertible transformation. Under the condition (25), the solution to the above problems can be shown to be written as follows:

$$\begin{aligned} \mathbf{D} = &\left(\mathbf{R} + \mathbf{S}_{0,0}^0 \mathbf{S}_{0,0}^{0H}\right)^{-1} \mathbf{S}_{0,0}^0 \\ &\times \left[\left(\mathbf{S}_{0,0}^0 \left(\mathbf{R} + \mathbf{S}_{0,0}^0 \mathbf{S}_{0,0}^{0H}\right)^+ \mathbf{S}_{0,0}^{0H}\right) \odot \mathbf{I}\right]^{-1} \mathrm{diag}(\boldsymbol{\alpha}), \end{aligned} \tag{33}$$

where $\boldsymbol{\alpha} \in \mathbb{C}^{(M+1)Nn_r}$ is an arbitrary vector with strictly positive entries and $\mathbf{R}$ can be either $\mathbf{R}_{rr}$ or $\mathbf{R}_{qq}$. If $\mathbf{R} = \mathbf{R}_{rr}$, $\mathbf{D}$ is the solution to the former problem in (32) and subsumes, as the special case of nonfading channel with known timing, the minimum MOE solution equivalent to the MMSE receiver; accordingly, we refer to this solution as an MMSE-like receiver. Otherwise, if $\mathbf{R} = \mathbf{R}_{qq}$, $\mathbf{D}$ is the solution to the latter problem in (32) and subsumes in the same way the linear ZF receiver; we thus refer to this solution as ZF-like receiver. Since scalar multiplicative constants have no influence on the decision rule (see [19]), the matrix $\mathbf{D}$ can be also expressed as follows:

$$\mathbf{D} = \left(\mathbf{R} + \mathbf{S}_{0,0}^0 \mathbf{S}_{0,0}^{0H}\right)^+ \mathbf{S}_{0,0}^0. \tag{34}$$

Before proceeding in the system derivation, it is worth commenting on condition (25), which was advocated to support solution (33). Indeed, the constraints in (32) just ensure that the output useful signature is nonzero with probability one, but they do not offer any guarantee that all of the interference be blocked before further processing. On the other hand, defining

$$\mathbf{X} = \left(\mathbf{s}_{0,-1}^0 \cdots \mathbf{s}_{t,l}^k \cdots \mathbf{s}_{n_t-1,+1}^{K-1} \mathbf{S}_{0,0}^0\right), \tag{35}$$

that is, the matrix containing all the $3Kn_t$ signatures $\mathbf{s}_{t,l}^k$ and $\mathbf{S}_{0,0}^0$, and noticing that

$$\mathbf{R}_{qq} + \mathbf{S}_{0,0}^0 \mathbf{S}_{0,0}^{0H} = \mathbf{X}\mathbf{X}^H, \qquad \mathbf{D}_{\text{ZF-like}} = \left(\mathbf{X}\mathbf{X}^H\right)^+ \mathbf{S}_{0,0}^0, \tag{36}$$

it is seen that a necessary condition for

$$\mathbf{D}_{\text{ZF-like}}^H \mathbf{s}_{t,l}^k = \mathbf{S}_{0,0}^{0H} \left(\mathbf{X}\mathbf{X}^H\right)^+ \mathbf{s}_{t,l}^k = \mathbf{0} \quad \text{for } (k,t,l) \neq (0,0,0), \tag{37}$$

(i.e., for all the interferers to be nullified and the useful signal to survive) is that $\mathbf{s}_{t,l}^k$ and the columns of $\mathbf{S}_{0,0}^0$ be linearly independent with respect to $\mathbf{X}$ for all $(k,t,l) \neq (0,0,0)$ (see [19] for more details). Ensuring that $\mathbf{s}_{0,0}^0$ is the only signature linearly dependent on the columns of $\mathbf{S}_{0,0}^0$ with respect to $\mathbf{X}$ amounts to forcing $\mathbf{s}_{0,0}^0 = \mathbf{S}_{0,0}^0 \mathbf{g}_0^0$ to be the only direction which belongs both to $\mathrm{Im}(\mathbf{S}_{0,0}^0 \mathbf{S}_{0,0}^{0H})$ and to $\mathrm{Im}(\mathbf{R}_{qq})$, that is, to forcing (25) to hold true. This condition will be, in the following, referred to as *identifiability condition*, a term we borrow from [17]: notice however that, while in the subspace-based detectors such a condition is a necessary one in order to ensure the channel identification—and indeed its violation would result in a useless receiver—in our approach, (25) is not a precondition, even though its violation usually results in a performance degradation and in the loss of the near-far resistance properties.

It is also worth pointing out here that, in the considered scenario, (25) cannot be relaxed through signal-space oversampling, as suggested in [16], and implemented in [19], where *rectangular* chip waveforms were adopted. The MC modulation format, instead, requires avoiding the intercarrier interference, which, for asynchronous systems, can be accomplished through the use of strictly bandlimited chip waveforms: obviously, no further sampling beyond the Nyquist rate may be advantageous in this situation.

3.2.2. Blind implementation of $\mathbf{D}$

In order to implement in a blind fashion the MMSE-like receiver, the covariance matrix $\mathbf{R}_{rr}$ is to be replaced in practice by its sample estimate $\hat{\mathbf{R}}_{rr}$; the blocking matrix is then

$$\hat{\mathbf{D}}_{\text{MMSE-like}} = \left(\hat{\mathbf{R}}_{rr} + \mathbf{S}_{0,0}^0 \mathbf{S}_{0,0}^{0H}\right)^+ \mathbf{S}_{0,0}^0. \tag{38}$$

The implementation of the ZF-like receiver requires, instead, more attention since an estimate of $\mathbf{R}_{qq} + \mathbf{S}_{0,0}^0 \mathbf{S}_{0,0}^{0H}$ is needed. To this end, first note that, based on (25),

$$\begin{aligned} &\dim\left(\mathrm{Im}\left(\mathbf{R}_{qq} + \mathbf{S}_{0,0}^0 \mathbf{S}_{0,0}^{0H}\right)\right) \\ &\quad = \dim\left(\mathrm{Im}\left(\mathbf{R}_{qq}\right)\right) + \dim\left(\mathrm{Im}\left(\mathbf{S}_{0,0}^0 \mathbf{S}_{0,0}^{0H}\right)\right) - 1 \\ &\quad = 3Kn_t + (M+1)Nn_r - 1; \end{aligned} \tag{39}$$

whereby, upon eigendecomposition, we obtain

$$\hat{\mathbf{R}}_{qq} + \mathbf{S}_{0,0}^{0}\mathbf{S}_{0,0}^{0H} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{H} = \mathbf{U}_1\boldsymbol{\Lambda}_1\mathbf{U}_1^{H} + \mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^{H}, \quad (40)$$

where $\mathbf{U} = [\mathbf{U}_1 \ \ \mathbf{U}_2]$, $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2)$, $\boldsymbol{\Lambda}_1 = \mathrm{diag}(\lambda_1, \ldots, \lambda_{3Kn_t+(M+1)Nn_r-1})$ contains the $3Kn_t + (M+1)Nn_r - 1$ largest eigenvalues and $\mathbf{U}_1$ the corresponding orthonormal eigenvectors. An estimate of $\mathbf{R}_{qq} + \mathbf{S}_{0,0}^{0}\mathbf{S}_{0,0}^{0H}$ is thus

$$\widehat{\mathbf{R}_{qq} + \mathbf{S}_{0,0}^{0}\mathbf{S}_{0,0}^{0H}} = \mathbf{U}_1\boldsymbol{\Lambda}_1\mathbf{U}_1^{H} \quad (41)$$

and the blind implementation of the ZF-like filter is

$$\hat{\mathbf{D}}_{\text{ZF-like}} = \left(\widehat{\mathbf{R}_{qq} + \mathbf{S}_{0,0}^{0}\mathbf{S}_{0,0}^{0H}}\right)^{+}\mathbf{S}_{0,0}^{0}. \quad (42)$$

3.2.3. Synthesis of the second stage e

Assuming that the blocking matrix $\mathbf{D}$ has suppressed all of the interference (the term $\mathbf{D}^H\mathbf{z}(i)$ is very small if the MMSE-like solution is adopted, while it is exactly zero for the ZF-like one), the observables at the output of the second stage can be written as

$$\mathbf{y}(i) = b_0^0(i)\mathbf{D}^H\mathbf{S}_{0,0}^{0}\mathbf{g}_0^0 + \mathbf{D}^H\mathbf{w}(i). \quad (43)$$

The vector $\mathbf{e}$ can be now chosen so as to minimize the BER, that is, it is the cascade of a whitening filter and of a filter matched to the warped useful signal. Upon considering the "economy size" singular value decomposition $\mathbf{D} = \mathbf{U}_D\boldsymbol{\Lambda}\mathbf{V}^H$, the whitening filter is $\mathbf{V}\boldsymbol{\Lambda}^{-1}$, with $\boldsymbol{\Lambda} \in \mathcal{M}_{(M+1)Nn_r\times(M+1)Nn_r}(\mathbb{C})$ a diagonal matrix and $\mathbf{V} \in \mathcal{M}_{(M+1)Nn_r\times(M+1)Nn_r}(\mathbb{C})$ a unitary square matrix. Accordingly, the whitened observables are given by

$$\begin{aligned}\mathbf{y}_w(i) &= \left(\mathbf{V}\boldsymbol{\Lambda}^{-1}\right)^H\mathbf{D}^H\mathbf{r}(i)\\ &= \boldsymbol{\Lambda}^{-1}\mathbf{V}^H\mathbf{V}\boldsymbol{\Lambda}\mathbf{U}_D^H\mathbf{r}(i) = \mathbf{U}_D^H\mathbf{r}(i)\\ &= b_0^0(i)\mathbf{U}_D^H\mathbf{S}_{0,0}^{0}\mathbf{g}_0^0 + \mathbf{U}_D^H\mathbf{w}(i)\end{aligned} \quad (44)$$

and the matched filter is $\mathbf{U}_D^H\mathbf{S}_0^0\mathbf{g}_0^0$. The second stage is then

$$\mathbf{e} = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{U}_D^H\mathbf{S}_{0,0}^{0}\mathbf{g}_0^0 \quad (45)$$

and the expression of the complete receiver is given by

$$\mathbf{m} = \mathbf{D}\mathbf{e} = \mathbf{U}_D\boldsymbol{\Lambda}\mathbf{V}^H\mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{U}_D^H\mathbf{S}_{0,0}^{0}\mathbf{g}_0^0 = \mathbf{U}_D\mathbf{U}_D^H\mathbf{S}_{0,0}^{0}\mathbf{g}_0^0. \quad (46)$$

3.2.4. Blind implementation of e

Since in practice the vector $\mathbf{g}_0^0$ is not known, a further processing is needed to obtain an estimate of the second stage (45). To this end, notice that the correlation matrix of $\mathbf{y}_w(i)$ can be written as

$$\mathbf{R}_{y_w y_w} = \mathbf{U}_D^H\mathbf{S}_{0,0}^{0}\mathbf{g}_0^0\left(\mathbf{U}_D^H\mathbf{S}_{0,0}^{0}\mathbf{g}_0^0\right)^H + 2\mathcal{N}_0\mathbf{I}_{(M+1)Nn_r}, \quad (47)$$

that is, it consists of the sum of a full-rank matrix and of a unit rank one, the latter admitting $\mathbf{U}_D^H\mathbf{S}_{0,0}^{0}\mathbf{g}_0^0$ as its unique eigenvector. Consequently, the eigenvector $\mathbf{u}_{\max}$ corresponding to the largest eigenvalue of $\mathbf{R}_{y_w y_w}$ is parallel to $\mathbf{U}_D^H\mathbf{S}_{0,0}^{0}\mathbf{g}_0^0$, and the receiver's second stage is $\mathbf{e} = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{u}_{\max}$. Thus the receiver is given by

$$\mathbf{m} = \mathbf{U}_D\mathbf{u}_{\max}. \quad (48)$$

In practice, the vector $\mathbf{u}_{\max}$ is estimated through an eigendecomposition of the sample covariance matrix $\hat{\mathbf{R}}_{y_w y_w}$ of the whitened observables $\mathbf{y}_w(i)$ with

$$\hat{\mathbf{R}}_{y_w y_w} = \frac{1}{Q}\sum_{i=0}^{Q-1}\mathbf{y}_w(i)\mathbf{y}_w(i)^H = \hat{\mathbf{U}}_D^H\hat{\mathbf{R}}_{rr}\hat{\mathbf{U}}_D. \quad (49)$$

3.3. Channel estimation

As a by-product of the previous derivations, an estimate (up to a complex scalar factor) of the discrete-time channel impulse response $\mathbf{g}_0^0$ can be obtained, based on the consideration that $\mathbf{u}_{\max}$ is parallel to $\hat{\mathbf{U}}_D^H\mathbf{S}_{0,0}^{0}\mathbf{g}_0^0$. Accordingly, the estimate $\hat{\mathbf{g}}_0^0$ of $\mathbf{g}_0^0$ is

$$\hat{\mathbf{g}}_0^0 = \left(\hat{\mathbf{U}}_D^H\mathbf{S}_{0,0}^{0}\right)^{-1}\hat{\mathbf{u}}_{\max} = \mathbf{d}. \quad (50)$$

This estimate (and, in the same way, the subspace-based one) can be further improved based on (16), which shows that $\mathbf{g}_0^0 = \mathbf{h}_0^0 \otimes \boldsymbol{\varphi}^0$ is a structured vector. Thus we can look for the nearest vector to $\mathbf{d}$ having this structure, that is, we can consider the following optimization problem:

$$\|\mathbf{h}\otimes\boldsymbol{\varphi} - \mathbf{d}\|^2 = \min, \quad \mathbf{h} \in \mathbb{C}^{Nn_r}, \ \boldsymbol{\varphi} \in \mathbb{R}^{M+1}. \quad (51)$$

Unfortunately, the cost function in (51) can be shown to have multiple minima, and no closed-form solution can be devised to compute its global minimum. A suitable strategy is to minimize this function alternately with respect to $\mathbf{h}$ and $\boldsymbol{\varphi}$, which yield the following iterative rule:

$$\begin{aligned}\mathbf{h}_n &= \frac{1}{||\boldsymbol{\varphi}_{n-1}||^2}\left(\mathbf{I}_{Nn_r}\otimes\boldsymbol{\varphi}_{n-1}\right)^H\mathbf{d},\\ \boldsymbol{\varphi}_n &= \frac{1}{||\mathbf{h}_n||^2}\Re\left(\left(\mathbf{h}_n\otimes\mathbf{I}_{M+1}\right)^H\mathbf{d}\right),\\ \hat{\mathbf{g}}_0^0(n) &= \mathbf{h}_n\otimes\boldsymbol{\varphi}_n,\end{aligned} \quad (52)$$

where we have denoted by $\hat{\mathbf{g}}_0^0(n)$ the estimate of $\mathbf{g}_0^0$ at the nth iteration. Note that convergence of this procedure to the global minimum is not guaranteed; however, experimental evidence has shown that after few iteration (i.e., 3–4), a fixed point is reached.

3.4. Gain vector estimation

If a noncooperative scheme with maximal ratio combining is adopted, after we have realized the n_r receivers, one for each antenna, a further processing is needed in order to get an estimate of the gain vector $\tilde{\mathbf{a}}$.

Assuming again complete suppression of all of the interference, (21) becomes

$$\boldsymbol{\vartheta}(i) = \tilde{\mathbf{a}}b_0^0(i) + \tilde{\mathbf{w}}. \quad (53)$$

A simple blind method for estimating $\widetilde{\mathbf{a}}$ (see [21]) can be developed noticing that the correlation matrix of $\boldsymbol{\vartheta}(i)$ is given by[5]

$$\mathbf{R}_{\vartheta\vartheta} = \widetilde{\mathbf{a}}\widetilde{\mathbf{a}}^H + 2\mathcal{N}_0\mathbf{I}_{n_r}. \tag{54}$$

Thus, the eigenvector corresponding to the largest eigenvalue of $\mathbf{R}_{\vartheta\vartheta}$ is parallel to $\widetilde{\mathbf{a}}$ and so, except for a complex scaling factor, it is an estimate of the gain vector $\widetilde{\mathbf{a}}$ (note that the phase ambiguity introduced by this complex constant is removed by the differential detection rule). Finally, note that this estimation technique can be easily made adaptive using the tracking algorithm suggested in [21].

3.5. Maximum number of users and system complexity

The identifiability condition sets a limit on the maximum rank of $\mathbf{R}_{qq}$ and, consequently, on the maximum number of users, $K_{\max}$ say, that the system can accommodate reliably. Since, based on (39),

$$\begin{aligned} 2MNn_r &\geq \dim\left(\operatorname{Im}\left(\mathbf{R}_{qq} + \mathbf{S}^0_{0,0}\mathbf{S}^{0H}_{0,0}\right)\right) \\ &= 3Kn_t + (M+1)Nn_r - 1, \end{aligned} \tag{55}$$

we have

$$K \leq \left\lfloor \frac{(M-1)Nn_r + 1}{3n_t} \right\rfloor. \tag{56}$$

Recalling that each user is assigned n_t spreading sequences, the maximum number of active users is

$$\begin{aligned} K_{\max} &= \left\lfloor \frac{(M-1)N+1}{3n_t} \right\rfloor, \\ K_{\max} &= \min\left\{ \left\lfloor \frac{(M-1)Nn_r+1}{3n_t} \right\rfloor, \frac{MN}{n_t} \right\} \end{aligned} \tag{57}$$

for noncooperative and cooperative detection, respectively. Note that the cooperative detection scheme, jointly elaborating the signals received at the n_r antennae, achieves better BER performance and, at the same time, can accommodate a larger number of users than the noncooperative scheme, as expected, at the price of some complexity increase. In fact, due to the matrix inversion in the first stage and to the singular value decomposition in the second one, the receiver complexity is cubic with the dimension of $\widehat{\mathbf{R}}_{rr}$, that is, the complexity is $\mathcal{O}((MNn_r)^3)$. Noncooperative receivers, instead, rely on n_r parallel operations conducted on matrices of order $2MN$ and entail a complexity $\mathcal{O}(n_r(MN)^3)$. Note, however, that, coupling a recursive least squares (RLS) procedure with subspace tracking techniques as in [18, 19], the overall complexity can be limited to be quadratic, that is, $\mathcal{O}((n_rMN)^2)$ and $\mathcal{O}(n_r(MN)^2)$ for cooperative and noncooperative detection, respectively. Moreover, since n_r is not very large for real applications, the complexity increase involved by cooperative over the noncooperative detection is often negligible.

[5]Note that the channel attenuations and thermal noise are "spatially" uncorrelated and that the receiver filters $\mathbf{m}_r$ have unit energy.

A final key remark is now in order. Conditions (57) represent the extension to the case of MC DS-CDMA employing multiple transmit and receive antennae of the condition reported in [19] for single-antenna DS-CDMA systems employing rectangular chip pulses. As already anticipated, such an identifiability condition cannot be relaxed through signal-space oversampling, once bandlimited waveforms are employed. Indeed, adopting rectangular pulses corresponds to enlarging the bandwidth beyond $1/T_c$ and to using infinite effective bandwidth which in turn corresponds to a theoretically infinite precision in delay estimation (see [20]). Thus, in the case of asynchronous systems with unknown delays, the DS-CDMA multiplex actually spans, in the *ensemble* of the delays realizations, an infinite-dimensional space whose principal directions can be in principle resolved by progressively enlarging the front-end bandwidth (i.e., "oversampling" by a factor L, which corresponds to chip-matched filtering through a unit-height pulse of duration T_c/L and sampling at rate L/T_c). In the considered strictly bandlimited scenario, instead, the signal span is strictly finite, whereby there appear to be just two alternatives in order to increase the maximum user number: the former is obviously an increase of the number of receive antennae, while the latter, that we just mention here, is to enlarge the processing window.

Before moving on to the statistical analysis of the proposed detection schemes, it is worth commenting on the two-stage receiver family introduced in this section. First, notice that the functional split between the interference cancellation and the BER maximization stages results in a greater flexibility at a design level; indeed, the blocking matrix $\mathbf{D}$ may be designed according to several different criteria, mainly depending on the intensity of the interfering users, without affecting the structure of the BER optimization stage. Additionally, even though we do not dwell on this issue here, it is natural to investigate the feasibility of *adaptive* (on a bit-by-bit scale) blind systems. Notice that, in our scenario, several different time-scales can be envisaged for channel variations: the abrupt changes in the MAI, wherein new users may enter the network and former users may abandon it, short-term variations in the channel tap-weights, and long-term variations in the temporal and spatial signatures of the active users. Notice also that the MAI structure affects only the interference-blocking stage of the proposed receiver, and would in principle require a self-recovering updating of the blocking matrix $\mathbf{D}$, which is indeed the focus of current research. As for the long-term variations, it is reasonable to assume that their time scale is large enough so as to allow batch processing with offline estimation of the relevant statistical measures. An open problem is, instead, the handling of short-term variations, which have an impact on both stages of the receiver. At an intuitive level, one might expect that the interference-blocking matrix design criterion should be modified in order to ensure nonzero output signal in the *ensemble* of the channel tap-weights realizations, which expectedly results in a set of constraints dictated by the covariance matrix of the channel taps. Additionally, constrained-complexity tracking procedures should be

introduced in order to adapt the BER optimization stage in such a time-varying scenario. All of the above issues form the objects of current investigations.

4. STATISTICAL ANALYSIS

In this section, we develop a statistical performance analysis of the proposed receiver and, in particular, we derive analytical expressions for the conditional error probability and near-far resistance, given the timing and the channel realizations of all of the users, that is, conditioned on the vector

$$\mathbf{g} = \left(\mathbf{g}_0^{0T} \cdots \mathbf{g}_t^{kT} \cdots \mathbf{g}_{n_t-1}^{(K-1)T}\right)^T \in \mathbb{C}^{Kn_t(M+1)Nn_r}. \tag{58}$$

4.1. Probability of error

First of all, recall that the decision rule is written as

$$\begin{aligned}\widehat{d}_0^0(i) &= \operatorname{sgn}\left[\Re\left(\langle \mathbf{r}(i), \mathbf{m}\rangle \overline{\langle \mathbf{r}(i-1), \mathbf{m}\rangle}\right)\right] \\ &= \operatorname{sgn}\left(\frac{1}{2}x_i\overline{x}_{i-1} + \frac{1}{2}\overline{x}_i x_{i-1}\right),\end{aligned} \tag{59}$$

where

$$\begin{aligned}x_i &= \langle \mathbf{r}(i), \mathbf{m}\rangle = b_0^0(i)\mathbf{m}^H\mathbf{S}_{0,0}^0\mathbf{g}_0^0 + \mathbf{m}^H\mathbf{z}(i) + \mathbf{m}^H\mathbf{w}(i) \\ &= b_0^0(i)\mathbf{m}^H\mathbf{s}_{0,0}^0 + \zeta(i).\end{aligned} \tag{60}$$

Assuming that the MAI plus ISI contribution $\mathbf{m}^H\mathbf{z}(i)$ at the output of the filter is approximately Gaussian with zero mean (see [22]), the term $\zeta(i)$ in (60) can be modeled as a complex Gaussian random variate with zero mean. Thus, given $\mathbf{g}$ and $b_0^0(i)$, the random variable x_i is itself Gaussian and

$$\begin{aligned}\mu &= E\left[x_i \,|\, \mathbf{g}, b_0^0(i)\right] = b_0^0(i)\mathbf{m}^H\mathbf{s}_{0,0}^0, \\ c(0) &= \operatorname{Var}\left(x_i \,|\, \mathbf{g}, b_0^0(i)\right) = E\left[\,|\zeta_i|^2\right] \\ &= \mathbf{m}^H\left(2\mathcal{N}_0\mathbf{I}_{2NMn_r} + \mathbf{R}_z\right)\mathbf{m}, \\ c(1) &= \operatorname{Cov}\left(x_i, x_{i-1} \,|\, \mathbf{g}, b_0^0(i), b_0^0(i-1)\right) = E\left[\zeta_i\overline{\zeta}_{i-1}\right] \\ &= \mathbf{m}^H\left(\mathbf{R}_{w(i)w(i-1)} + \mathbf{R}_{z(i)z(i-1)}\right)\mathbf{m},\end{aligned} \tag{61}$$

where

$$\begin{aligned}\mathbf{R}_{z(i)z(i)} &= E\left[\mathbf{z}(i)\mathbf{z}^H(i)\right] = \mathbf{s}_{0,-1}^0\mathbf{s}_{0,-1}^{0H} + \mathbf{s}_{0,+1}^0\mathbf{s}_{0,+1}^{0H} \\ &\quad + \sum_{t=1}^{n_t-1}\left[\mathbf{s}_{t,-1}^0\mathbf{s}_{t,-1}^{0H} + \mathbf{s}_t^0\mathbf{s}_t^{0H} + \mathbf{s}_{t,+1}^0\mathbf{s}_{t,+1}^{0H}\right] \\ &\quad + \sum_{k=1}^{K-1}\sum_{t=0}^{n_t-1}\left[\mathbf{s}_{t,-1}^k\mathbf{s}_{t,-1}^{kH} + \mathbf{s}_t^k\mathbf{s}_t^{kH} + \mathbf{s}_{t,+1}^k\mathbf{s}_{t,+1}^{kH}\right], \\ \mathbf{R}_{z(i)z(i-1)} &= \sum_{t=1}^{n_t-1}\left[\mathbf{s}_{t,-1}^0\mathbf{s}_t^{0,H} + \mathbf{s}_t^0\mathbf{s}_{t,+1}^{0H}\right] \\ &\quad + \sum_{k=1}^{K-1}\sum_{t=0}^{n_t-1}\left[\mathbf{s}_{t,-1}^k\mathbf{s}_t^{kH} + \mathbf{s}_t^k\mathbf{s}_{t,+1}^{kH}\right], \\ \mathbf{R}_{w(i)w(i-1)} &= \mathbf{I}_{n_r} \otimes \begin{pmatrix}\mathbf{O}_{MN,MN} & 2\mathcal{N}_0\mathbf{I}_{MN} \\ \mathbf{O}_{MN,MN} & \mathbf{O}_{MN,MN}\end{pmatrix}.\end{aligned} \tag{62}$$

Notice also that $\mathbf{R}_{z(i)z(i)}$ and $\mathbf{R}_{z(i)z(i-1)}$ are equal to the null matrix if the ZF-like receiver is adopted. Since the probability of error can be written as

$$\begin{aligned}P_{e|\mathbf{g}} &= P_{e|\mathbf{g},d_0^0(i)=1} \\ &= P\left(\frac{1}{2}x_i\overline{x}_{i-1} + \frac{1}{2}\overline{x}_i x_{i-1} < 0 \,|\, \mathbf{g}, d_0^0(i) = 1\right),\end{aligned} \tag{63}$$

and since $(1/2)x_i\overline{x}_{i-1} + (1/2)\overline{x}_i x_{i-1}$ is a quadratic form in correlated complex-valued Gaussian random variables, upon defining

$$\begin{aligned}a &= |\mu|\sqrt{\frac{c(0) - \sqrt{c^2(0) - \Im^2(c(1))}}{c^2(0) - \Im^2(c(1))}}, \\ b &= |\mu|\sqrt{\frac{c(0) + \sqrt{c^2(0) - \Im^2(c(1))}}{c^2(0) - \Im^2(c(1))}}, \\ \alpha &= \frac{1}{2}\left(1 + \frac{\Re(c(1))}{\sqrt{c^2(0) - \Im^2(c(1))}}\right),\end{aligned} \tag{64}$$

and using the results in [23], we obtain

$$\begin{aligned}P_{e|\mathbf{g}} &= Q_1(a,b) - \alpha I_0(ab)e^{-(a^2+b^2)/2}, \\ P_e &= E_{\mathbf{g}}[P_{e|\mathbf{g}}].\end{aligned} \tag{65}$$

Notice that (65) is the expression of the probability of error of any linear receiver employing differential data detection. In order to obtain the unconditional error probability, we should carry out the expectation with respect to the vector $\mathbf{g}$; however, this task cannot be easily accomplished, whereby we resort to a numerical average over a finite number of random realizations of $\mathbf{g}$.

So far, the case of a cooperative reception has been analyzed; moving to the noncooperative receiving scheme with hard integration, denote by p_e the conditional probability of error over each of the n_r receive antennae (note that p_e can be computed with the same approach as in the case of cooperative detection); since the channel gains and the thermal noise are assumed independent across the receive antennae, the hard integration strategy amounts to a Bernoulli counting and the overall probability of error is easily shown to be written as follows:

(i) $P_{e|\mathbf{g}}^{HI} = \sum_{i=(n_r+1)/2}^{n_r}\binom{n_r}{i}p_e^i(1-p_e)^{n_r-i}$ for n_r odd;

(ii) $P_{e|\mathbf{g}}^{HI} = \sum_{i=n_r/2+1}^{n_r}\binom{n_r}{i}p_e^i(1-p_e)^{n_r-i} + (1/2)\binom{n_r}{n_r/2}p_e^{n_r/2}(1-p_e)^{n_r/2}$ for n_r even.

Determining an analytical expression for the error probability in the case of noncooperative reception with soft integration is quite involved a task. Indeed, in this case, the test statistic can be expressed through the quadratic form $\sum_{j=0}^{n_r-1}((1/2)x_i^j\overline{x}_{i-1}^j + (1/2)\overline{x}_i^j x_{i-1}^j)$, where the n_r pairs $\{x_i^j, x_{i-1}^j\}$

are Gaussian variates, statistically independent of each other but not identically distributed, thus implying that the results in [23] cannot be directly applied. For the sake of brevity, we do not dwell any further on this issue, and just point out that the system error probability in this scenario is lower and upper bounded by those of the cooperative scheme and noncooperative scheme with hard integration, respectively.

4.2. Near-far resistance

For a multiuser detector, the asymptotic efficiency and the near-far resistance for the 0th transmit antenna of the 0th user are defined as follows:

$$\eta = \sup\left\{r \in [0,1] : \lim_{\mathcal{N}_0 \to 0} \frac{P_e(\mathcal{E}_b^0/\mathcal{N}_0)}{P_e^o(r\mathcal{E}_b^0/\mathcal{N}_0)} < +\infty\right\},$$
$$\overline{\eta} = \inf_{\substack{\mathcal{E}_b^i \geq 0 \\ i \neq 0}} \{\eta\}, \tag{66}$$

respectively, where P_e^o is the probability of error of the optimum receiver (maximum likelihood) for an isolated system (i.e., with no other user except the 0th one); the performance measures in (66) determine the loss due to the presence of the MAI in the limit of very low background noise. We just focus on the ZF-type receiver, since the MMSE-like solution converges to the ZF-like one as $\mathcal{N}_0$ vanishes.

First of all, note that if (25) is met, the proposed receiver achieves asymptotic multiuser efficiency, since the first stage is able to completely suppress interference (see (37)). However, as P_e cannot be easily computed in a closed form, in the sequel we condition on the vector $\mathbf{g}$ and consider the following *conditional* near-far resistance:

$$\overline{\eta}(\mathbf{g}) = \inf_{\substack{\mathcal{E}_b^i \geq 0 \\ i \neq 0}} \left\{ \sup\left\{r \in [0,1] : \lim_{\mathcal{N}_0 \to 0} \frac{P_{e|\mathbf{g}}(\mathcal{E}_b^0/\mathcal{N}_0)}{P_{e|\mathbf{g}}^o(r\mathcal{E}_b^0/\mathcal{N}_0)} < +\infty\right\}\right\}. \tag{67}$$

Note that even though $\overline{\eta}(\mathbf{g})$ does not coincide with the actual system near-far resistance $\overline{\eta}$, it is still a measure of the receiver capability to combat interference with arbitrarily large strength in the low-noise region: precisely, $\overline{\eta}(\mathbf{g})$ is the near-far resistance that the receiver experiences during the transmission of a frame.

Now, since a closed-form expression of $P_{e|\mathbf{g}}^o$ is not available, a lower bound for $\overline{\eta}(\mathbf{g})$ can be obtained by replacing $P_{e|\mathbf{g}}^o$ itself with the error probability $Q(\sqrt{\|\mathbf{s}_{0,0}^0\|^2/\mathcal{N}_0})$ of a synchronous single-antenna system employing binary phase-shift keying; thus, we have

$$\overline{\eta}(\mathbf{g}) \geq \inf_{\substack{\mathcal{E}_b^i \geq 0 \\ i \neq 0}} \left\{ \sup\left\{r \in [0,1] : \lim_{\mathcal{N}_0 \to 0} \frac{P_{e|\mathbf{g}}(\mathcal{E}_b^0/\mathcal{N}_0)}{Q\left(\sqrt{r||\mathbf{s}_{0,0}^0||^2/\mathcal{N}_0}\right)} < +\infty\right\}\right\}. \tag{68}$$

Now, we evaluate this parameter. For the ZF-like receiver, the quantities in (61) and (64) simplify to

$$\begin{aligned}
\mu &= b_0^0(i)\mathbf{m}^H\mathbf{s}_{0,0}^0,\\
c(0) &= 2\mathcal{N}_0\|\mathbf{m}\|^2,\\
c(1) &= \mathbf{m}^H\mathbf{R}_{n(i)n(i-1)}\mathbf{m} = 2\mathcal{N}_0\nu,\\
a &= \frac{1}{\sqrt{2\mathcal{N}_0}}\left|\mathbf{m}^H\mathbf{s}_{0,0}^0\right|\sqrt{\frac{\|\mathbf{m}\|^2 - \sqrt{\|\mathbf{m}\|^4 - \Im^2(\nu)}}{\|\mathbf{m}\|^4 - \Im^2(\nu)}} = \frac{\xi}{\sqrt{\mathcal{N}_0}},\\
b &= \frac{1}{\sqrt{2\mathcal{N}_0}}\left|\mathbf{m}^H\mathbf{s}_{0,0}^0\right|\sqrt{\frac{\|\mathbf{m}\|^2 + \sqrt{\|\mathbf{m}\|^4 - \Im^2(\nu)}}{\|\mathbf{m}\|^4 - \Im^2(\nu)}} = \frac{\phi}{\sqrt{\mathcal{N}_0}},\\
\alpha &= \frac{1}{2}\left(1 + \frac{\Re(\nu)}{\sqrt{\|\mathbf{m}\|^4 - \Im^2(\nu)}}\right),
\end{aligned} \tag{69}$$

respectively, and the probability of error for the 0th transmit antenna of the 0th user given $\mathbf{g}$ in (65) can be also written as follows:

$$P_{e|\mathbf{g}} = Q_1\left(\frac{\xi}{\sqrt{\mathcal{N}_0}}, \frac{\phi}{\sqrt{\mathcal{N}_0}}\right) - \alpha I_0\left(\frac{\xi\phi}{\mathcal{N}_0}\right)e^{-(\xi^2+\phi^2)/(2\mathcal{N}_0)}. \tag{70}$$

Since $Q_1(\xi/\sqrt{\mathcal{N}_0}, \phi/\sqrt{\mathcal{N}_0})$ and $I_0(\xi\phi/\mathcal{N}_0)e^{-(\xi^2+\phi^2)/(2\mathcal{N}_0)}$ are both asymptotic functions, for $\mathcal{N}_0 \to 0$, to $Q((\phi-\xi)/\sqrt{\mathcal{N}_0})$ (see [24]), the conditional near-far resistance admits the following lower bound:

$$\overline{\eta}(\mathbf{g}) \geq \begin{cases} \dfrac{(\phi-\xi)^2}{||\mathbf{s}_{0,0}^0||^2} = \dfrac{\left|\langle \mathbf{m}, \mathbf{s}_{0,0}^0\rangle\right|^2}{||\mathbf{s}_{0,0}^0||^2(\|\mathbf{m}\|^2 + \Im(\nu))}, & \text{if } K \leq K_{\max},\\ 0, & \text{otherwise.}\end{cases} \tag{71}$$

It is obviously understood that averaging the above quantity with respect to $\mathbf{g}$ leads to a sort of *average near-far resistance*, that is, the near-far resistance experienced, on the average, by the receiver during the transmission of many (theoretically infinite) packets; in this case too, the expectation with respect to the vector $\mathbf{g}$ can be evaluated numerically.

5. NUMERICAL RESULTS

In this section, we discuss numerical results illustrating the performance of the proposed receivers. We use both semi-analytical procedures exploiting the previously derived analytical formulas, and plain Monte Carlo simulations. In both situations, the curves shown will be the result of an average over 10^4 channels and delays realizations. We assume that

(a) each user is equipped with two transmit antennae;
(b) the convolution $(\psi_{tx} * \psi_{rx})(\tau) = \varphi(\tau)$ is a raised cosine with duration $4T_c$ ($\Delta = 4$) and roll-off factor 0.22;
(c) the number of subcarriers is $N = 4$ and the spreading over each one is $M = 8$ (the composite spreading gain is then $PG = 32$ and the spreading sequences are PN $\in \{-1, 1\}$ of length 31 stretched out with a -1);
(d) the sample correlation matrix $\hat{\mathbf{R}}_{rr}$ is obtained through a sample estimate over $Q = 1300$ samples.

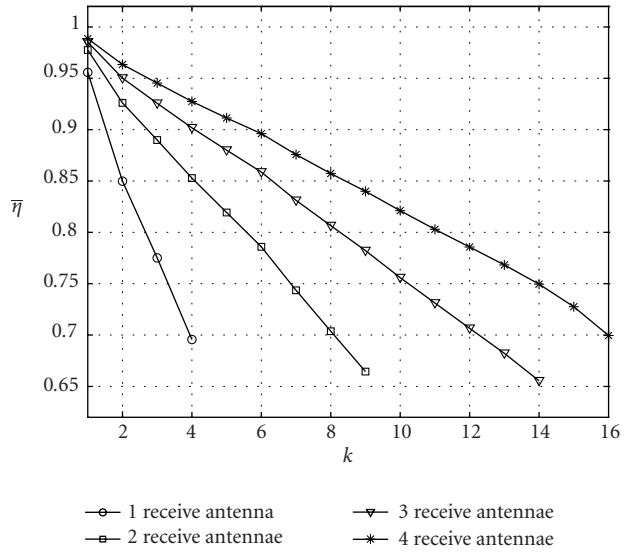


FIGURE 4: Lower bound for the average near-far resistance of the two-stage receiver versus the number of users for several n_r: $M = 8, N = 4$, and $n_t = 2$.

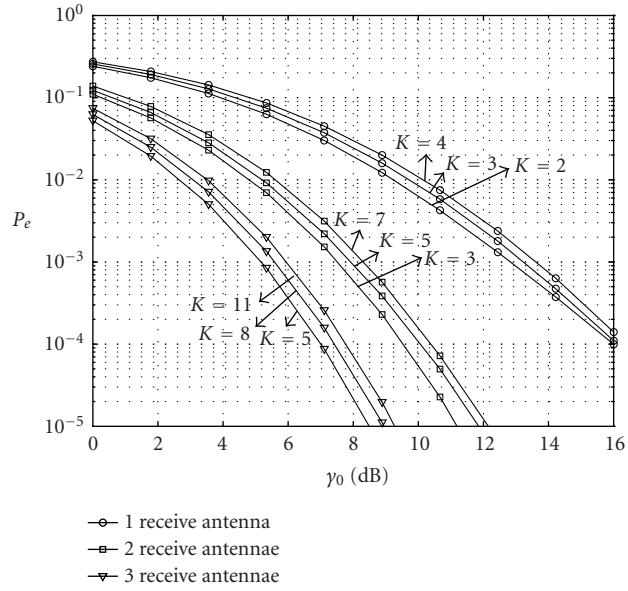


FIGURE 5: Probability of error for the nonblind ZF-like receiver versus γ_0 for different number of users and of receive antennae: $M = 8, N = 4$, and $n_t = 2$.

In Figure 4, the computed lower bound for the *average near-far resistance* of the two-stage receiver with cooperative detection is represented versus the number of active users for different number of receive antennae ($n_r = 1, \ldots, 4$). Results show that the proposed receiver is near-far resistant and, also, that increasing the number of receive antennae yields a remarkable performance improvement (note that for $n_r = 4$, the limiting factor of the number of users is no longer dictated by (25), but by the number of available spreading sequences, that is, sixteen times two). Figures 5 and 6 show the probability of error (obtained through the semianalytical procedure) of the nonblind receivers with cooperative detection versus the ratio $\gamma_0 = \mathcal{E}_b^0/\mathcal{N}_0$, for several values of the number of receive antennae and of active users. It is here assumed that perfect average power control has been pursued, even though, due to the said near-far resistance feature of the proposed receivers, the system performance is only slightly degraded in a near-far scenario. It is seen from Figure 5 that as the number of receive antennae grows, the receiver performance improves and, for a fixed error probability value, a higher number of users can be accommodated. On the other hand, Figure 6 shows the error probability for different receivers and fixed number of users, that is, $K = 4$. It can be seen that the MMSE-like receiver behaves slightly worse than the MMSE one for $n_r = 1$, while for $n_r = 2$, all the nonblind receivers exhibit the same performance. Simulation results, not provided here for the sake of brevity, have also confirmed a perfect agreement between the semianalytical procedure and the Monte Carlo-based performance evaluation technique.

With regard to the performance of the blind receivers, results of Monte Carlo simulations are presented in Figures 7, 8, 9, 10, 11, and 12 for a severe near-far scenario (the interfering users are 15 dB above the user of interest) and with $K = 4$ active users, for both cooperative and noncooperative

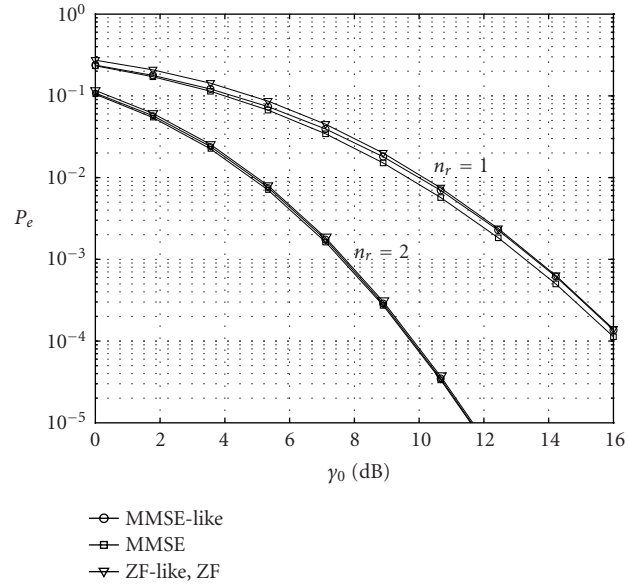


FIGURE 6: Probability of error for the nonblind receivers: MMSE-like, MMSE, ZF-like, and ZF: $K = 4, M = 8, N = 4$, and $n_t = 2$.

reception (observe that the maximum number of user $K_{\max}$ for the noncooperative scheme is 4 implying that the network is fully loaded). Figures 7 and 8 show the performance of the proposed subspace-based channel estimation procedure for a noncooperative and a cooperative reception scheme, respectively. In particular, the correlation coefficient

$$\rho = \frac{|\langle \hat{\mathbf{g}}_0^0, \mathbf{g}_0^0 \rangle|}{||\hat{\mathbf{g}}_0^0||\,||\mathbf{g}_0^0||} \tag{72}$$

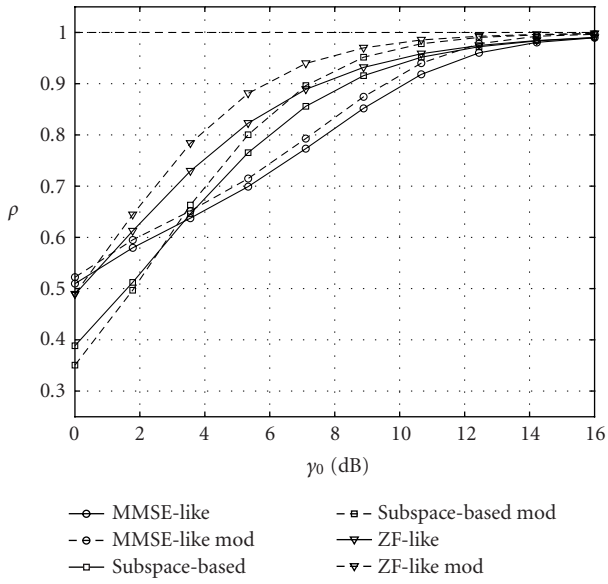


Figure 7: Channel estimation for a noncooperative reception scheme: $K = 4$, $M = 8$, $N = 4$, $n_t = 2$, $n_r = 2$, $Q = 1300$, and $P = 1500$.

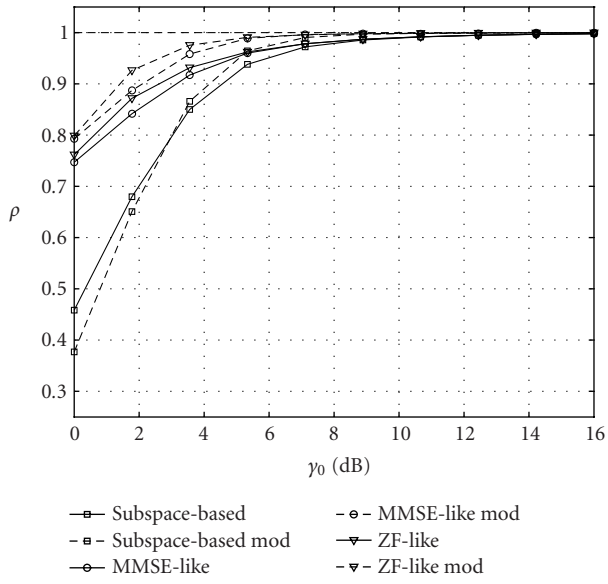


Figure 8: Channel estimation for a cooperative reception scheme: $K = 4$, $M = 8$, $n_t = 2$, $n_r = 2$, $Q = 1300$, and $P = 1500$.

is reported versus γ_0. Here, the word "mod" in the legends refers to the improved channel estimation rule in (52). Figures 9 to 11 show the system error probability for the noncooperative scheme with hard and soft integration, and for the cooperative scheme, respectively. Here, the curve labeled as "MMSE-like limit" reports the performance of the MMSE-like receiver in the limit of increasingly large size Q of the sample set used to estimate the covariance matrix of the data. Inspecting the figures, it is seen that in the noncooperative case, with the network fully loaded, the best channel esti-

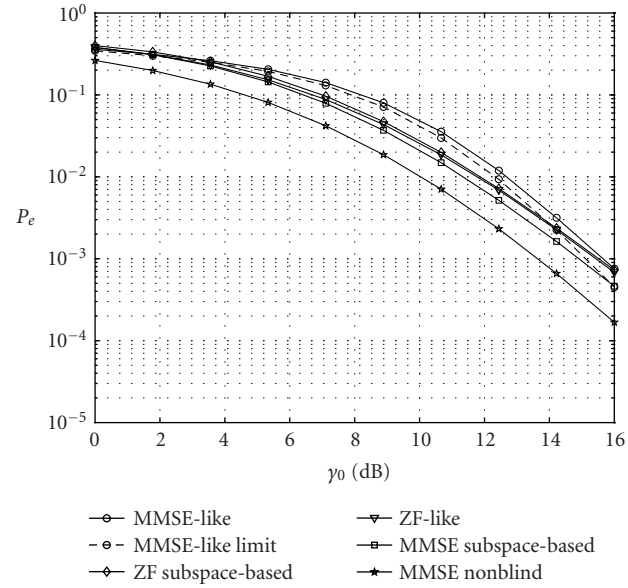


Figure 9: Probability of error for the blind receivers; noncooperative reception with hard integration: $K = 4$, $M = 8$, $n_t = 2$, $n_r = 2$, $Q = 1300$, and $P = 1500$.

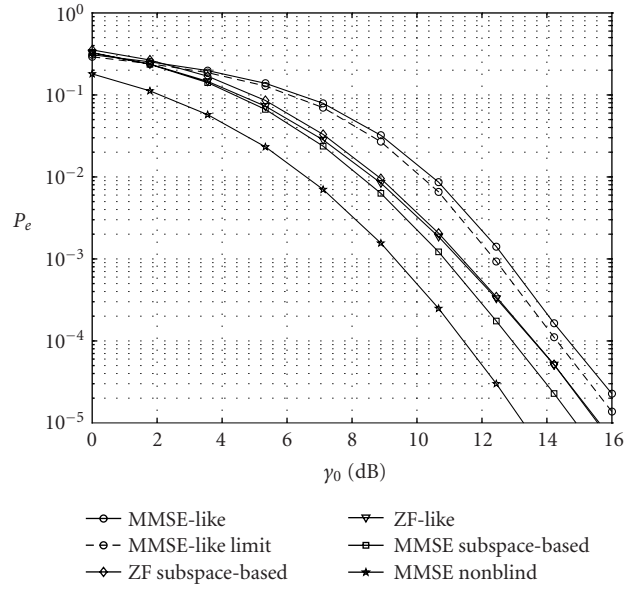


Figure 10: Probability of error for the blind receivers; noncooperative reception with soft integration: $K = 4$, $M = 8$, $N = 4$, $n_t = 2$, $n_r = 2$, $Q = 1300$, and $P = 1500$.

mation is achieved by the ZF-like receiver, immediately followed by the subspace-based one, while for the cooperative case, both the ZF-like and the MMSE-like receivers outperform the subspace-based one. This trend is confirmed in the plots showing the error probability; indeed, the ZF-like receiver performs slightly better then the ZF subspace-based one in both cases while the MMSE-like receiver outperforms the subspace-based receiver only in the cooperative case. It is

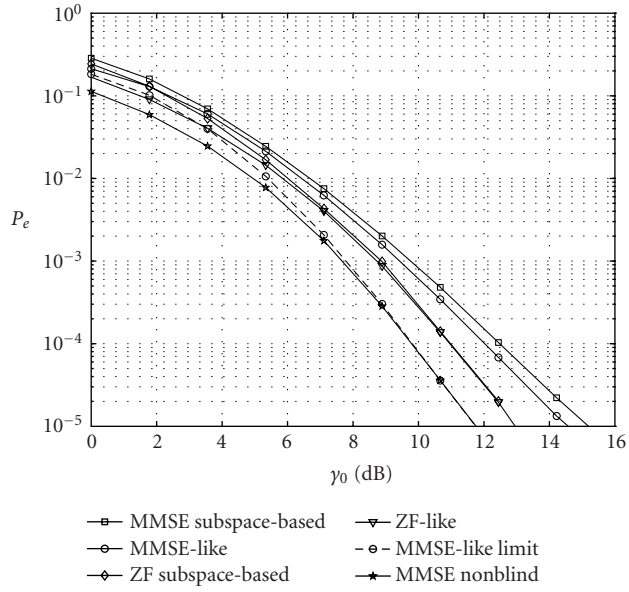


Figure 11: Probability of error for the blind receivers; cooperative reception scheme: $K = 4$, $M = 8$, $N = 4$, $n_t = 2$, $n_r = 2$, $Q = 1300$, and $P = 1500$.

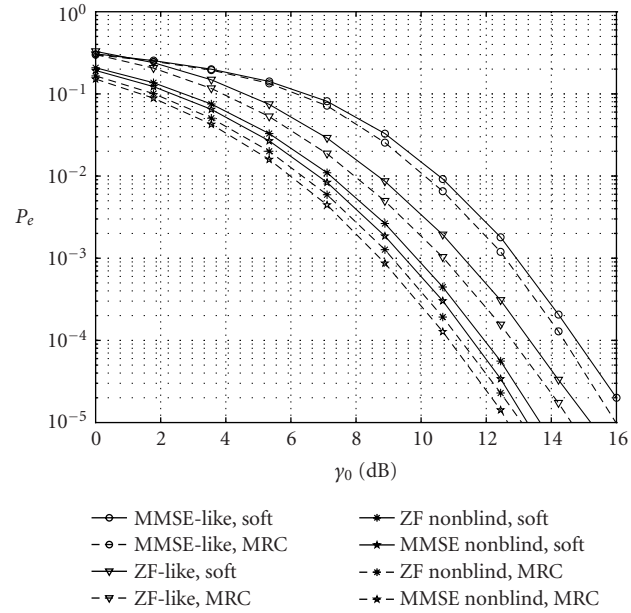


Figure 12: Probability of error for both soft integration and maximal ratio combiner in a noncooperative scheme: $K = 4$, $M = 8$, $N = 4$, $n_t = 2$, $n_r = 2$, $Q = 1300$, $P = 1500$, and $Q_2 = 1000$.

also seen that the soft integration achieves superior performance with respect to the hard integration scheme and that both of them incur a loss with respect to the cooperative reception. Notice that for the noncooperative receiver, due to the network full load, the MMSE-like limit performance is not coincident with that of the ideal MMSE receiver; conversely, for cooperative reception, since now the users number is smaller than the maximum one, the MMSE-like limit curve is quite coincident with the ideal MMSE receiver performance. Finally, in Figure 12 a comparison between the error probability of the soft integration and MRC techniques in a noncooperative reception scheme is provided. Notice that, at the price of some complexity increase, the MRC scheme achieves better results with respect to the soft integration one for the nonblind receivers; on the other hand, concerning the blind receivers, the performance improvement is less evident due to the not perfect estimation of the vector gain ($\hat{R}_{\vartheta\vartheta}$ was obtained though a sample estimate over $Q_2 = 1000$ samples).

6. CONCLUSIONS

In this paper, we have considered the problem of blind multiuser detection for asynchronous MC DS-CDMA systems equipped with multiple transmit and receive antennae. This is nowadays an interesting research topic, since MC modulation formats coupled with the use of multiple antennae represent a suitable means to achieve high data rates on the wireless channel at a reasonable computational and practical implementation cost.

The receivers that have been proposed here are code-aided in the sense that they require knowledge of the spreading code for the user of interest only, while no prior knowledge on the channel state and on the timing offset is needed. Several combining rules for the statistics obtained at the output of each antenna have been considered and assessed. A thorough statistical analysis has been derived for the proposed receivers (and for any linear receiver employing binary differential transmission), while the performance of the blind version has been evaluated through Monte Carlo simulations. Results have shown that these receivers exhibit performance levels close to those of the MMSE and ZF ones and that the use of multiple receive antennae has a beneficial impact on the system performance.

Future work on this topic comprises the consideration of space-time and space-frequency codes, as well as the extension of the proposed detection strategy to the situation in which the channel is time-dispersive, that is, it does not remain constant over the whole transmitted frame.

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewer A for his constructive comments and for suggesting the MRC combining rule reported in Section 3.4.

REFERENCES

[1] F. Adachi, M. Sawahashi, and H. Suda, "Wideband DS-CDMA for next-generation mobile communication systems," *IEEE Communications Magazine*, vol. 36, no. 9, pp. 56–69, 1998.

[2] L. Vandendorpe, "Multitone direct-sequence CDMA system in an indoor wireless environment," in *Proc. IEEE 1st Symp. Commun. and Veh. Technol.*, pp. 4.1.1–4.1.8, Delft, Netherlands, October 1993.

[3] L. Vandendorpe, "Multitone spread spectrum multiple access communication systems in a multipath Rician fading channel," *IEEE Trans. Vehicular Technology*, vol. 44, no. 2, pp. 327–337, 1995.
[4] N. Yee, J. P. Linnartz, and G. Fettweis, "Multicarrier CDMA in indoor wireless radio networks," in *Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications*, pp. 109–113, Yokohama, Japan, September 1993.
[5] X. Gui and T. S. Ng, "Performance of asynchronous orthogonal multicarrier CDMA system in frequency selective fading channel," *IEEE Trans. Communications*, vol. 47, no. 7, pp. 1084–1091, 1999.
[6] P. Zong, K. Wang, and Y. Bar-Ness, "Partial sampling MMSE interference suppression in asynchronous multicarrier CDMA system," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 8, pp. 1605–1613, 2001.
[7] S. Kondo and L. B. Milstein, "Performance of multicarrier DS CDMA systems," *IEEE Trans. Communications*, vol. 44, no. 2, pp. 238–246, 1996.
[8] S. L. Miller and B. J. Rainbolt, "MMSE detection of multicarrier CDMA," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 11, pp. 2356–2363, 2000.
[9] J. Namgoong, T. F. Wong, and J. S. Lehnert, "Subspace multiuser detection for multicarrier DS-CDMA," *IEEE Trans. Communications*, vol. 48, no. 11, pp. 1897–1908, 2000.
[10] L. Fang and L. B. Milstein, "Performance of successive interference cancellation in convolutionally coded multicarrier DS/CDMA systems," *IEEE Trans. Communications*, vol. 49, no. 12, pp. 2062–2067, 2001.
[11] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.
[12] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: performance criteria and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.
[13] G. J. Foschini and M. J. Gans, "On limits of wireless communication in a fading environment when using multiple antennae," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.
[14] H. Huang, H. Viswanathan, and G. J. Foschini, "Multiple antennae in cellular CDMA systems: transmission, detection, and spectral efficiency," *IEEE Trans. Wireless Commun.*, vol. 1, no. 3, pp. 383–392, 2002.
[15] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, 2003.
[16] X. Wang and H. V. Poor, "Blind equalization and multiuser detection in dispersive CDMA channels," *IEEE Trans. Communications*, vol. 46, no. 1, pp. 91–103, 1998.
[17] M. K. Tsatsanis and Z. D. Xu, "Performance analysis of minimum variance CDMA receivers," *IEEE Trans. Signal Processing*, vol. 46, no. 11, pp. 3014–3022, 1998.
[18] S. Buzzi, M. Lops, and H. V. Poor, "Timing-free code aided blind adaptive joint MAI and ISI suppression in dispersive CDMA channels," in *Proc. IEEE Wireless Communications and Networking Conference*, pp. 542–548, Orlando, Fla, March 2002.
[19] S. Buzzi, M. Lops, and H. V. Poor, "Blind adaptive joint multiuser detection and equalization in dispersive CDMA channels," *IEEE Trans. Signal Processing*, vol. 51, no. 7, pp. 1880–1893, 2003.
[20] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, vol. 1–3, John Wiley & Sons, 1971.
[21] X. Wang and H. V. Poor, "Bind multiuser detection: a subspace approach," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 677–690, 1998.
[22] H. V. Poor and S. Verdú, "Probability of error in MMSE multiuser detection," *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 858–871, 1997.
[23] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, 4th edition, 2001.
[24] S. Verdú, *Multiuser Detection*, Cambridge University Press, Cambridge, UK, 1998.

Stefano Buzzi was born in Piano di Sorrento, Italy, on December 10, 1970. He received with honors the Dr. Eng. degree in 1994, and the Ph.D. degree in electronic engineering and computer science in 1999, both from the University of Naples "Federico II." In 1996, he spent six months at Centro Studi e Laboratori Telecomunicazioni *(CSELT)*, Turin, Italy, while from November 1999 through December 2001, he spent eight months at the Department of Electrical Engineering, Princeton University, as a Visiting Research Fellow. He is currently an Associate Professor at the University of Cassino, Italy. His current research and study interests lie in the area of statistical signal processing, with emphasis on signal detection in non-Gaussian noise and multiple access communications. Dr. Buzzi is currently serving as an Associate Editor for the *Journal of Communications and Networks*. He was awarded by the Associazione Elettrotecnica ed Elettronica Italiana (AEI) the "G. Oglietti" scholarship in 1996, and was the recipient of a NATO/CNR advanced fellowship in 1999 and of a CNR short-term mobility Grant in 2000 and 2001.

Emanuele Grossi was born in Sora, Italy, on May 10, 1978. He received the Dr. Eng. degree in telecommunication engineering from the University of Cassino in 2002. He is currently a Ph.D. student at the University of Cassino, Italy. His current research interests concern wireless multiuser communication systems in CDMA applications, MIMO systems with space-time coding, and signal detection for radar systems.

Marco Lops was born in Naples, Italy, on March 16, 1961. He received the Dr. Eng. degree in electronic engineering from the University of Naples in 1986. From 1986 to 1987, he was in Selenia, Roma, Italy, as an Engineer in the Air Traffic Control Systems Group. In 1987, he joined the Department of Electronic and Telecommunications Engineering of the University of Naples as a Ph.D. student in electronic engineering. He received the Ph.D. degree in electronic engineering from the University of Naples in 1992. From 1991 to 2000, he was an Associate Professor of radar theory and digital transmission theory at the University of Naples, while, since March 2000, he is a Full Professor at the University of Cassino, engaged in research in the field of statistical signal processing, with emphasis on radar processing and spread spectrum multiuser communications. He also held teaching positions at the University of Lecce and, during 1991, 1998, and 2000, he was on sabbatical leaves at University of Connecticut, Rice University, and Princeton University, respectively. Dr. Lops is currently serving as an Associate Editor for the *Journal of Communications and Networks.*

EURASIP Journal on Applied Signal Processing 2004:5, 629–639

Signal Reception for Space-Time Differentially Encoded Transmissions over FIR Rich Multipath Channels

Zhan Zhang
Department of Electrical and Computer Engineering, Dalhousie University, 1360 Barrington Street, P.O. Box 1000, Halifax, NS, Canada B3J 2X4
Email: zhangz@dal.ca

Jacek Ilow
Department of Electrical and Computer Engineering, Dalhousie University, 1360 Barrington Street, P.O. Box 1000, Halifax, NS, Canada B3J 2X4
Email: j.ilow@dal.ca

Received 1 January 2003; Revised 28 November 2003

With sophisticated signal and information processing algorithms, air interfaces with space-time (ST) coding and multiple reception antennas substantially improve the reliability of wireless links. This paper proposes a new receiver algorithm for differential ST coded transmissions over the finite-impulse-response (FIR) rich multipath fading channels. The symbol detection introduced in this paper is a deterministic subspace-based approach in a multiple-input and multiple-output (MIMO) system framework. The receiver (i) operates in a blind fashion without estimating the channel or its inverse and (ii) is able to work with a small number of signal samples and hence can be applied in the quasistatic channels. The proposed scheme employs multiple antennas at both sides of the transceiver and exploits both the antenna diversity and the multiple constant modulus (MCM) characteristics of the signaling. The receiver is able to blindly mitigate the intersymbol interference (ISI) in a rich multipath propagation environment, and this has been verified through the extensive Monte Carlo simulations.

Keywords and phrases: rich multipath channels, space-time processing, transmit diversity, unitary group codes, signal subspace, constant modulus.

1. INTRODUCTION

Space-time (ST) multiple-input multiple-output (MIMO) transmission and reception is now regarded as one of the most effective approaches for increasing channel capacity or system fading-resistance [1, 2, 3, 4, 5, 6, 7]. Among a variety of ST coding schemes, differential ST modulation (DSTM) and differential space-code modulation (DSCM) are among the most promising ST coding schemes in wireless fading channels because of their efficient differential encoding and detection features [8, 9, 10, 11, 12]. Of particular interest to this paper is differential unitary group codes introduced in [8, 9, 12].

The differential schemes can work whether the channel state information (CSI) is available or not, and this is what makes them very attractive. When an accurate estimation of the CSI is difficult or costly, the DSTM schemes are obviously preferable than other schemes which assume full knowledge of the CSI.

As a recent development, a new type of ST block code is the Khatri-Rao ST code (KRST) proposed in [13], which possesses a built-in channel identifiability. It relies on the blind identifiability properties of the trilinear models and parallel factor analysis to estimate the channel states and to detect the ST symbols. However, there are some concerns about the convergence speed of its iterative algorithm. DSTM does not have such an issue. Compared to DSTM, KRST has a higher computational complexity at the receiver. The DSTM was designed to maximize the diversity advantage of the code while maintaining a receiver implementation to be as simple as possible.

The common point of the DSTM, DSCM, and the KRST is that they all assume a frequency-flat fading channel modeling in their design and analysis. In this paper, we consider reception of the DSTM signals under more realistic and complex channel conditions in rich multipath environment. Multipath scattering and reflection effects characterize most wireless channels. They cause both time and angle spreads. As a result, most wireless channels are selective in time, space, and frequency, and this is a reason why this paper addresses multipath frequency-selective impairments in the design of the ST receiver.

In contrast to the method presented in this paper, a combination of orthogonal frequency division multiplexing (OFDM) scheme with one of DSTM, DSCM, and KRST is feasible for transceiver designs over MIMO frequency-selective channels, because OFDM is capable of converting the frequency-selective channels into frequency-flat fading channels. Besides, to achieve the maximum diversity gain, a direct design of the frequency-ST coding scheme based on OFDM is also possible. An example is the transceiver proposed in [14]. However, the OFDM scheme has its own limitations: it has a very large peak to average power ratio, which demands a high linearity on the transmitter power amplifier. Nonlinearity of the system causes the intercarrier interference, which gives rise to the drastic degradation of the system performance. Moreover, performance of OFDM is more vulnerable to the frequency synchronization error than the conventional schemes, such as the single-carrier M-ary PSK, which the DSTM employs [15].

For channel equalizers requiring the channel estimation, the channel identification precision substantially affects the system performance. Small estimation bias may cause a severe performance degradation. In mobile communications, the channel changes quickly so that channel estimation is inefficient. Therefore, in this paper, channel estimation is neither assumed nor conducted in the algorithm. At the receiver, the transmitted data are recovered directly from the observed samples using an algebraic approach. Specifically, the new transceiver scheme consists of (i) a DSTM transmitter, (ii) an equalization algorithm based on direct input signal subspace estimation, and (iii) a differential ST symbol detector.

In general, the proposed receiver mitigates the multipath time-spread impairments without channel estimation provided that the channel is of rich multipath type so that its characterization matrix meets certain column-rank conditions. The approach used in this paper to recover the data relies on a modified version of signal subspace-based method introduced in [16]. The novelty of this paper stems from integrating subspace method based signal deconvolution and the exploitation of constant modulus property of the transmitted symbols to facilitate the noncoherent detection of DSTM signaling in a frequency-selective environment.

2. REVIEW OF DIFFERENTIAL ST MODULATION

In this section, the DSTM and unitary group codes [8] are briefly described for transmissions over frequency-flat fading channels. A transmitter equipped with K antennas and a receiver equipped with M antennas are assumed to constitute the transceiver system. A unitary ST codeword matrix $\mathbf{C}_j$ of size $K \times K$ is transmitted in the jth time slot T_j of duration $T_c = K \cdot T_s$, where j is the time index and T_s is the symbol duration. Each code matrix $\mathbf{C}_j$ is of the form $\mathbf{C}_j = \mathbf{C}_{j-1}\mathbf{G}_j$. Matrix $\mathbf{G}_j$ is chosen from a specific code set $\mathcal{G} = \{\mathbf{G}_{(m)} \mid \mathbf{G}_{(m)}\mathbf{G}_{(m)}^H = \mathbf{I}\}$ to represent user data, where m is the codeword index ($m = 1, 2, \ldots, \mathcal{M}$). The code has the property

$$\mathbf{C}_j\mathbf{C}_j^H = K\mathbf{I}_{K\times K}. \tag{1}$$

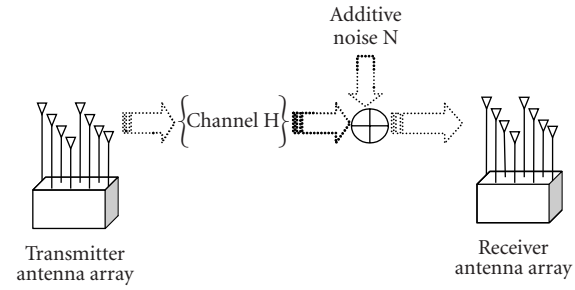


Figure 1: The general modeling of a multiantenna transceiver system.

It was proved in [9] that a full-rank unitary group code with $\mathcal{M} = 2^n$ codewords is equivalent to either a cyclic group code or a dicyclic group code. Assuming that the unknown frequency-flat fading channel is characterized by matrix $\mathbf{H} \in \mathcal{C}_{M\times K}$, the received data of the differentially ST coded signals at multiple receiving antennas are given as [8]

$$\mathbf{Y}_j = \mathbf{H}\mathbf{X}_j + \mathbf{N}_j, \tag{2}$$

where (i) the transmitted ST code is represented by $\mathbf{X}_j$, $j = 1, 2, \ldots, J$; (ii) J is a frame length in codewords; and (iii) $\mathbf{N}_j$ stands for the matrix version of additive white Gaussian noise (AWGN). With such modeling in a frequency-flat fading environment, a maximum likelihood (ML) decoder derived in [8] is

$$\widehat{\mathbf{G}}_j = \arg\max_{\mathbf{G}_{(m)}} \Re\{\operatorname{Tr}\{\mathbf{G}_{(m)}\mathbf{Y}_j^H\mathbf{Y}_{j-1}\}\}, \tag{3}$$

where $\Re$ stands for *real part* of the value and **Tr** denotes a *trace* computation. Hence, without knowing $\mathbf{H}$, the $\mathbf{G}_j$ can be estimated by observing the last two received data blocks $[\mathbf{Y}_{j-1}, \mathbf{Y}_j]$.

3. THE NEW RECEIVER ALGORITHM FOR TRANSMISSION OVER FIR RICH MULTIPATH FADING CHANNELS

3.1. Basis representations of the transmitted signals

In what follows, after a frame-by-frame DSTM transmitter is proposed, the discussion will focus on an equalization algorithm based on direct input signal subspace estimation.

The transmission scenario proposed in this paper for MIMO rich multipath channels is a frame-by-frame transmission/reception scheme illustrated in Figures 1 and 2, where T_c is a time slot for a codeword and $T_G \geqslant LT_s$ is a frame guard interval to avoid the interframe interference (L is the maximum channel length of the subchannels).

At the receiver, the continuous-time received signal vector $Y(t)$ is sampled at the symbol rate $(1/T_s)$ after down-

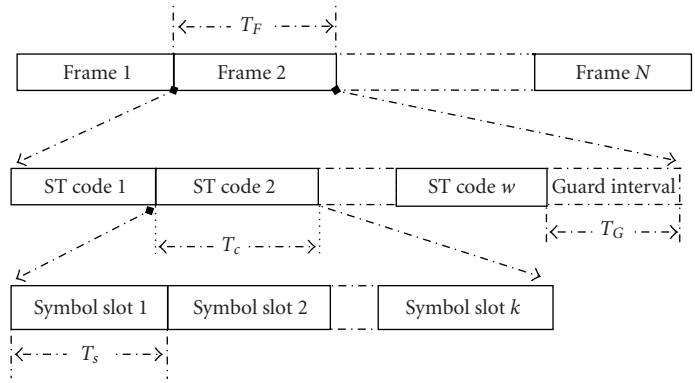


Figure 2: Transmitted signal frame structure and timing.

converting and reception filtering. For a period of signal frame (T_F), the sampled data sequence of $Y(t)$ at a receiver is arranged in a matrix form as follows:

$$\mathbf{Y}_{M\times(N+L)} = [\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{N+L-1}] = \begin{bmatrix} y_1(0) & y_1(1) & \cdots & y_1(N+L-1) \\ y_2(0) & y_2(1) & \cdots & y_2(N+L-1) \\ \vdots & \vdots & \vdots & \vdots \\ y_M(0) & y_M(1) & \cdots & y_M(N+L-1) \end{bmatrix}, \quad (4)$$

where (i) N is the frame length in symbols and (ii) $\mathbf{y}_i$ is a column vector of sampled data. We assume the quasistatic channel, namely, over the duration of one frame, the MIMO channel is time invariant.

According to the general modeling of MIMO channels, to capture the channel states, a matrix sequence $\{\mathbf{h}(i),\ i = 0, 1, \ldots, L\}$ is used. If the noise effects are temporarily disregarded and with the proper arrangement of data, we get the following input-output relation in a matrix format for the qth frame:

$$\mathbf{Y}^{[q]}_{M\times(N+L)} = \mathbf{H}_{M\times K(L+1)}\mathcal{X}^{[q]}_{K(L+1)\times(N+L)}, \quad (5)$$

where

$$\mathbf{H}_{M\times K(L+1)} = [\mathbf{h}(0), \mathbf{h}(1), \ldots, \mathbf{h}(L)];$$

$$\mathcal{X}^{[q]}_{K(L+1)\times(N+L)} = \begin{bmatrix} \mathbf{x}(0) & \mathbf{x}(1) & \cdots & \mathbf{x}(N-1) & 0 & \cdots & 0 \\ 0 & \mathbf{x}(0) & \cdots & \mathbf{x}(N-2) & \mathbf{x}(N-1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \mathbf{x}(0) & \mathbf{x}(1) & \cdots & \mathbf{x}(N-1) \end{bmatrix}; \quad (6)$$

and $\mathbf{x}(i)$ is a column vector $\mathbf{x}(i) = [x_1(i), x_2(i), \ldots, x_K(i)]^T$.

In order to retrieve input (transmitted) signals from the observation of convoluted received signals, first, a matrix sequence $\{\mathfrak{Y}^{(p)} \mid p = 0, 1, 2, \ldots, L\}$ is formed such that

$$\mathfrak{Y}^{(p)} = [\mathbf{y}_p, \mathbf{y}_{p+1}, \ldots, \mathbf{y}_{p+N-1}]; \quad p = 0, 1, 2, \ldots, L, \quad (7)$$

where $\mathfrak{Y}^{(p)}$ can be viewed as the vector subsequences of $[\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{N+L-1}]$ within a sliding window of width N corresponding to the shift $p = 0, 1, \ldots, L$.

For every $\mathfrak{Y}^{(p)}$, we calculate a matrix $\mathfrak{Z}^{(p)}$ which consists of the spanning row vector set, that is, the rows of $\mathfrak{Z}^{(p)}$ constitute the orthonormal basis for the subspace spanned by the rows of $\mathfrak{Y}^{(p)}$. The matrix $\mathfrak{Z}^{(p)}$ can be obtained by singular value decomposition (SVD) or some other efficient estimation methods. This processing is denoted in this paper by

$$\mathfrak{Y}^{(p)} \Longrightarrow \mathfrak{Z}^{(p)}, \quad p = 0, 1, 2, \ldots, L. \quad (8)$$

Proposition 1. *Let the row vector subspace of* $\mathbf{X}_{K\times N} = [\mathbf{x}(0)\ \mathbf{x}(1)\ \cdots\ \mathbf{x}(N-1)]$ *be denoted by* $\mathfrak{F}_X$. *In absence of the noise, the intersection of the row vector subspaces of* $\mathfrak{Z}^{(p)}$, $p = 0, 1, \ldots, L$, *is equivalent to* $\mathfrak{F}_X$ *with a probability of 1 for transmissions employing unitary ST group codes, provided* $\mathbf{H}$ *is of a full-column rank and the signal frame length* N *is sufficiently large for matrix* $\mathbf{X}$ *to have full-row rank.*

The proof of Proposition 1 is in the appendix. The full-column-rank assumption of $\mathbf{H}$ could be met with probability of 1 if it is a "tall" matrix with a row number larger than the column number if channel is of a rich multipath type. Evidently, if the subchannel lengths increase, accordingly, the number of reception antennas should be increased. Some auxiliary methods to facilitate meeting this assumption are discussed in Section 3.2. This assumption is a sufficient condition for the operation of the algorithm proposed in Section 4; however, it is not a necessary condition.

As a matter of fact, for the proposed algorithm, it is only assumed that some matrices among $\mathbf{h}(i)$, $i = 0, 1, \ldots, L$, individually have a full-column rank. This normally holds with probability of 1 for a rich multipath environment and the number of the reception antennas being larger than that of the transmission antennas. This assumption could be further relaxed by the data stacking method discussed in Section 3.2.

Defining a new matrix Ξ whose row vectors span the vector subspace intersection of $\Xi^{(p)}$, $p = 0, 1, \ldots, L$, and denoting it by $\Xi = \bigcap_{p=0}^{L} \Xi^{(p)}$, from Proposition 1, we have that the rows of Ξ also span subspace $\mathcal{S}_X$ with probability 1. Therefore,

$$\mathbf{X}_{K\times N} = \mathbf{W}_{K\times K}\Xi_{K\times N} \tag{9}$$

holds with probability 1, where $\mathbf{W}_{K\times K}$ is a weight matrix. Hence, with a proper $\mathbf{W}$, the transmitted signals could be recovered completely from $\mathcal{Y}^{(p)}$ by finding the spanning vector set using procedure of Proposition 1. In other words, the transmitted data could be recovered from $\mathcal{Y}^{(p)}$ within the ambiguity of a transformation $\mathbf{W}$.

The above observation is a fundamental point for the receiver algorithm design in this paper based on direct input signal subspace estimation. The estimation of $\mathbf{W}$ will be discussed in Section 4.2.

3.2. Column-rank assumption of channel matrices and oversampling

Regarding the assumption for the column rank of $\mathbf{h}(i)$, the following discussion is in order. As analyzed in [17], rich multipath scattering normally causes wide angle spreads. In these situations, the channels can be modeled using uncorrelated high-rank matrices. For MIMO frequency-flat fading channels, a formula suggested in [17] to predict a high-rank channel situation is

$$\frac{2D_t}{K-1}\frac{2D_r}{M-1} > \frac{R\lambda}{M}, \tag{10}$$

where (i) D_t, D_r stand for the transmission and reception scattering radius, respectively; (ii) R is the distance between transmitter and receiver; and (iii) λ is the wavelength. This formula indicates that a large number of scatters (and large antenna spacing), large angle spreading, and small range R will help in building up the high-rank MIMO channels in a frequency-flat fading modeling. High-rank MIMO channels can offer significant spatial multiplexing gain or diversity gain.

For MIMO frequency-selective channels, the above prediction method is applicable to channel matrices among $\mathbf{h}(i)$ that do not have zero columns. Therefore, it still brings insight to investigation of the MIMO frequency-selective channels and the scheme discussed in this paper.

To facilitate meeting the channel matrix column-rank requirements with minimum receiver antenna number, it is possible to arrange the received sample data for each frame by stacking the data ν times as follows:

$$\mathbb{Y}^{[q]}_{M\nu\times(N+L+\nu-1)} = \mathbb{H}_{M\nu\times K(L+\nu)}\mathbb{X}^{[q]}_{K(L+\nu)\times(N+L+\nu-1)}, \tag{11}$$

$$\mathbb{Y}^{[q]} = \begin{bmatrix} \mathbf{y}_0^{[q]} & \mathbf{y}_1^{[q]} & \cdots & \cdots & \mathbf{y}_{N+L-1}^{[q]} & 0 & 0 \\ 0 & \mathbf{y}_0^{[q]} & \mathbf{y}_1^{[q]} & \cdots & \mathbf{y}_{N+L-2}^{[q]} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \mathbf{y}_0^{[q]} & \cdots & \cdots & \mathbf{y}_{N+L-1}^{[q]} \end{bmatrix},$$

$$\mathbb{H} = \begin{bmatrix} \mathbf{h}(0) & \mathbf{h}(1) & \cdots & \cdots & \mathbf{h}(L) & 0 & 0 \\ 0 & \mathbf{h}(0) & \mathbf{h}(1) & \cdots & \mathbf{h}(L-1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \mathbf{h}(0) & \cdots & \cdots & \mathbf{h}(L) \end{bmatrix}, \tag{12}$$

$$\mathbb{X}^{[q]} = \begin{bmatrix} \mathbf{x}^{[q]}(0) & \mathbf{x}^{[q]}(1) & \cdots & \cdots & \mathbf{x}^{[q]}(N-1) & 0 & 0 \\ 0 & \mathbf{x}^{[q]}(0) & \mathbf{x}^{[q]}(1) & \cdots & \mathbf{x}^{[q]}(N-2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \mathbf{x}^{[q]}(0) & \cdots & \cdots & \mathbf{x}^{[q]}(N-1) \end{bmatrix}. \tag{13}$$

The arrangement of received data in the matrix above is different from that of [16] for improving signal detection at the first and last L symbols in each transmitted frame.

If a large receive antenna number is not feasible, oversampling with larger reception bandwidth could be considered as an alternative approach to meet the necessary channel matrix rank condition. If the oversampling rate is P, $(P-1)$ times more data can be obtained and arranged as follows:

$$[\bar{\mathbf{y}}_0, \bar{\mathbf{y}}_1, \ldots, \bar{\mathbf{y}}_{N+L-1}] = \begin{bmatrix} \mathbf{y}(0) & \mathbf{y}(1) & \cdots & \mathbf{y}(N+L-1) \\ \mathbf{y}\left(\frac{1}{P}\right) & \mathbf{y}\left(1+\frac{1}{P}\right) & \cdots & \mathbf{y}\left(N+L-1+\frac{1}{P}\right) \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{y}\left(\frac{P-1}{P}\right) & \mathbf{y}\left(1+\frac{P-1}{P}\right) & \cdots & \mathbf{y}\left(N+L-1+\frac{P-1}{P}\right) \end{bmatrix}, \tag{14}$$

where the index $i + j/P$ stands for the jth sample in the ith symbol duration. Therefore, with a MIMO channel characterized by

$$[\bar{\mathbf{h}}(0), \bar{\mathbf{h}}(1), \ldots, \bar{\mathbf{h}}(L)] = \begin{bmatrix} \mathbf{h}(0) & \mathbf{h}(1) & \cdots & \mathbf{h}(L) \\ \mathbf{h}\left(\frac{1}{P}\right) & \mathbf{h}\left(1+\frac{1}{P}\right) & \cdots & \mathbf{h}\left(L+\frac{1}{P}\right) \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{h}\left(\frac{P-1}{P}\right) & \mathbf{h}\left(1+\frac{P-1}{P}\right) & \cdots & \mathbf{h}\left(L+\frac{P-1}{P}\right) \end{bmatrix}, \tag{15}$$

provided that the effects of transmission shaping filtering and reception filtering are encompassed into the channel $[\bar{\mathbf{h}}(0), \bar{\mathbf{h}}(1), \ldots, \bar{\mathbf{h}}(L)]$, the input-output relation in the oversampling case becomes

$$\bar{\mathbb{Y}}_{PMv\times(N+L+v-1)} = \bar{\mathbb{H}}_{PMv\times K(L+v)}\mathbb{X}_{K(L+v)\times(N+L+v-1)}, \quad (16)$$

where

$$\bar{\mathbb{Y}} = \begin{bmatrix} \bar{\mathbf{y}}_0^{[q]} & \bar{\mathbf{y}}_1^{[q]} & \cdots & \cdots & \bar{\mathbf{y}}_{N+L-1}^{[q]} & 0 & 0 \\ 0 & \bar{\mathbf{y}}_0^{[q]} & \bar{\mathbf{y}}_1^{[q]} & \cdots & \bar{\mathbf{y}}_{N+L-2}^{[q]} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \bar{\mathbf{y}}_0^{[q]} & \cdots & \cdots & \bar{\mathbf{y}}_{N+L-1}^{[q]} \end{bmatrix},$$

$$\bar{\mathbb{H}} = \begin{bmatrix} \bar{\mathbf{h}}(0) & \bar{\mathbf{h}}(1) & \cdots & \cdots & \bar{\mathbf{h}}(L) & 0 & 0 \\ 0 & \bar{\mathbf{h}}(0) & \bar{\mathbf{h}}(1) & \cdots & \bar{\mathbf{h}}(L-1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \bar{\mathbf{h}}(0) & \cdots & \cdots & \bar{\mathbf{h}}(L) \end{bmatrix}, \quad (17)$$

and $\mathbb{X}_{K(L+v)\times(N+L+v-1)}$ is the same as the one in (13).

In the oversampling case, it is possible to meet the full-rank requirement with a receiver antenna number smaller than that of transmitter antennas at the cost of oversampling complexity and wider reception bandwidth. The latter factor also causes degradation in signal-to-noise ratio (SNR) to a certain extent.

4. ESTIMATION OF THE TRANSMITTED SIGNALS FROM RECEIVED DATA OVER RICH MULTIPATH CHANNELS IN THE PRESENCE OF NOISE

The ST subchannels can be of different lengths and the signals are usually contaminated by the noise. In the presence of noise, $\mathcal{S}_X$ may not necessarily be the subspace intersection of $\Xi^{(p)}$ discussed in Section 3. However, it is still possible to search for independent vectors whose linear combinations can approximate row vectors in $\mathcal{S}_X$ in a similar fashion. We propose the following algorithm for determining a spanning vector set from received signals to approximate the transmitted signal vectors. This scheme is verified to provide a robust performance through simulations, which is described in the next section.

4.1. The basis estimation and approximation of transmitted signals

In the description of the receiver algorithm, the following notation is adopted:

(a) $[\mathbf{A}; \mathbf{B}]$ stands for a matrix formed by stacking matrices $\mathbf{A}$ and $\mathbf{B}$;

(b) L is the maximum length of the MIMO subchannels and is assumed to be known to the receiver;

(c) $[n_\eta, \mathbf{q}] = \bigcap_{i=1}^{i_{\max}} \{\Xi^{(i)}\}|_\eta$ denotes the following computation routine:

(1) calculate SVD: $\mathbf{U}\boldsymbol{\Sigma}\mathbf{Q} = \mathrm{SVD}([\Xi^{(1)}; \Xi^{(2)}; \ldots; \Xi^{(i_{\max})}])$, where $\mathbf{U}$, $\boldsymbol{\Sigma}$, and $\mathbf{Q}$ are the resulting matrices of the SVD computation;

(2) $\mathbf{q} = \mathbf{Q}_{[1:n_\eta,:]}$, where n_η is the number of singular vectors whose corresponding singular values are not less than η. $\mathbf{Q}_{[a:b,:]}$ denotes a matrix consisting of the rows from ath to bth of matrix $\mathbf{Q}$. Parameters $[n_\eta, \mathbf{q}]$ are the computation results of this routine.

The proposed algorithm to estimate $\mathcal{S}_X$ proceeds in three steps as follows.

Algorithm Procedure.

Step a

(1) $i_{\max} = L + v$, $r = 0$;

(2) calculate $[n_\eta, \mathbf{q}] = \bigcap_{i=1}^{i_{\max}} \{\Xi^{(i)}\}|_{\eta=0.96(\lambda_{\max}-1)}$, where $\lambda_{\max}$ is the current largest singular value;

(3) $\mathbf{V}^{(r)} = \mathbf{q}$; $r + 1 \Rightarrow r$;

(4) if $n_\eta < K$, go to Step b; else go to Step c.

Step b if $i_{\max} > 1$,

(1) $i_{\max} = i_{\max} - 1$;

(2) calculate $[n_\eta, \mathbf{q}] = \bigcap_{i=1}^{i_{\max}} \{\Xi^{(i)}\}|_{\eta=0.96(\lambda_{\max}-1)}$, where $\lambda_{\max}$ is the current largest singular value;

(3) $\mathbf{V}^{(r)} = \mathbf{q}$; $r + 1 \Rightarrow r$;

(4) if $n_\eta < K$, repeat Step b, else go to Step c;

else go to Step c.

Step c

(1) Calculate $[n_\eta, \mathbf{q}] = \bigcap_{i=1}^{r} \{\mathbf{V}^{(i)}\}|_{\eta=0.96}$;

(2) $\hat{\Xi} = \mathbf{q}$.

In the above computation of the intersection of basis vectors by SVD analysis, $\lambda_{\max}$ is an important parameter because it is used to compute how many $\Xi^{(i)}$ share certain vectors as row basis vector. Computing and applying $\lambda_{\max}$ at each step instead of setting a constant value makes the algorithm adaptive to different channel-rank situations.

In the presence of noise and channel-rank deficiency, the above basis-vector searching algorithm may get more vectors than the desired basis vectors as computation results. However, this does not prevent approximating the transmitted signals. In this paper, the signaling property of multiple constant modulus (MCM) is taken advantage of to properly weight the estimated basis in approximating the original signal vectors. The "closer" vectors to the original signal vector basis are sorted out by their dominant weights obtained from the MCM constraint.

Specifically, once the matrix $\hat{\Xi}$ is obtained, the transmitted signal matrix $\mathbf{X}_{K\times N}$ can be approximated by exploiting the MCM property. Similarly as in (9), the relation between $\hat{\mathbf{X}}_{K\times N}$ and $\hat{\Xi}_{S\times N}$ can be expressed as follows:

$$\hat{\mathbf{X}}_{K\times N} = \widehat{\mathbf{W}}_{K\times S}\hat{\Xi}_{S\times N}, \quad (18)$$

where (i) $\hat{\Xi}$ stands for a matrix whose row vectors are the estimated bases and (ii) $\hat{\mathbf{X}}$ represents the estimate of signal frame after deconvolution. The number of row vectors in $\hat{\Xi}$ may be greater than the number of the signal vector basis due to the noise effects. Hence, the matrix $\widehat{\mathbf{W}}$ is not necessarily a square matrix as $\mathbf{W}$ in (9).

The noise components have direct influences on the algorithm in two aspects: (i) the noise degrades the estimation

accuracy of the Ξ and $\mathbf{W}$; (ii) the random noise makes the processing time of each estimation vary from frame to frame. The sensitivity of the algorithm to the noise was examined by the simulations elaborated in Section 5.

The weight matrix $\widehat{\mathbf{W}}$ is calculated using the alternating projection iterations algorithm presented in the next section.

4.2. Signal property projection

DSTM employs PSK signaling so that transmitted signals have MCM characteristics. Therefore, an alternating projection method from [18] is adopted here to calculate $\widehat{\mathbf{W}}$ in the following procedure.

Algorithm Procedure. For $j = 0, 1, \ldots, n$,

(1) $\hat{\mathbf{X}}_{K\times N}^{(j)} = \widehat{\mathbf{W}}_{K\times S}^{(j)} \hat{\Xi}_{S\times N}$,
(2) $\tilde{\tilde{\mathbf{X}}}_{K\times N}^{(j)} = \text{Proc}\_G\_S\{\hat{\mathbf{X}}^{(j)}\}$,
(3) $\bar{\mathbf{X}}^{(j)} = \boldsymbol{\lambda}^{(j)} \hat{\mathbf{X}}^{(j)} + (\mathbf{I} - \boldsymbol{\lambda}^{(j)}) \tilde{\tilde{\mathbf{X}}}^{(j)}$,
(4) $\bar{\mathbf{X}}^{(j+1)} = \bar{\mathbf{X}}^{(j)} \cdot / |\bar{\mathbf{X}}^{(j)}|$,
(5) $\widehat{\mathbf{W}}_{K\times S}^{(j+1)} = \bar{\mathbf{X}}_{K\times N}^{(j+1)} \hat{\Xi}_{S\times N}^{\dagger}$,

where Proc_G_S means the Gram-Schmidt orthogonalization procedure, and $\boldsymbol{\lambda}^{(j)}$ is a diagonal relaxation matrix. The initial matrix $\widehat{\mathbf{W}}^{(0)}$ could be either determined by pilot signals or choosing randomly a full-column-rank matrix. As mentioned in [18], the Gram-Schmidt orthogonalization procedure is applied here to prevent the algorithm from being biased to certain signals of strong power. The iteration stops when $\widehat{\mathbf{W}}^{(j)}$ reaches a stable state, that is, norm $(\widehat{\mathbf{W}}^{(j+1)} - \widehat{\mathbf{W}}^{(j)}) \le \varepsilon$, where ε is a small constant.

4.3. Signal detection

After the $\widehat{\mathbf{W}}$ is estimated by the above procedure, the transmitted signal could be approximated as in (18). The relation between the original coded signal frame $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_c]$ and the estimate $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \hat{\mathbf{x}}_3, \ldots, \hat{\mathbf{x}}_c]$ can be modeled as

$$\hat{\mathbf{X}} = \mathbf{A}\mathbf{X} + \mathbf{N}, \tag{19}$$

$$\hat{\mathbf{x}}_i = \mathbf{A}\mathbf{x}_i + \mathbf{n}_i, \quad i = 1, 2, \ldots, c, \tag{20}$$

where $\mathbf{A}$ is an admissible matrix and $\mathbf{x}_i$ is an ST group code matrix. Noise elements are assumed to have independent and identical circularly symmetric complex Gaussian distribution $\mathcal{CN}(0, \delta^2)$.

Definition 1 (see [18]). If $\alpha_k \in \{\alpha \mid |\alpha_k| = 1,\ k = 1, \ldots, d\} \subset \mathcal{C}$ and $\mathbf{P}$ is a permutation matrix, the matrix $\mathbf{A} = (\text{diag}(\alpha_1, \alpha_2, \ldots, \alpha_d)\mathbf{P})$ is an admissible transformation matrix.

The ambiguity between $\mathbf{X}$ and its estimate $\hat{\mathbf{X}}$, represented by $\mathbf{A}$, exists because the MCM signal property constraint used in estimating $\widehat{\mathbf{W}}$ does not contain any phase information. From equations

$$\begin{aligned} \hat{\mathbf{x}}_i &= \mathbf{A}\mathbf{x}_i + \mathbf{n}_i, \qquad \hat{\mathbf{x}}_{i+1} = \mathbf{A}\mathbf{x}_{i+1} + \mathbf{n}_{i+1}, \\ \mathbf{x}_{i+1} &= \mathbf{x}_i \mathbf{G}_{[m]}, \end{aligned} \tag{21}$$

we obtain the following relations:

$$\hat{\mathbf{x}}_{i+1} = \hat{\mathbf{x}}_i \mathbf{G}_{[m]} + \ddot{\mathbf{n}}_{i+1}, \tag{22}$$

where

$$\ddot{\mathbf{n}}_{i+1} = \mathbf{n}_{i+1} - \mathbf{n}_i \mathbf{G}_{[m]}. \tag{23}$$

The dependence between $\hat{\mathbf{x}}_{i+1}$ and $\hat{\mathbf{x}}_i$ indicates a differential relation with the multiplicative matrix $\mathbf{G}_{[m]}$. It can be observed that the ambiguity matrix $\mathbf{A}$ between $\hat{\mathbf{x}}_i$ and $\mathbf{x}_i$ is removed by the differential signaling and differential detection.

Hence, the detection of $\mathbf{G}^{[i]}$ can be carried out using a least square error detector:

$$\hat{\mathbf{G}}^{[i+1]} = \arg\min_{\mathbf{G}_{[r]}} \left|\left| \hat{\mathbf{x}}_{i+1} - \hat{\mathbf{x}}_i \mathbf{G}_{[r]} \right|\right|, \tag{24}$$

where, for the $\mathbf{G}$ matrices, the matrix subscript r is an ST codeword alphabet index, and the superscript i is a time index of the ST codeword.

From (24), we get

$$\begin{aligned} \hat{\mathbf{G}}^{[i+1]} &= \arg\min_{\mathbf{G}_{[r]}} \mathbf{Tr}\left\{ \left(\hat{\mathbf{x}}_{i+1} - \hat{\mathbf{x}}_i \mathbf{G}_{[r]}\right)^H \left(\hat{\mathbf{x}}_{i+1} - \hat{\mathbf{x}}_i \mathbf{G}_{[r]}\right) \right\} \\ &= \arg\min_{\mathbf{G}_{[r]}} \mathbf{Tr}\Big\{ \hat{\mathbf{x}}_{i+1}^H \left(\hat{\mathbf{x}}_{i+1}\right) - \left(\hat{\mathbf{x}}_i \mathbf{G}_{[r]}\right)^H \hat{\mathbf{x}}_{i+1} \\ &\qquad - \left(\hat{\mathbf{x}}_{i+1}\right)^H \left(\hat{\mathbf{x}}_i \mathbf{G}_{[r]}\right) + \left(\hat{\mathbf{x}}_i \mathbf{G}_{[r]}\right)^H \left(\hat{\mathbf{x}}_i \mathbf{G}_{[r]}\right) \Big\}. \end{aligned} \tag{25}$$

Because $\mathbf{Tr}\{(\hat{\mathbf{x}}_i \mathbf{G}_{[r]})^H (\hat{\mathbf{x}}_i \mathbf{G}_{[r]})\}$ is a constant for different $\mathbf{G}_{[r]}$, the detector for DSTM's differential signaling becomes

$$\hat{\mathbf{G}}^{[i+1]} = \arg\max_{\mathbf{G}_{[r]}} \Re\left\{ \mathbf{Tr}\left\{ \left(\hat{\mathbf{x}}_i \mathbf{G}_{[r]}\right)^H \left(\hat{\mathbf{x}}_{i+1}\right) \right\} \right\}. \tag{26}$$

Through the approximation of the signals with the estimated basis as in (18), the intersymbol interference (ISI) of the signal is mitigated. Hence, in the procedure proposed in this paper for MIMO frequency-selective channels, the final detection stage embodied through (26) is similar to that for DSTM signaling over frequency-flat fading MIMO channels as represented in (3). In the comparison of (26) and (3), the following property is useful: for square matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{Tr}\{\mathbf{AB}\} = \mathbf{Tr}\{\mathbf{BA}\}$.

4.4. Summary of the receiver algorithm

The complete receiver algorithm proposed for DSTM signaling over the finite-impulse-response (FIR) rich multipath channels proceeds on a frame-by-frame basis according to the following four steps:

(1) estimate the direct input signal subspace basis and signal approximations according to the method in Section 3.1;
(2) calculate $\widehat{\mathbf{W}}$ by iterating the alternating projections exploiting MCM using the algorithm presented in Section 3.2;
(3) determine $\hat{\mathbf{X}}$ by $\hat{\mathbf{X}} = \widehat{\mathbf{W}}\hat{\Xi}$;

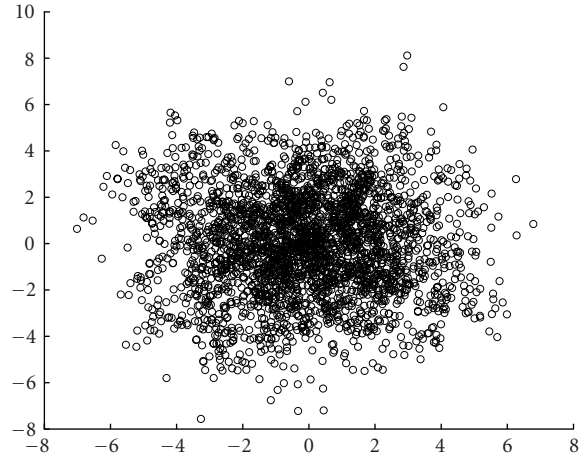


Figure 3: Received signal constellation diagram ($L = 7$, $M = 6$, $K = 4$, $P = 1$, $N = 256$, SNR/bit/antenna = 18.5 dB).

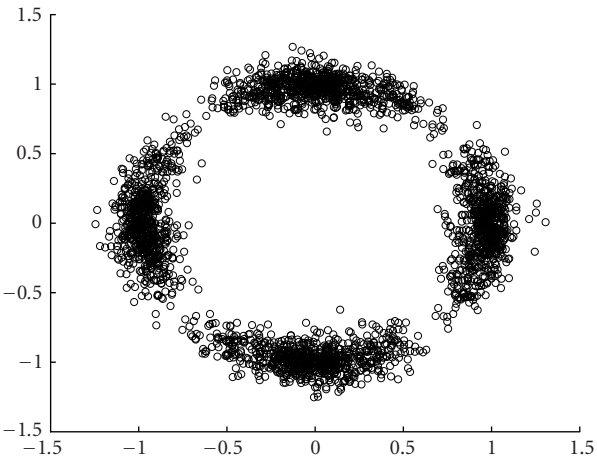


Figure 4: Signal constellation diagram after equalization ($L = 7$, $M = 6$, $K = 4$, $P = 1$, $N = 256$, SNR/bit/antenna = 18.5 dB).

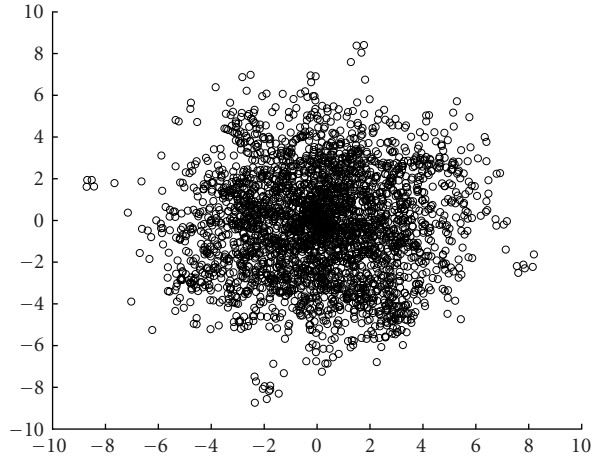


Figure 5: Received signal constellation diagram ($L = 7$, $M = 6$, $K = 4$, $P = 1$, $N = 256$, SNR/bit/antenna = 19.3 dB).

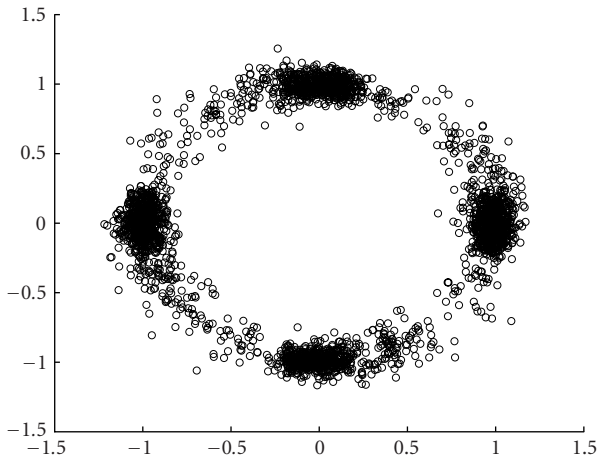


Figure 6: Signal constellation diagram after equalization ($L = 7$, $M = 6$, $K = 4$, $P = 1$, $N = 256$, SNR/bit/antenna = 19.3 dB).

(4) perform signal detection according to detection criteria (26) as described in Section 4.3.

Provided that the maximum delay spread is less than T_G, the block Toeplitz signal structure and data processing procedures in Sections 3.1, 3.2, and 4.3 enable the algebraic data recovery without channel knowledge and channel estimation. The procedures in Sections 3.1 and 3.2 mitigate frequency-selective effects in rich multipath environment, and the differential detection of ST symbols described in Section 4.3 removes the ambiguity of transformation $\mathbf{A}$ in (19).

Regarding the proposed algorithm, it should be noted that the receiver algorithm proposed in this paper exploits both block Toeplitz structure of the received signals and the MCM property of M-ary PSK signaling. It is not directly applicable to the schemes with a signaling without constant envelope. When employing other signalings that do not have the MCM property, the part of the receiver algorithm described in Section 3.2 for estimating $\mathbf{W}$ must be modified.

5. PERFORMANCE SIMULATIONS

With different parameter settings of the transceiver and the channels, simulations of the new receiver algorithm were conducted to verify the bit error rate (BER) performance over Rayleigh FIR fading channels in the presence of AWGN. Figures 3, 4, 5, 6, 7, and 8 illustrate the signal constellation before and after the equalization for different values of SNR per antenna. From Figures 4, 6, and 8, it is evident that enforcing the MCM property in our algorithm causes the signal constellation after equalization to have a circular appearance.

The representative BER simulation results with the parameters $K = 4$, $M = 5, 6$, $N = 256$, and $P = 1$ are illustrated in Figures 9, 10, and 11 for $L = 3, 5, 7$, respectively.

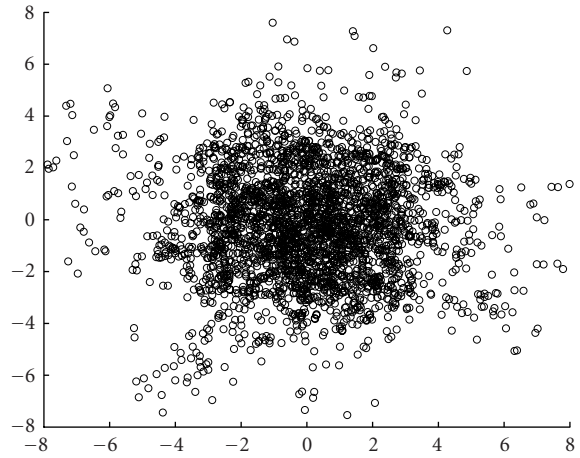


Figure 7: Received signal constellation diagram ($L = 7$, $M = 6$, $K = 4$, $P = 1$, $N = 256$, SNR/bit/antenna = 21.4 dB).

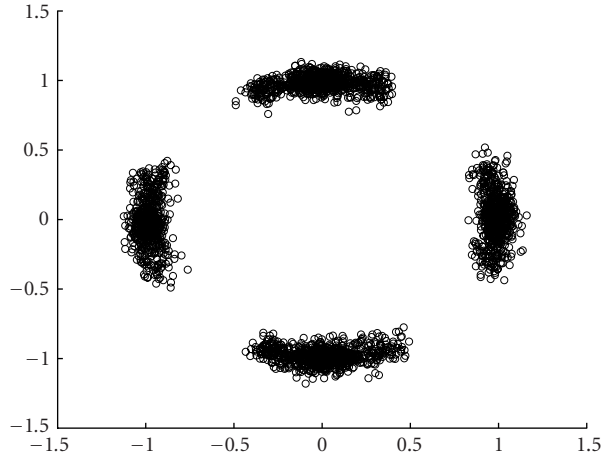


Figure 8: Signal constellation diagram after equalization ($L = 7$, $M = 6$, $K = 4$, $P = 1$, $N = 256$, SNR/bit/antenna = 21.4 dB).

The multiple channels were simulated to be the FIR Rayleigh fading channels.

The simulations were carried out by employing a $(\mathcal{M}; k_1, \ldots, k_4) = (4; 1, 1, 1, 1)$ cyclic group code [9] and Q-PSK signaling. The results were statistically averaged over all possible cases of random path delays of the subchannels, random ST channel states, random bitstreams, and random AWGN. The SNR values in Figures 9, 10, and 11 are the spatially and temporally averaged SNR per antenna over all the frames received.

For comparison purposes, the performance of DSTM signaling with the previous receiver's algorithm was simulated with the same fading channels. From the figures, it can be observed that the receiver (without equalization) derived under the assumption of the frequency-flat fading channels fail in the frequency-selective fading channels considered in the simulations (curves are labeled as "without equalization" in the figures). On the other hand, the proposed algorithm

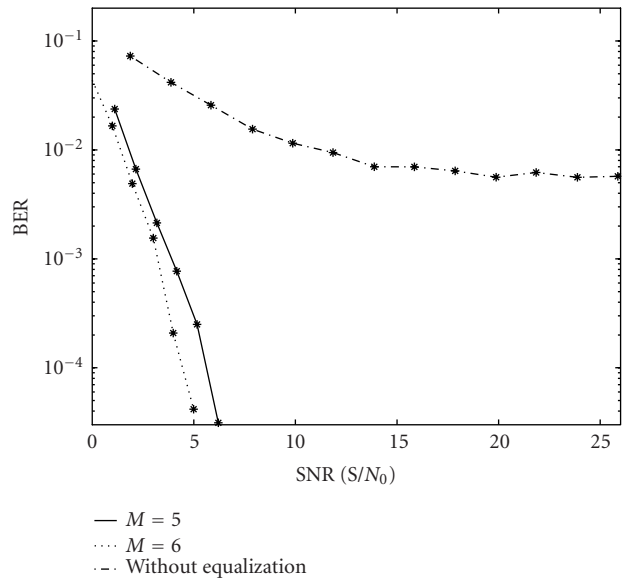


Figure 9: System BER performance in time-dispersive fading channel ($L = 3$, $K = 4$, $P = 1$, $N = 256$).

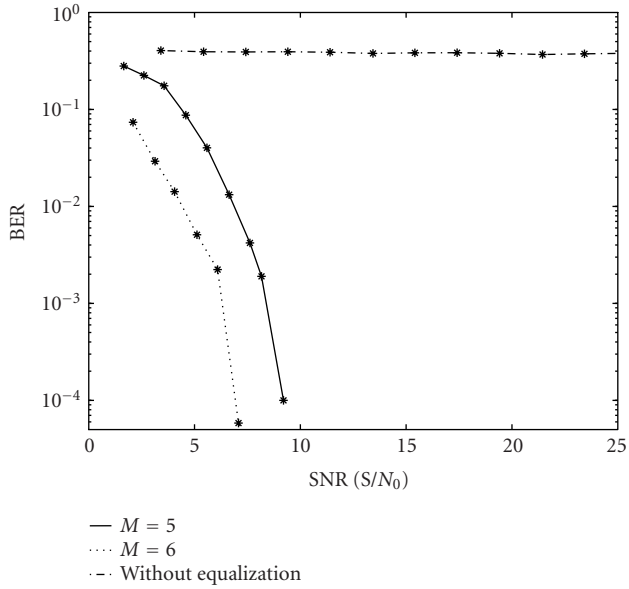


Figure 10: System BER performance in time-dispersive fading channel ($L = 5$, $K = 4$, $P = 1$, $N = 256$).

(with equalization) maintains a robust performance in rich multipath quasistatic FIR fading channels.

When the channel length is increased, it is more difficult to remove the ISI effects. This is evident by comparing the performance curves in Figures 9, 10, and 11, where $L = 3, 5$, and 7, respectively. From these figures, we can observe that in order to obtain the same performance of BER at 10^{-3} using the same transceiver setup, the SNR has to be increased from 4 dB to 7 dB and 14.1 dB for $K = 4$, $M = 5$. Additionally,

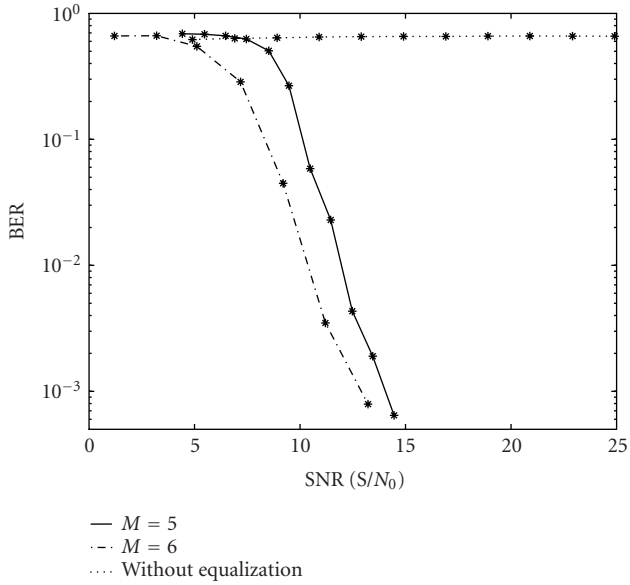


Figure 11: System BER performance in time-dispersive fading channel ($L = 7, K = 4, P = 1, N = 256$).

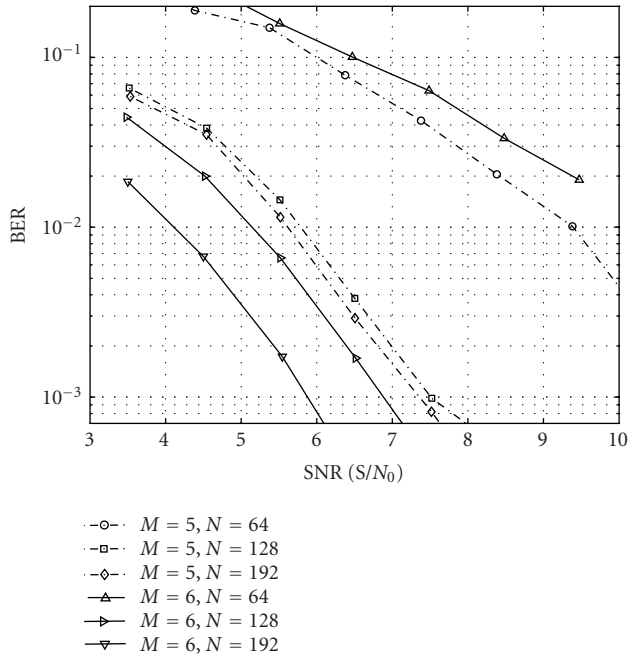


Figure 12: System BER performance in time-dispersive fading channel ($L = 5, K = 4, M = 5, 6, P = 1$).

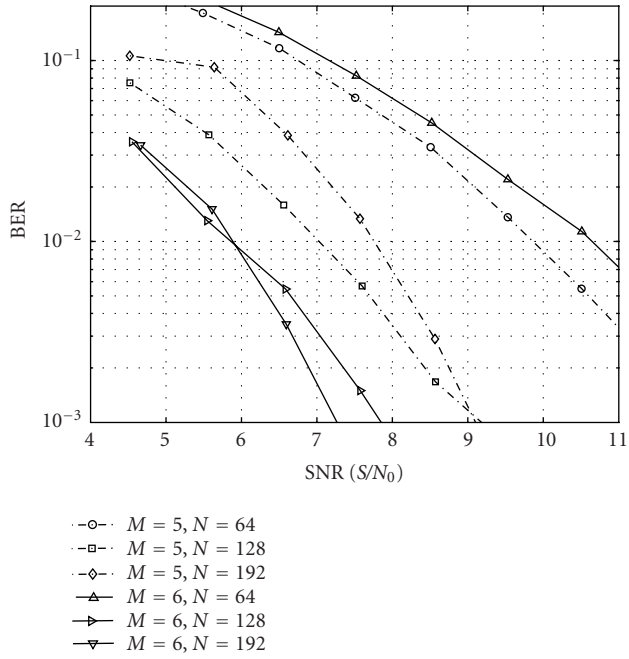


Figure 13: System BER performance in time-dispersive fading channel ($L = 6, K = 4, M = 5, 6, P = 1$).

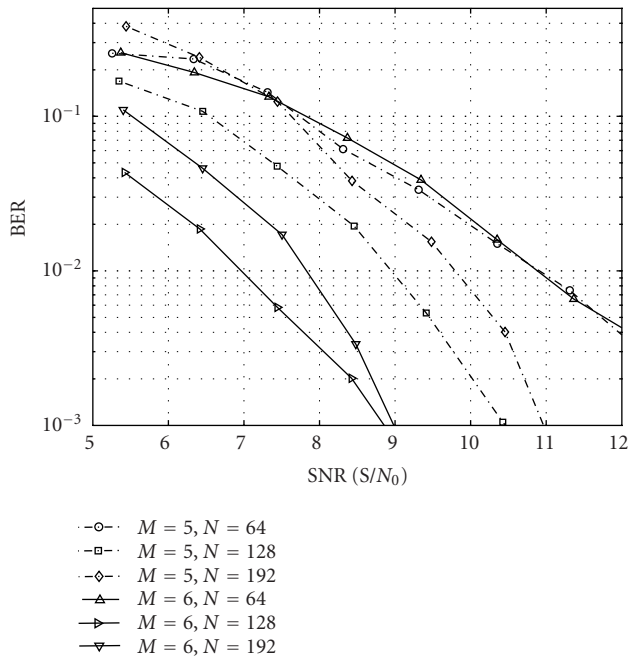


Figure 14: System BER performance in time-dispersive fading channel ($L = 7, K = 4, M = 5, 6, P = 1$).

the power savings by increasing the receiver antenna number depends on the BER operating point of the system.

Similarly, for different $N = 64, 128, 192$, the simulation results with the parameters $K = 4, M = 5, 6$, and $P = 1$ are illustrated in Figures 12, 13, and 14 for $L = 5, 6, 7$, respectively. From these figures, we could observe that the choices of N exhibit a considerable influence on the system performance. To some extent, for short channel length cases, a relatively larger N within a certain range facilitates higher performance. The improvements are achieved at the expense of the increased computational complexity. But, for the cases of long channel lengths, this trend does not exist.

6. CONCLUSIONS

This paper proposes a blind ST receiver algorithm for DSTM transmissions over quasistatic FIR fading channels. The algorithm is applicable in the transmission scenarios with different numbers of antennas at both the transmitter and receiver sides. Simulation results demonstrate its robust performance over unknown rich multipath FIR fading channels. With a proper design of the transceiver parameters and the frame guard time T_G in the new scheme, the ST symbol detection error drops significantly when SNR passes certain thresholds despite the delay spread of the channels.

Particularly, the new detection algorithm does not rely on the channel estimation. Secondly, the proposed receiver is not subjected to the channel changes provided the channel is invariant within one frame time slot. Furthermore, in contrast to the methods based on the statistics of a large amount of signal samples, the proposed scheme is capable of operating when a relatively small number of received data samples are available.

APPENDIX

PROOF OF PROPOSITION 1

Proof. Let $\mathcal{S}_{\mathcal{Y}^{(p)}}$ denote the row span of $\Xi^{(p)}$, $p = 0, 1, 2, \ldots, L$. If $\mathbf{H}$ is of a full-column rank, from (5), it could be concluded that

For $p = 0$,

$$\mathcal{S}_{\mathcal{Y}^{(0)}} = \text{row\_span}\begin{Bmatrix} \mathbf{x}(0) & \mathbf{x}(1) & \mathbf{x}(2) & \mathbf{x}(3) & \cdots & \cdots & \mathbf{x}(N-1) \\ 0 & \mathbf{x}(0) & \mathbf{x}(1) & \mathbf{x}(2) & \cdots & \cdots & \mathbf{x}(N-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \mathbf{x}(0) & \mathbf{x}(1) & \cdots & \mathbf{x}(N-L-1) \end{Bmatrix}; \tag{A.1}$$

For $p = 1$,

$$\mathcal{S}_{\mathcal{Y}^{(1)}} = \text{row\_span}\begin{Bmatrix} \mathbf{x}(1) & \mathbf{x}(2) & \cdots & \cdots & \mathbf{x}(N-1) & 0 \\ \mathbf{x}(0) & \mathbf{x}(1) & \cdots & \cdots & \cdots & \mathbf{x}(N-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \mathbf{x}(0) & \mathbf{x}(1) & \cdots & \mathbf{x}(N-L) \end{Bmatrix}; \tag{A.2}$$

$$\cdots$$

For $p = L$,

$$\mathcal{S}_{\mathcal{Y}^{(L)}} = \text{row\_span} \times \begin{Bmatrix} \mathbf{x}(L) & \mathbf{x}(L+1) & \cdots & \mathbf{x}(N-1) & 0 & 0 \\ \mathbf{x}(L-1) & \mathbf{x}(L-2) & \cdots & \cdots & \mathbf{x}(N-1) & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{x}(0) & \mathbf{x}(1) & \mathbf{x}(2) & \cdots & \cdots & \mathbf{x}(N-1) \end{Bmatrix}. \tag{A.3}$$

By observing the above relationship, it is evident that $\mathcal{S}_X \subset \mathcal{S}_{\mathcal{Y}^{(i)}}$, respectively, for $i = 0, 1, 2, \ldots, L$. Therefore, according to set theory,

$$\mathcal{S}_X \subset \left\{ \bigcap_{i=0}^{L} \mathcal{S}_{\mathcal{Y}^{(i)}} \right\}. \tag{A.4}$$

Consider $\mathcal{S}_{\mathcal{Y}^{(1)}} \bigcap \mathcal{S}_{\mathcal{Y}^{(2)}}$, which is equivalent to the intersection of row subspaces of

$$\begin{Bmatrix} \mathbf{x}(0) & \mathbf{x}(1) & \mathbf{x}(2) & \cdots & \cdots & \cdots & \cdots & \mathbf{x}(N-1) \\ 0 & \mathbf{x}(0) & \mathbf{x}(1) & \mathbf{x}(2) & \cdots & \cdots & \cdots & \mathbf{x}(N-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \mathbf{x}(0) & \mathbf{x}(1) & \cdots & \cdots & \mathbf{x}(N-L-1) \end{Bmatrix},$$

$$\begin{Bmatrix} \mathbf{x}(1) & \mathbf{x}(2) & \cdots & \cdots & \cdots & \cdots & \mathbf{x}(N-1) & 0 \\ \mathbf{x}(0) & \mathbf{x}(1) & \mathbf{x}(2) & \cdots & \cdots & \cdots & \cdots & \mathbf{x}(N-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \mathbf{x}(0) & \mathbf{x}(1) & \cdots & \cdots & \cdots & \mathbf{x}(N-L) \end{Bmatrix}. \tag{A.5}$$

If frame length N is sufficiently large, the rows of $\mathcal{X}^{[q]}$ are linear independent with probability of 1. Observing the block Toeplitz structure of the above matrices, the row rank of the intersection is $(K(L+1)-K)$. Therefore, the number of basis vectors of $\mathcal{S}_{\mathcal{Y}^{(1)}} \bigcap \mathcal{S}_{\mathcal{Y}^{(2)}}$ is also $(K(L+1)-K)$.

Following the similar verification procedure, it could be observed that the number of row basis vectors of $\mathcal{S}_{\mathcal{Y}^{(1)}} \bigcap \mathcal{S}_{\mathcal{Y}^{(2)}} \bigcap \mathcal{S}_{\mathcal{Y}^{(3)}}$ is $(K(L+1)-2K)$.

Moreover, the number of basis vectors of $\{\bigcap_{i=0}^{L} \mathcal{S}_{\mathcal{Y}^{(i)}}\}$ is K, which is equal to the number of row basis vectors for $\mathcal{S}_X$. Hence, from (A.4), it is concluded that

$$\mathcal{S}_X = \left\{ \bigcap_{i=0}^{L} \mathcal{S}_{\mathcal{Y}^{(i)}} \right\}. \tag{A.6}$$

□

ACKNOWLEDGMENT

Part of the work described in this paper was presented during the Fourth IEEE International Workshop on Mobile and Wireless Communications Network, September 2002, Stockholm, Sweden.

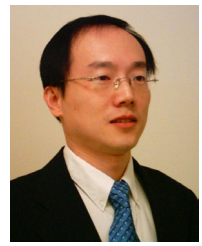
REFERENCES

[1] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Magazine*, vol. 14, no. 6, pp. 49–83, 1997.

[2] C. Schlegel and Z. Bagley, *Efficient processing for high-capacity MIMO channels*, preprint, http://www.ee.ualberta.ca/%7eschlegel/publications.html.

[3] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[4] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. Journal*, vol. 1, no. 2, pp. 41–59, 1996.
[5] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. URSI International Symposium on Signals, Systems, and Electronics*, pp. 295–300, Pisa, Italy, September–October 1998.
[6] T. H. Liew and L. Hanzo, "Space-time codes and concatenated channel codes for wireless communications," *Proceedings of the IEEE*, vol. 90, no. 2, pp. 187–219, 2002.
[7] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.
[8] B. L. Hughes, "Differential space-time modulation," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2567–2578, 2000.
[9] B. L. Hughes, "Space-time group codes," in *Proc. 34th Asilomar Conference on Signals, Systems, and Computers*, vol. 1, pp. 699–704, Pacific Grove, Calif, USA, October–November 2000.
[10] J. Liu, J. Li, H. Li, and E. G. Larsson, "Differential space-code modulation for interference suppression," *IEEE Trans. Signal Processing*, vol. 49, no. 8, pp. 1786–1795, 2001.
[11] J. Liu, E. G. Larsson, J. Li, and H. Li, "High-rate differential space-code modulation for interference suppression," in *Proc. 3rd IEEE Signal Processing Workshops on Signal Processing Advances in Wireless Communications*, pp. 283–286, Taoyuan, Taiwan, March 2001.
[12] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Communications*, vol. 48, no. 12, pp. 2041–2052, 2000.
[13] N. D. Sidiropoulos and R. S. Budampati, "Khatri-Rao space-time codes," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2396–2407, 2002.
[14] Q. Ma, C. Tepedelenlioglu, and Z. Liu, "Differential space-time-frequency coded OFDM with maximum diversity," in *Proc. 37th Annual Conference on Information Sciences and Systems*, Baltimore, Md, USA, March 2003.
[15] B. L. Hughes, "Further results on differential space-time modulation," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, pp. 163–167, Cambridge, Mass, USA, March 2000.
[16] A.-J. van der Veen, S. Talwar, and A. Paulraj, "A subspace approach to blind space-time signal processing for wireless communication systems," *IEEE Trans. Signal Processing*, vol. 45, no. 1, pp. 173–190, 1997.
[17] D. Gesbert, H. Bolcskei, D. A. Gore, and A. J. Paulraj, "MIMO wireless channels: capacity and performance prediction," in *Proc. IEEE Global Communications Conference*, vol. 2, pp. 1083–1088, San Francisco, Calif, USA, November 2000.
[18] Y. Wang, Y. C. Pati, Y. M. Cho, A. Paulraj, and T. Kailath, "A matrix factorization approach to signal copy of constant modulus signals arriving at an antenna array," in *Proc. 28th Annual Conference on Information Sciences and Systems*, pp. 178–183, Princeton, NJ, USA, March 1994.



Zhan Zhang received his B.S. degree in electronic engineering from Northwestern Polytechnic University, Xi'an, China, in 1990, and the M.A.S. degree from Beijing Institute of Technology, Beijing, China, in 1996. He is now working toward the Ph.D. degree at the Department of Electrical & Computer Engineering, Dalhousie University, Halifax, Canada. His main research interests are in the areas of signal processing for digital communications, with emphases on multiuser detection, space-time channel coding and signal processing, and channel identification and equalization.

Jacek Ilow is an Associate Professor at the Department of Electrical and Computer Engineering, Dalhousie University, Halifax, Canada. He joined Dalhousie as an Assistant Professor in April 1997, and was promoted to the Associate Professor with tenure in July 2002. He received the B.Eng. degree in electronics from the Wroclaw University of Technology, Poland, in 1987 and the M.A.S. and Ph.D. degrees in electrical and computer engineering from the University of Toronto, Canada, in 1992 and 1996, respectively. From 1987 to 1988, he worked as a Researcher at the Wroclaw University of Technology in the Institute of Telecommunication and Acoustics. From December 1995 until March 1997, Dr. Ilow held an NSERC postdoctoral fellowship at the Defense Research Establishment, Ottawa, Ontario. His research interests lie in the areas of wireless networks, digital communications, and statistical signal processing.

EURASIP Journal on Applied Signal Processing 2004:5, 640–648

Improved Multiuser Detectors Employing Genetic Algorithms in a Space-Time Block Coded System

Yinggang Du
Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong
Email: ygdu@ee.cuhk.edu.hk

Kam Tai Chan
Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong
Email: ktchan@ee.cuhk.edu.hk

Received 31 December 2002; Revised 15 November 2003

Enhanced genetic algorithms (GA) applied in space-time block coded (STBC) multiuser detection (MUD) systems in Rayleigh flat-fading channels are reported in this paper. Firstly, an improved objective function, which is designed to help speed up the search for the optimal solution, is introduced. Secondly, a decorrelating detector (DD) and a minimum mean square error (MMSE) detector have been added to the GA STBC MUD receiver to create the seed chromosome in the initial population. This operation has improved the receiver performance further because some signal information has been intentionally embedded in the initial population. Simulation results show that the receiver employing the improved objective function and the DD or MMSE detector can converge faster with the same bit error rate (BER) performance than the receiver with the initial population chosen randomly. The total signal-to-noise ratio (SNR) improvement contributed by these two modifications can reach 4 dB. Hence the proposed GA receiver is a promising solution of the STBC MUD problem.

Keywords and phrases: genetic algorithms, multiuser detection, decorrelating detector, minimal mean square error detector, space-time block coding, objective function.

1. INTRODUCTION

In wireless communications, space-time block coding (STBC) with diversity gains has been widely studied in multiuser detections (MUDs) [1, 2, 3] because STBC can utilize the information in the spatial and time domains simultaneously [4]. From the coding point of view, the single-user performance of STBC has been studied in [5, 6] and some STBC code designs have been established. Generally speaking, STBC is a coding technique designed particularly for the application with multiple transmit antennas [5, 6]. It introduces temporal and spatial correlation into signals transmitted from different antennas. The signals transmitted from the different transmit antennas in an STBC transmitter can be considered and calculated as if they were originated from different virtual users so that STBC detection in the case of single user can be regarded simply as a MUD problem.

Conventional detection (CD) of multiuser signals utilizes a bank of chip-level matched filters to detect each user signal separately while treating all other user signals as interference. However, owing to multiaccess interference (MAI), this single-user detection method suffers much from the near-far effect. Consequently, the MUD technique that employs a combinatorial optimization process to exploit the information of all the users in order to detect a target user signal has been proposed to mitigate the near-far effect [7]. Among all the MUD techniques, the maximum likelihood (ML) optimal detector can achieve satisfactory bit error rate (BER) performance but the computational complexity varies exponentially with the number of users. Even with the maximum a posteriori (MAP) method described in [8], the computational complexity still varies in the order of 2^l, where "l" is the length of the codeword and is a linear function of the number of users. Therefore, the MUD technique is a nondeterministic polynomial (NP)-hard problem [9], which requires unforeseeable huge computing power in order to find a global optimum solution. For such NP-hard problems, it is necessary to search for good approximation algorithms that yield solutions close to the optimum, although they do not guarantee that a global optimum can be obtained for every instance. Such approximate algorithms are also based on the ML method but the final decision is reached from a simpler route at the expense of performance degradation. Some general approximation algorithms that can achieve a reasonable balance of system performance and computational speed have been reported: simulated annealing (SA)

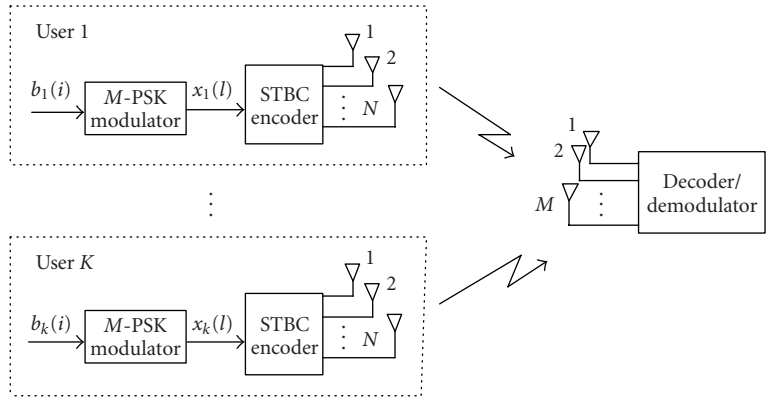


Figure 1: Schematic of an uplink STBC MUD system.

[10], tabu search (TS) [11], and genetic algorithms (GA) [9, 12, 13, 14, 15, 16]. In the TS method, the move attribute, which is a set of important parameters stored in the tabu list, must be chosen in such a way that it is neither too permissive nor too restrictive. Otherwise, the method will likely converge to a local optimum with higher probabilities. On the other hand, the GA method, which can be transformed into the SA method by gradually modifying some GA parameters, will only lead to a longer convergence time but not a higher probability of converging to a local optimum even if the GA parameters have been improperly selected. It is because of this more tolerable selection of parameters that GA has been chosen to investigate alternative MUD solutions in this paper.

The GA comes originally from the schemata theory [12]. It was inspired by the observations of the natural process of evolution of plant and animal species. These living species constantly explore new possibilities in building new living organisms as well as skillfully exploit the "knowledge" accumulated in the current living organisms to create new species that are as capable of surviving as their ancestors [9]. These remarkable characteristics of the process of creating new forms of life have caught the interest of computer science researchers and led to the creation of the GA in 1975 [12].

It has been shown in [13, 14, 15, 16] that the application of GA in MUD systems can significantly reduce the computation complexity with comparable BER performance in Rayleigh flat-fading channels. In particular, the application of GA in an STBC MUD system is proposed for the first time in [13], where all the chromosomes in the initial population are randomly chosen. However, the results shown in [13] indicate that the number of generations needed for the output to converge to a satisfactory performance is still fairly large, which makes the computation too long even though it is already much shorter than the ML method.

In this paper, a modified objective function is first introduced to shorten the GA computation. Furthermore, a decorrelating multiuser detector (DD) and a minimum mean square error (MMSE) detector [17] are also proposed to provide the seed chromosome of the initial population in the STBC MUD receiver in a wireless uplink transmission system. The DD decouples the received signal and the linear MMSE detector maximizes the receiver output signal-to-noise ratio (SNR). The seed chromosome provided by the DD or MMSE detector is perturbed randomly to generate all the chromosomes of the initial population. Hence some signal information has been embedded in the initial population. Consequently, we expect this algorithm to converge more quickly and the computation burden should be reduced. The only cost of this technique is the time consumed to obtain the seed chromosome, which is small compared with the benefit it brings. The simulation results have confirmed the validity of the proposed receiver.

Notations

Superscripts $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ denote the complex conjugate, transpose, and Hermit operation respectively; $(\cdot)^{-1}$ refers to the matrix inverse operation; $\|\cdot\|$ is for the matrix/vector Frobenius norm; and $\cdot \times \cdot$ represents the dimension of a complex matrix.

2. SYSTEM MODEL

The schematic of a multiuser STBC system is shown in Figure 1. Just to simplify the analysis, we cut the description of the outer code and the channel estimation with the assumption that the channel state information (CSI) is perfectly known and remains constant during an entire block period. In real cases, it is not difficult to include such technologies in STBC MUD systems. There are K users in the uplink channel. For each user, the bit stream is MPSK modulated and STBC encoded before being transmitted from N antennas. The size of the PSK modulation set is Q. All users are assumed to be synchronous and mutually independent. The signals from all users reach the M receive antennas through a Rayleigh flat-fading channel.

Without loss of generality, the STBC $\mathcal{G}_2$ code described in [5, 6] is selected in this study. The $\mathcal{G}_2$ code represents one category of orthogonal STBC, where the subscript "2" refers to N, the number of transmit antennas. It is straightforward to extend this study to the case of more transmit antennas at

other transmission rates, no matter whether the STBC design is orthogonal or not. $\mathcal{G}_2$ is given as

$$\mathcal{G}_2 = \begin{bmatrix} x_1^k & x_2^k \\ -x_2^{k*} & x_1^{k*} \end{bmatrix} = \begin{bmatrix} c_{1,k}^1 & c_{1,k}^2 \\ c_{2,k}^1 & c_{2,k}^2 \end{bmatrix}, \tag{1}$$

where the superscript "k" represents the kth user. x_1^k and x_2^k are the symbols to be transmitted per block (i.e., the number of symbols per block is $K_0 = 2$). The symbol $c_{t,k}^i$ (where $i = 1,\ldots,N$ and $t = 1,\ldots,P$) in the ith column for the kth user with $P = 2$ time slots per block is transmitted by the ith transmit antenna. The transmission rate of $\mathcal{G}_2$ is $R = K_0/P = 1$, which is the highest in all STBC designs. In all of the other cases, R is always less than 1.

The detected signal from all the users in the jth ($j = 1,2,\ldots,M$) receive antenna (in the base station) in the tth time slot is

$$r_t^j = \sum_{k=1}^{K} \left(\sum_{i=1}^{N} \alpha_{i,j}^k c_{t,k}^i \right) + \eta_t^j, \tag{2}$$

where $\alpha_{i,j}^k$ is the path gain between the ith transmit antenna and the jth receive antenna for the kth user and η_t^j is the noise. As usual, $\alpha_{i,j}^k$ is taken as an independent complex Gaussian random variable with zero mean and a variance of 0.5 per dimension.

For the kth user, the path gain matrix for the jth receive antenna is (the superscript k is dropped till (9) to simplify the notation)

$$\vec{h}_j = \begin{bmatrix} \alpha_{1,j} & \alpha_{2,j} & \cdots & \alpha_{N,j} \end{bmatrix}^T. \tag{3}$$

The signal symbols per block can be denoted as

$$\vec{x} = [\mathrm{Re}\,(x_1)\ \mathrm{Re}\,(x_2)\ \cdots\ \mathrm{Re}\,(x_{K_0})\ \mathrm{Im}\,(x_1)\ \mathrm{Im}\,(x_2)\ \cdots\ \mathrm{Im}\,(x_{K_0})]^T. \tag{4}$$

The signal in the jth receive antenna is

$$\underline{r}^j = \begin{bmatrix} r_1^j & r_2^j & \cdots & r_P^j \end{bmatrix}^T \tag{5}$$

and the noise in the jth receive antenna is

$$\eta^j = \begin{bmatrix} \eta_1^j & \eta_2^j & \cdots & \eta_P^j \end{bmatrix}^T. \tag{6}$$

Using STBC encoding, the detected signal in the jth antenna can be obtained from the following general matrix transformation which is applicable to any STBC design:

$$\underline{r}^j = \mathcal{G}_2 \vec{h}_j + \eta^j = H^j \vec{x} + \eta^j, \tag{7}$$

where H^j for the $\mathcal{G}_2$ design is defined as

$$H^j = \begin{bmatrix} \alpha_{1,j} & \alpha_{2,j} & \alpha_{1,j} & \alpha_{2,j} \\ \alpha_{2,j} & -\alpha_{1,j} & -\alpha_{2,j} & \alpha_{1,j} \end{bmatrix}. \tag{8}$$

The purpose of introducing the redundant vector $\vec{x}$ in (4) with the conjugation of the signal symbols is to enable it to treat the generalized STBC design. When a nonredundant signal vector $[x_1, x_2, \ldots, x_{K0}]^T$ is used, the cost of conjugating half of the received signals and CSI has to be taken into account as well [3]. For example, instead of the transformation in (7), the received signal for the Alamouti design $\mathcal{G}_2$ [6] will be represented as [3] follows:

$$\begin{bmatrix} r_1^j \\ r_2^{j*} \end{bmatrix} = \begin{bmatrix} \alpha_{1,j} & \alpha_{2,j} \\ \alpha_{2,j}^* & -\alpha_{1,j}^* \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \eta^j. \tag{9}$$

However, it is impossible to apply (9) for the following STBC design:

$$S = \begin{bmatrix} x_1 & x_2 & x_3 & 0 \\ -x_2^* & x_1^* & 0 & -x_3 \\ x_3^* & 0 & -x_1^* & -x_2 \\ 0 & -x_3^* & x_2^* & -x_1^* \end{bmatrix}. \tag{10}$$

The $\mathcal{H}_{3,4}$ design in [5] is similar to S. In this situation, the redundant form given by (4) and the transformation shown in (7) have to be adopted, which then eliminates the need to conjugate the received signals and the CSI [18]. Hence the combined received signal of the M receive antennas can be written as

$$\underset{MP\times 1}{\mathbf{r}} = \begin{bmatrix} \underline{r}^{1^T} & \underline{r}^{2^T} & \cdots & \underline{r}^{M^T} \end{bmatrix}^T \tag{11}$$

where $MP \times 1$ and similar notations in the following equations refer to the dimensions of the matrices. Therefore, including all of the K users, the signal in the jth receive antenna should be

$$\underline{r}^j = \sum_{k=1}^{K} H_k^j \vec{x}^k + \eta^j, \quad j = 1,2,\ldots,M. \tag{12}$$

Hence the received signal vector is

$$\underset{MP\times 1}{\mathbf{r}} = \begin{bmatrix} \underline{r}^1 \\ \underline{r}^2 \\ \vdots \\ \underline{r}^M \end{bmatrix} = \underbrace{\begin{bmatrix} H_1 & H_2 & \cdots & H_K \end{bmatrix}}_{\mathbf{H}} \times \underbrace{\begin{bmatrix} \vec{x}^1 \\ \vec{x}^2 \\ \vdots \\ \vec{x}^K \end{bmatrix}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} \eta^1 \\ \eta^2 \\ \vdots \\ \eta^M \end{bmatrix}}_{\boldsymbol{\eta}} \tag{13}$$

$MP \times 1 \qquad MP \times 2K_0K \qquad 2K_0K \times 1 \quad MP \times 1$

where

$$H_k = \begin{bmatrix} H_k^1 & H_k^2 & \cdots & H_k^M \end{bmatrix}^T \ MP \times 2K_0. \tag{14}$$

In matrix form, the received signal is

$$\mathbf{r} = \mathbf{H}\mathbf{x} + \boldsymbol{\eta}. \tag{15}$$

In the conventional receiver, each user is separately detected [14]. Therefore, the detection rule [6] for single user can also be applied to the MUD with matched filtering. The estimation of $r_{t,k}^{j}(\hat{r}_{t,k}^{j})$ for the kth user based on (7) is obtained after the matched filter from the received signal $\mathbf{r}$. For example, the decisions based on the estimation in [6] for two receive antennas with the $\mathcal{G}_2$ design are

$$\begin{aligned}\hat{x}_{k,1} &= \alpha_{1,1}^{k^*}\hat{\underline{r}}_{1,k}^{1} + \alpha_{2,1}^{k}\hat{\underline{r}}_{2,k}^{1^*} + \alpha_{1,2}^{k^*}\hat{\underline{r}}_{1,k}^{2} + \alpha_{2,2}^{k}\hat{\underline{r}}_{2,k}^{2^*},\\ \hat{x}_{k,2} &= \alpha_{2,1}^{k^*}\hat{\underline{r}}_{1,k}^{1} - \alpha_{1,1}^{k}\hat{\underline{r}}_{2,k}^{1^*} + \alpha_{2,2}^{k^*}\hat{\underline{r}}_{1,k}^{2} - \alpha_{1,2}^{k}\hat{\underline{r}}_{2,k}^{2^*}.\end{aligned} \tag{16}$$

In this paper, if we let $\hat{\mathbf{x}}$ be the estimation of the original signal $\mathbf{x}$, the estimation error can be obtained assuming perfect CSI:

$$f_1 = \left\|\mathbf{r} - \mathbf{H}\hat{\mathbf{x}}\right\|^2. \tag{17}$$

As mentioned earlier, a DD or MMSE detector is applied at the receiver side in the MUD so as to create the seed chromosomes for the subsequent GA operations in order to improve the receiver performance. Firstly, the estimation of the original signal is to be calculated by either the DD or the MMSE detector and designated as the seed chromosome. Then this seed chromosome is perturbed randomly to form the initial population, which will undergo further GA operations. The DD estimation of the original signal is

$$\hat{\mathbf{x}}_{\mathrm{DD}} = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{r}, \tag{18}$$

where $\mathbf{H}$ is assumed to be full column rank. The corresponding MMSE detector estimation is

$$\hat{\mathbf{x}}_{\mathrm{MMSE}} = \left(\mathbf{R}_{rr}^{-1}\mathbf{H}\right)^H\mathbf{r}, \tag{19}$$

where $\mathbf{R}_{rr}$ is the self-correlation of the received signals from all the receive antennas. The inversion operations in (18) and (19) only appear once and the subsequent operations do not require such inversions so that the total computation time is little affected if the number of evaluations of (17) is fairly large.

3. THE GENETIC ALGORITHM

In general, a GA is composed of the following steps.

(1) Initial population generation: all initial chromosomes are encoded in bit level to simplify the following GA recombination operations. They are either generated randomly or derived from the DD decision or the MMSE decision. Each chromosome is a combination of the probable solution for all users. Normally, the population size is taken as the product between K, the number of users, and Q^{K_0}, the number of all possible solutions of each user.

(2) Fitness value calculation: the MSE shown in (17) can act as the objective function to evaluate the fitness of each chromosome. The optimal solution of (17) should yield a minimum value. In fact, the ML method evaluates every possible combination of bits so the computation time varies exponentially with the number of users. As explained before, the use of GA in an STBC MUD system can reduce the computation times of (17) significantly. To make the search more effective, an improved version of the objective function over that defined by (17) is proposed here:

$$f_2 = \left\| \frac{\mathbf{r}}{\|\mathbf{r}\|} - \frac{\mathbf{H}\hat{\mathbf{x}}}{\|\mathbf{H}\hat{\mathbf{x}}\|} \right\|^2. \tag{20}$$

The expression calculates the phase difference between $\mathbf{r}$ and $\mathbf{H}\hat{\mathbf{x}}$ and should therefore be more sensitive to the changes of either or both vectors. The best chromosome in a generation should have the least value of the objective function. If the value of the best chromosome in the present generation is larger than its counterpart in the previous generation, the chromosome with the largest value of the objective function in the present generation will be replaced by the best chromosome of the previous generation. This operation ensures that at least the useful information contained in the present generation is passed on to the next. A chromosome is usually considered to be better if it has a larger fitness value. Hence a fitness value can be defined with the help of (17) or (20) as follows:

$$f = f_0 - f_i, \quad i = 1, 2, \tag{21}$$

where f_0 is a sufficiently large constant and can be taken as the largest value of f_1 or f_2 within the whole population. Obviously, the larger the fitness value, the better the chromosome is.

(3) If the optimal criterion is satisfied, that is, when any one of all f_i in the population is less than a predetermined threshold, or if the generation number has exceeded a predefined value, which is also commonly taken as the product between K and Q^{K_0}, then go to Step (9). Otherwise, go to Step (4).

(4) Selection: this operation is based on the Roulette-wheel rule [9] and the probability of each chromosome being selected is calculated using the fitness value obtained with (21) in Step (2). It serves to provide the chromosomes for the subsequent recombination operations.

(5) Reproduction: this step is intended to replace the chromosome of the largest objective function by the best chromosome of the same generation.

(6) Crossover: this operation exchanges some parts of the chromosomes to provide a chance for a chromosome to include more signal information. Since the objective function is calculated in symbol (8 PSK) level and normally erroneous symbols are detected adjacent to the correct symbols, the crossover operation is carried in symbol level. In this paper, a single-point crossover is adopted.

(7) Mutation: this operation can enhance the convergence of the GA in case the original signal information has

Table 1: The time needed for different detectors.

Receiver	RR-f_1	RR-f_2	RDD-f_1	RDD-f_2	RMD-f_1	RMD-f_2	ML ($K = 1$)	ML ($K = 4$)
Time needed per block(s)	3.287	3.259	1.124	1.092	1.123	1.090	6.479	998.559

not been included in the initial population. For example, if the second bit of all chromosomes is "0" whereas the real signal bit is "1," then the only way leading to the right solution is by mutation. The mutation operation is also carried out in symbol level, where a symbol may be mutated to its adjacent symbol in the constellation according to a certain selected mutation probability. Generally, the crossover probability p_c is close to 1 and the mutation probability p_m is close to 0.

(8) Go to Step (2) for the next generation.

(9) End and output the decisions.

Sometimes the bit inversion operation is performed in the GA but it can be regarded as a special kind of crossover so that it is not considered here. The operations of Step (6) and Step (7) may also lead to chromosomes with less information but the operation in Step (5) can reverse this degenerative effect. For given predefined number of generations and population size, the computation times of (17) vary linearly with $(KQ^{K_0})^2$, which is much smaller than the factor of Q^{KK_0} in the ML detection. The improvement is clearly significant.

The significant features of the GA proposed in this paper are the introduction of an improved objective function and the preparation of an initial population that already contains some signal information from the output of a DD or MMSE detector given by (18) or (19) instead of a blind and random selection of the initial population as suggested in [13].

4. SIMULATIONS

In the following simulations, 8 PSK modulation and $\mathcal{G}_2$ STBC of rate 1 have been adopted and the number of users is $K = 4$. Hence $Q = 8$, $N = 2$, and the signal vector length, that is, the chromosome length, is $L = K^* K_0^* \log_2 Q = 24$. The number of receive antennas is $M = 2$. The predefined number of generations and the population size are both 256 since the number of all possible solutions for each user is $Q^{K_0} = 64$. The improvement of the computation time in the GA STBC MUD is therefore 256 times over the ML detection. The recombination operation parameters are $p_m = 0.05$ and $p_c = 0.95$. The Roulette-wheel selection rule and the single-point crossover are adopted. The channel is Rayleigh flat fading and is maintained constant during the whole block period. The path gain is taken as an independent complex Gaussian random variable with zero mean and a variance of 0.5 per dimension.

The computation times for each block averaged from 1000 times of Monte Carlo simulations in Matlab are given in Table 1 for various simulation schemes. In the table, "f_1" and "f_2" represent the receiver based on the objective function given by (17) and (20) respectively. "RR" refers to the receiver with the seed chromosome chosen randomly. "RDD" refers to the receiver with its seed chromosome created from the DD detector. "RMD" refers to the receiver with its seed chromosome created from an MMSE detector. "ML ($K = 1$)" represents the ML detector for the single-user case. The time per block needed for "ML ($K = 4$)" is so large that only several blocks are simulated for reference and therefore the BER performance of this case is not shown in the following figures. The same notations are also used in the following figures. The criterion for terminating the algorithm is when the objective function from (17) or (20) is less than a predefined threshold. This threshold is selected from the smallest value of (17) or (20), where the training signal is obtained by setting $\hat{\mathbf{x}}$ to $\mathbf{x}$. The resulting threshold is 0.5 for f_1 or 0.1 for f_2 in this paper. The table shows clearly that the time needed for our proposed GA STBC MUD is only about 1/5 of that of the single-user receiver with ML detection, and the improvement over the MUD is more than 256 times since some block detection may have the GA operations terminated before the last generation. The computation time required when the seed chromosome is prepared from either the DD detector or the MMSE detector is a few times further smaller than that when the initial chromosomes are randomly chosen. The computation time for the modified objective function f_2 is about 10% less than that for the original function f_1 because of its quicker convergence.

The BER performance versus the number of generations for the various detection schemes employing the objective function f_1 or f_2 is shown in Figure 2 at SNR = 6 dB. For comparison, the BER curves for the iterative MAP method suggested in [3] after the 6th iteration (6th iter. of IMAP in the figure) for the CD, DD, MMSE, and ML detectors are also given. By comparing Figure 2a with Figure 2b, it can be observed that the improvement of RDD and RMD over RR is more significant with f_2 than with f_1. Besides, the performance of RDD is comparable to that of RMD. All proposed GA receivers outperform the CD after 30 or 10 generations with f_1 or f_2, respectively. RDD or RMD detector initially performs the same as DD or MMSE detector but gradually achieves a better BER performance after GA operations.

Figure 3 gives a comparison between the two objective functions when SNR = 6 dB for RR, RDD, and RMD, respectively. The receiver with the objective function f_2 converges to a BER of 10^{-2} about 80 generations sooner than that with f_1 for both RDD and RMD, which is also the reason why the computation time for the final decision is smaller with f_2 than with f_1. A similar behavior is observed with RR but the improvement is not so significant. It is also obvious that the performance of the receivers with f_2 approaches to that of the single-user ML detection much faster and nearer than those with f_1.

Figure 4 shows the BER versus SNR performance of the final GA output of various receivers with the objective

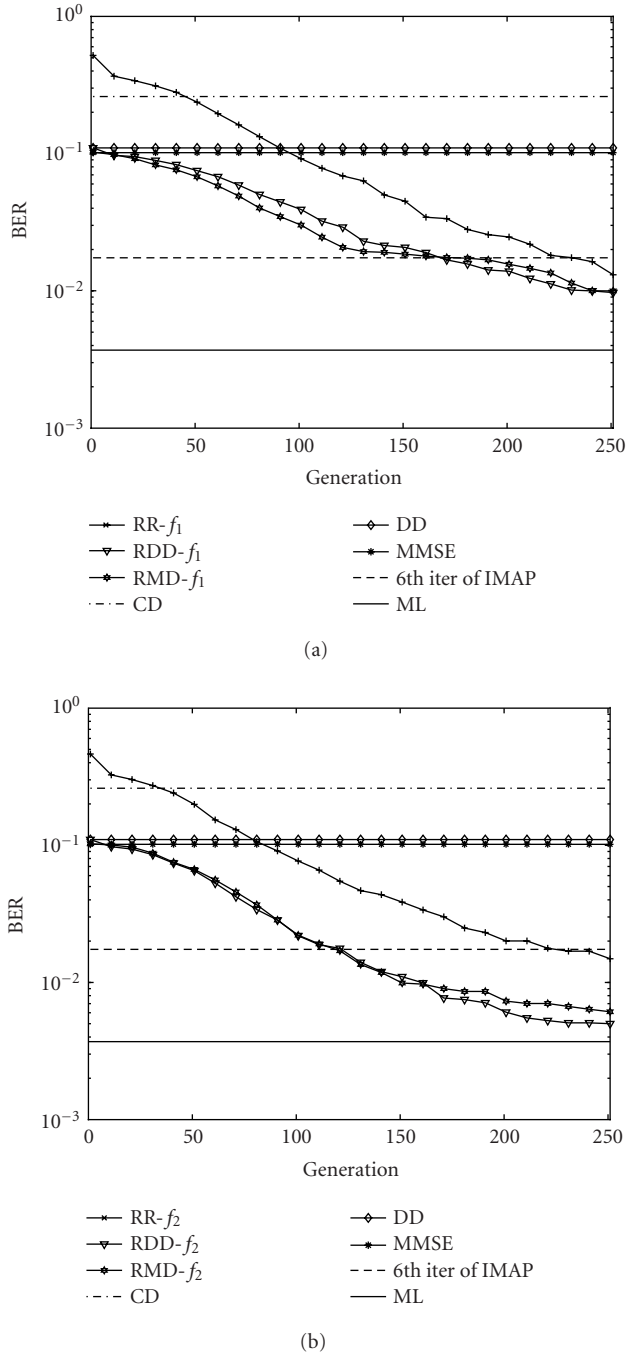


FIGURE 2: Performance comparison of various receivers at SNR = 6 dB with the objective function (a) f_1 and (b) f_2.

function f_1 or f_2, respectively. Figure 4a shows that RDD and RMD with the objective function f_1 outperform RR by about 2 dB at the BER of 10^{-2}. Here the comparison is referenced at the BER of 10^{-2} instead of 10^{-3} just because the results cannot approach 10^{-3}. Figure 4b shows that RDD and RMD with the objective function f_2 outperform RR by about 3 dB at the BER of 10^{-2} and can outperform 6th iteration results of the iterative MAP detection. The SNR degradation compared

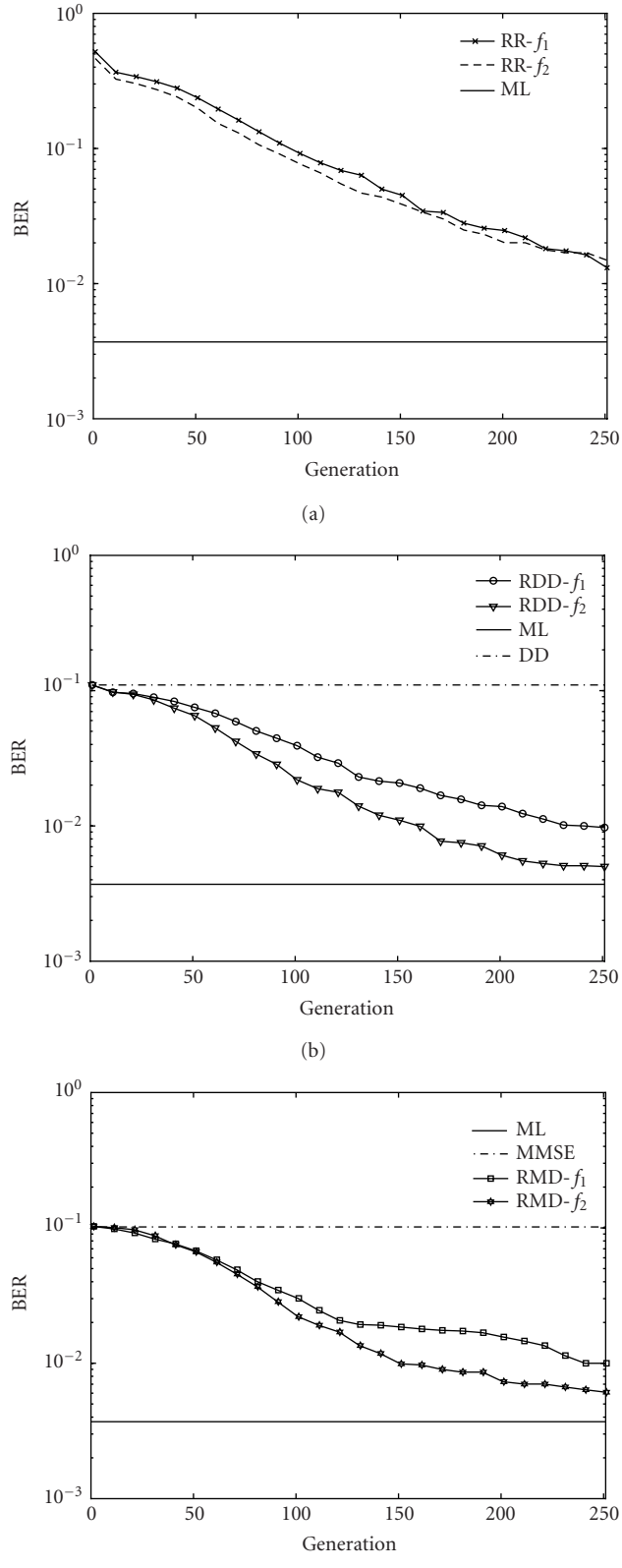


FIGURE 3: Performance comparison at SNR = 6 dB of different objective functions for receiver with the seed chromosome (a) randomly chosen, (b) created by a DD detector, and (c) created by an MMSE detector.

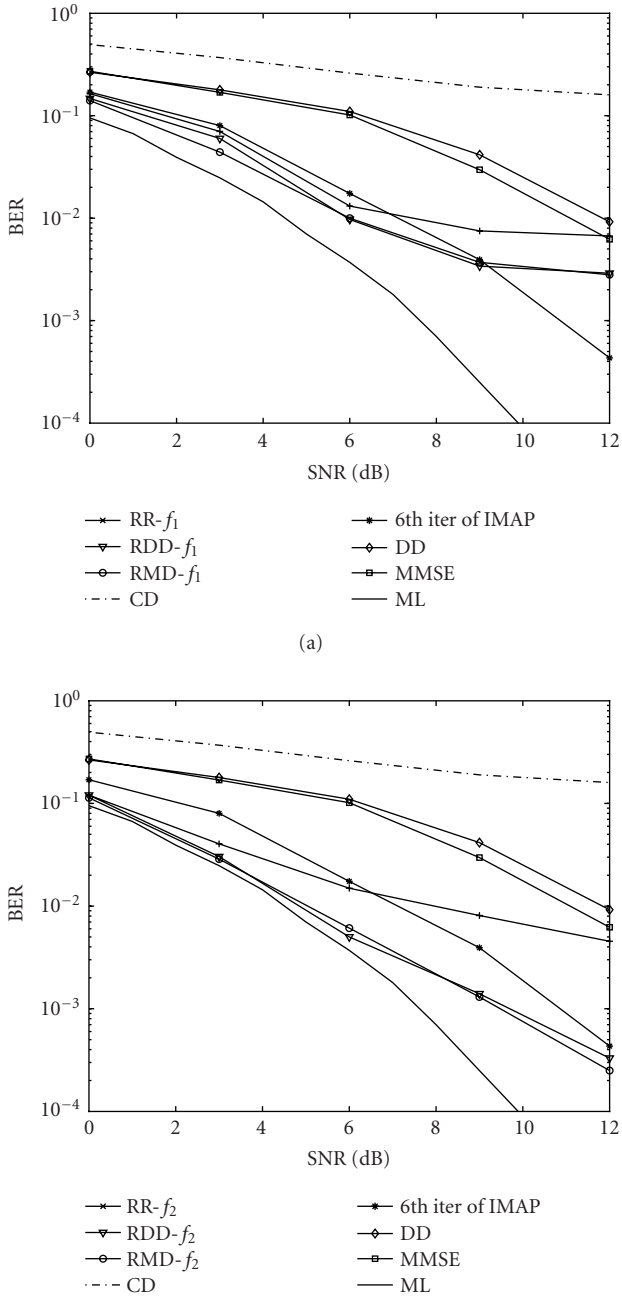


Figure 4: BER versus SNR performance of various receivers with the objective function (a) f_1 and (b) f_2.

with the single-user ML performance is about 2 dB for RMD and RDD at the BER of 10^{-3}. Furthermore, RMD outperforms RDD by a very small amount with both objective functions. The improvement of all the proposed receivers over the CD, DD, and MMSE is significant. However, Figure 4 also shows that there may be a bound for those detectors with f_1, which will limit any further application of such detectors.

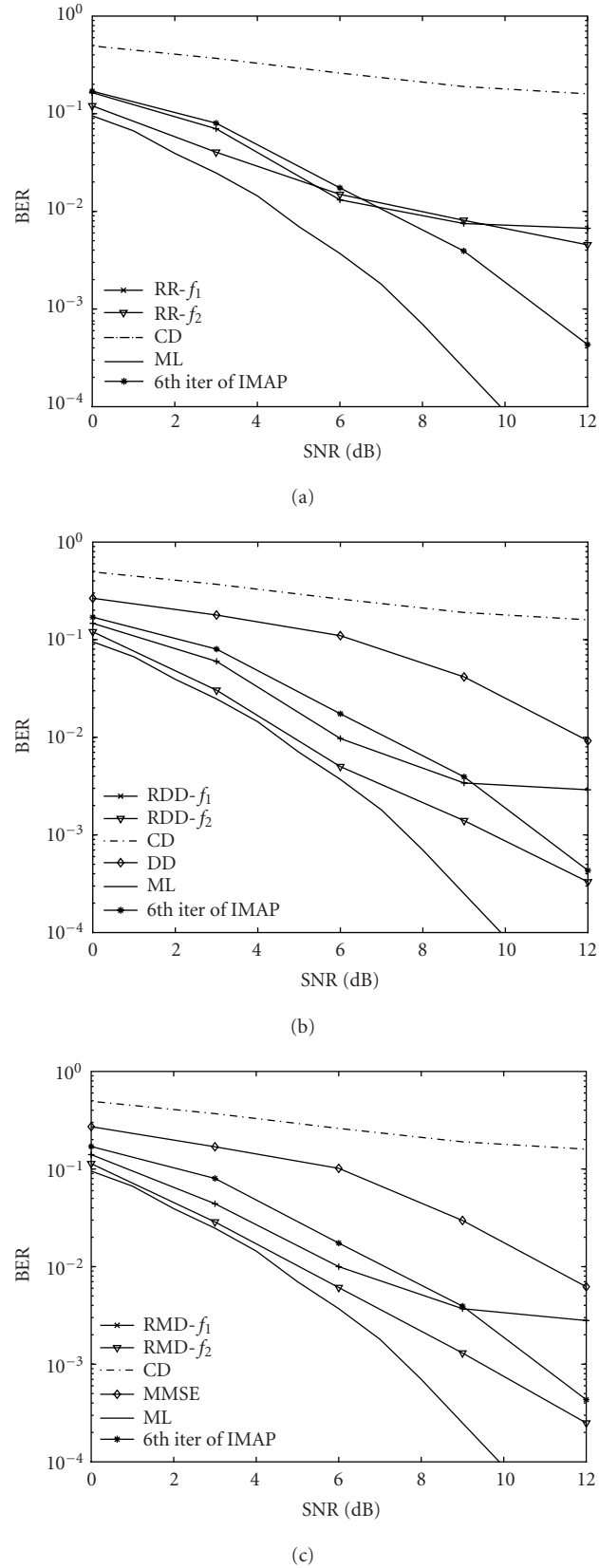


Figure 5: BER performance comparison of different objective functions for receiver with the seed chromosome (a) randomly chosen; (b) created by a DD detector, and (c) created by an MMSE detector.

Figure 5 shows the detailed performance comparison between f_1 and f_2 for RR, RDD, and RMD, respectively. Clearly the receivers with f_2 outperform those with f_1. The SNR improvement at the BER of 10^{-2} is about 1.3 dB or 1 dB for RDD or RMD, respectively.

5. CONCLUSIONS

The GA has previously been shown to be a feasible technique to solve the STBC MUD problem requiring less computing resources. To further improve its performance, two modifications have been proposed in this paper. Firstly, a new objective function is introduced which includes the phase information of the relevant signal vectors in order to make the decision more accurate. It contributes to about 10% reduction of detection time and about 1.3 or 1 dB SNR improvement at the BER of 10^{-2} for the GA receiver with DD or MMSE detector, respectively. It also requires fewer generations to converge. Secondly, the DD and MMSE detectors have been embedded into the GA STBC MUD system to generate the seed chromosome and thus provide some signal information to the first generation. The resulting simulations confirm that the receivers thus designed can converge faster than that with the initial population randomly chosen. The improvement in SNR is about 2–3 dB at the BER of 10^{-2}. Therefore, the total SNR improvement of the best receiver proposed here can reach 3–4 dB at the BER of 10^{-2} when compared with the previous reported GA STBC MUD receiver. This receiver performance is also better than the DD, MMSE, or MAP detector. The degradation in SNR when compared with a single-user ML detector is limited to about 2 dB at the BER of 10^{-3}. All the above results suggest that the proposed improved GA receiver is a promising solution of the STBC MUD problem with reasonable computation complexity and fairly good performance.

REFERENCES

[1] S. Barbarossa and F. Cerquetti, "Simple space-time coded SS-CDMA systems capable of perfect MUI/ISI elimination," *IEEE Communications Letters*, vol. 5, no. 12, pp. 471–473, 2001.

[2] S. Iraji and J. Lilleberg, "EM-based multiuser detection and parallel interference cancellation for space-time block coded WCDMA systems employing 16-QAM over multipath fading channels," in *Proc. 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 2, pp. 688–692, Lisboa, Portugal, September 2002.

[3] B. Lu and X. Wang, "Iterative receivers for multiuser space-time coding systems," in *Proc. IEEE International Conference on Communications*, vol. 1, pp. 302–306, New Orleans, La, USA, June 2000.

[4] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criteria in the presence of channel estimation errors, mobility, and multiple paths," *IEEE Trans. Communications*, vol. 47, no. 2, pp. 199–207, 1999.

[5] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: performance results," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 3, pp. 451–460, 1999.

[6] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.

[7] S. Verdú, *Multiuser Detection*, Cambridge University Press, New York, NY, USA, 1998.

[8] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 284–287, 1974.

[9] E. Falkenauer, *Genetic Algorithms and Grouping Problems*, John Wiley & Sons, New York, NY, USA, 1998.

[10] P. J. M. van Laarhoven and E. H. L. Aarts, *Simulated Annealing: Theory and Applications*, vol. 37 of *Mathematics and Its Applications*, D. Reidel, Dordrecht, Netherlands, 1987.

[11] F. Glover and M. Laguna, *Tabu Search*, Kluwer Academic Publishers, Boston, Mass, USA, 1997.

[12] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, MIT Press, Cambridge, Mass, USA, 1st edition, 1975.

[13] Y. Du and K. T. Chan, "Feasibility of applying genetic algorithms in space-time block coding multiuser detection systems," in *Proc. 3rd IASTED International Conference on Wireless and Optical Communications*, pp. 469–473, Banff, Albert, Canada, July 2003.

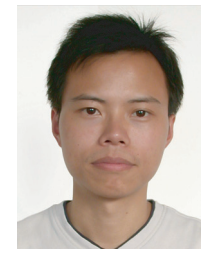
[14] C. Ergun and K. Hacioglu, "Multiuser detection using a genetic algorithm in CDMA communications systems," *IEEE Trans. Communications*, vol. 48, no. 8, pp. 1374–1383, 2000.

[15] M. J. Juntti, T. Schlosser, and J. O. Lilleberg, "Genetic algorithms for multiuser detection in synchronous CDMA," in *Proc. IEEE International Symposium on Information Theory*, p. 492, Ulm, Germany, June–July 1997.

[16] K. Yen and L. Hanzo, "Genetic algorithm assisted joint multiuser symbol detection and fading channel estimation for synchronous CDMA systems," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 6, pp. 985–998, 2001.

[17] A. F. Naguib, N. Seshadri, and A. R. Calderbank, "Increasing data rate over wireless channels," *IEEE Signal Processing Magazine*, vol. 17, no. 3, pp. 76–92, 2000.

[18] Y. Du and K. T. Chan, "A multiuser space-time block coding system with squaring method," in *Proc. IEEE International Conference on Neural Networks & Signal Processing*, vol. 2, pp. 1450–1454, Nanjing, China, December 2003.



Yinggang Du received his B.Eng. and M. Eng. degrees in 1997 and 2000, respectively, both from the Department of Electronic Engineering, Nanjing University of Science & Technology (NJUST). Now he is pursuing his Ph.D. degree at the Department of Electronic Engineering, the Chinese University of Hong Kong (CUHK). He acted as an Assistant Professor from April to September 2000 at NJUST and a Teaching Assistant from October 2000 to September 2003 in CUHK. He has been a Research Assistant since October 2003. He is a student member of IEEE and his present research interests include space-time coding, radar signal processing, wireless communications, genetic algorithm, and digital signal processing.

Kam Tai Chan received his Ph.D. degree from Cornell University in applied physics in March 1986. His thesis research involved the preparation of ultrathin compound semiconductor materials for optoelectronics and quantum-size effect devices. He stayed on at Cornell University as a Postdoctoral Research Associate to work on high-power lasers and integrated photodetectors after graduation. He joined Hewlett-Packard Company in July 1986 at its Microwave Technology Division. He participated in projects that were related to photodetectors and high electron mobility transistors. In 1989, he was invited by Lawrence Berkeley Laboratory, University of California at Berkeley, to serve as a Visiting Industrial Fellow to develop industrial applications of the extensive sophisticated instrumentation in the Laboratory. He resigned from Hewlett-Packard at the end of 1991 to assume his present position at the Chinese University of Hong Kong. He is a member of IEEE and his present research interests include ultrafast lasers, novel photonic devices, optical switches, optical CDMA, wireless communication, and quantum cryptography.

EURASIP Journal on Applied Signal Processing 2004:5, 649–661

Performance Comparisons of MIMO Techniques with Application to WCDMA Systems

Chuxiang Li
Department of Electrical Engineering, Columbia University, New York, NY 10027, USA
Email: cxli@ee.columbia.edu

Xiaodong Wang
Department of Electrical Engineering, Columbia University, New York, NY 10027, USA
Email: wangx@ee.columbia.edu

Received 11 December 2002; Revised 1 August 2003

Multiple-input multiple-output (MIMO) communication techniques have received great attention and gained significant development in recent years. In this paper, we analyze and compare the performances of different MIMO techniques. In particular, we compare the performance of three MIMO methods, namely, BLAST, STBC, and linear precoding/decoding. We provide both an analytical performance analysis in terms of the average receiver SNR and simulation results in terms of the BER. Moreover, the applications of MIMO techniques in WCDMA systems are also considered in this study. Specifically, a subspace tracking algorithm and a quantized feedback scheme are introduced into the system to simplify implementation of the beamforming scheme. It is seen that the BLAST scheme can achieve the best performance in the high data rate transmission scenario; the beamforming scheme has better performance than the STBC strategies in the diversity transmission scenario; and the beamforming scheme can be effectively realized in WCDMA systems employing the subspace tracking and the quantized feedback approach.

Keywords and phrases: BLAST, space-time block coding, linear precoding/decoding, subspace tracking, WCDMA.

1. INTRODUCTION

Multiple-input multiple-output (MIMO) communication technology has received significant recent attention due to the rapid development of high-speed broadband wireless communication systems employing multiple transmit and receive antennas [1, 2, 3]. Many MIMO techniques have been proposed in the literature targeting at different scenarios in wireless communications. The BLAST system is a layered space-time architecture originally proposed by Bell Labs to achieve high data rate wireless transmissions [4, 5, 6]. Note that the BLAST systems do not require the channel knowledge at the transmitter end. On the other hand, for some applications, the channel knowledge is available at the transmitter, at least partially. For example, an estimate of the channel at the receiver can be fed back to the transmitter in both frequency division duplex (FDD) and time division duplex (TDD) systems, or the channel can be estimated directly by the transmitter during its receiving mode in TDD systems. Accordingly, several channel-dependent signal processing schemes have been proposed for such scenarios, for example, linear precoding/decoding [7]. The linear precoding/decoding schemes achieve performance gains by allocating power and/or rate over multiple transmit antennas, with partially or perfectly known channel state information [7]. Another family of MIMO techniques aims at reliable transmissions in terms of achieving the full diversity promised by the multiple transmit and receive antennas. Space-time block coding (STBC) is one of such techniques based on orthogonal design that admits simple linear maximum likelihood (ML) decoding [8, 9, 10]. The trade-off between diversity and multiplexing gain are addressed in [11, 12], which are from a signal processing perspective and from an information theoretic perspective, respectively.

Some simple MIMO techniques have already been proposed to be employed in the third-generation (3G) wireless systems [13, 14]. For example, in the 3GPP WCDMA standard, there are open-loop and closed-loop transmit diversity options [15, 16]. As more powerful MIMO techniques emerge, they will certainly be considered as enabling techniques for future high-speed wireless systems (i.e., 4G and beyond).

The purpose of this paper is to compare the performance of different MIMO techniques for the cases of two and four transmit antennas, which are realistic scenarios for MIMO applications. For a certain transmission rate, we

compare the performance of three MIMO schemes, namely, BLAST, STBC, and linear precoding/decoding. Note that both BLAST and STBC do not require channel knowledge at the transmitter, whereas linear precoding/decoding does. For each of these cases, we provide an analytical performance analysis in terms of the receiver output average signal-to-noise ratio (SNR) as well as simulation results on their BER performance. Moreover, we also consider the application of these MIMO techniques in WCDMA systems with multipath fading channel. In particular, when precoding is used, a subspace tracking algorithm is needed to track the eigenspace of the MIMO system at the receiver and feed back this information to the transmitter [17, 18, 19, 20]. Since the feedback channel typically has a very low bandwidth [21], we contrive an efficient and effective quantized feedback approach.

The main findings of this study are as follows.

(i) In the high data rate transmission scenario, for example, four symbols per transmission over four transmit antennas, the BLAST system actually achieves a better performance than the linear precoding/decoding schemes, even though linear precoding/decoding make use of the channel state information at the transmitter.

(ii) In the diversity transmission scenario, for example, one symbol per transmission over two or four transmit antennas, beamforming offers better performance than the STBC schemes. Hence the channel knowledge at the transmitter helps when there is some degree of freedom to choose the eigen channels.

(iii) By employing the subspace tracking technique with an efficient quantized feedback approach, the beamforming scheme can be effective and feasible to be employed in WCDMA systems to realize reliable data transmissions.

The remainder of this paper is organized as follows. In Section 2, performance analysis and comparisons of different MIMO techniques are given for the narrowband scenario. Section 3 describes the WCDMA system based on the 3GPP standard, the channel estimation method, the algorithm of tracking the MIMO eigen-subspace, as well as the quantized feedback approach. Simulation results and further discussions are given in Section 4. Section 5 contains the conclusions.

2. PERFORMANCE ANALYSIS AND COMPARISONS OF MIMO TECHNIQUES

In this section, we analyze the performance of several MIMO schemes under different transmission rate assumptions, for the cases of two and four transmit antennas. BLAST and linear precoding/decoding schemes are studied and compared for high-rate transmissions in Section 2.1. Section 2.2 focuses on the diversity transmission scenario, where different STBC strategies are investigated and compared with beamforming and some linear precoding/decoding approaches.

2.1. BLAST versus linear precoding for high-rate transmission

Assume that there are n_T transmit and n_R receive antennas, where $n_R \geq n_T$. In this section, we assume that the MIMO system is employed to achieve the highest data rate, that is, n_T symbols per transmission. When the channel is unknown to the transmitter, the BLAST system can be used to achieve this; whereas when the channel is known to the transmitter, the linear precoding/decoding can be used to achieve this.

2.1.1. BLAST

In the BLAST system, at each transmission, n_T data symbols $s_1, s_2, \ldots, s_{n_T}$, $s_i \in \mathcal{A}$, where $\mathcal{A}$ is some unit-energy (i.e., $\mathbb{E}\{|s_i|^2\} = 1$) constellation signal set (e.g., PSK, QAM), are transmitted simultaneously from all n_T antennas. The received signal can be represented by

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_R} \end{bmatrix}}_{\mathbf{y}} = \sqrt{\frac{\rho}{n_T}} \underbrace{\begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,n_T} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,n_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_R,1} & h_{n_R,2} & \cdots & h_{n_R,n_T} \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{n_T} \end{bmatrix}}_{\mathbf{s}} + \underbrace{\begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_{n_R} \end{bmatrix}}_{\mathbf{n}}, \tag{1}$$

where y_i denotes the received signal at the ith receive antenna; $h_{i,j}$ denotes the complex channel gain between the ith receive antenna and the jth transmit antenna; ρ denotes the total transmit SNR; and $\mathbf{n} \sim \mathcal{N}_c(\mathbf{0}, \mathbf{I}_{n_R})$.

The received signal is first matched filtered to obtain $\mathbf{z} = \mathbf{H}^H\mathbf{y} = \sqrt{\rho/n_T}\mathbf{H}^H\mathbf{H}\mathbf{s} + \mathbf{H}^H\mathbf{n}$. Denote $\mathbf{\Omega} \triangleq \mathbf{H}^H\mathbf{H}$ and $\mathbf{w} \triangleq \mathbf{H}^H\mathbf{n}$, and thus, $\mathbf{w} \sim \mathcal{N}_c(\mathbf{0}, \mathbf{\Omega})$. The matched-filter output is then whitened to get

$$\mathbf{u} = \mathbf{\Omega}^{-1/2}\mathbf{z} = \sqrt{\frac{\rho}{n_T}}\mathbf{\Omega}^{1/2}\mathbf{s} + \tilde{\mathbf{v}}, \tag{2}$$

where $\tilde{\mathbf{v}} \triangleq \mathbf{\Omega}^{-1/2}\mathbf{w} \sim \mathcal{N}_c(\mathbf{0}, \mathbf{I}_{n_R})$. Based on (2), several methods can be used to detect the symbol vector $\mathbf{s}$. For example, the ML detection rule is given by

$$\hat{\mathbf{s}}_{\mathrm{ML}} = \arg\min_{\mathbf{s}\in\mathcal{A}^{n_T}} \left\| \mathbf{u} - \sqrt{\frac{\rho}{n_T}}\mathbf{\Omega}^{1/2}\mathbf{s} \right\|^2, \tag{3}$$

which has a computational complexity exponential in the number of transmit antennas n_T. On the other hand, the sphere decoding algorithm offers a near-optimal solution to (2) with an expected complexity of $\mathcal{O}(n_T^3)$ [22]. Moreover, a linear detector makes a symbol-by-symbol decision $\hat{\mathbf{s}} = \mathbb{Q}(\mathbf{x})$, where $\mathbf{x} = \mathbf{G}\mathbf{u}$ and $\mathbb{Q}(\cdot)$ denotes the symbol slicing operation. Two forms of linear detectors can be used [5, 6], namely, the linear zero-forcing detector, where $\mathbf{G} = \mathbf{\Omega}^{-1/2}$, and the linear MMSE detector, where $\mathbf{G} = (\mathbf{\Omega}^{1/2} + (n_T/\rho)\mathbf{I})^{-1}$. Finally, a method based on interference cancellation with ordering offers improved performance over the linear detectors with comparable complexity [22]. Note that among

the above-mentioned BLAST decoding algorithms, the linear zero-forcing detector has the worst performance. The decision statistics of this method is given by

$$\mathbf{x} = \mathbf{G}\mathbf{u} = \boldsymbol{\Omega}^{-1/2}\mathbf{u} = \sqrt{\frac{\rho}{n_T}}\mathbf{s} + \boldsymbol{\Omega}^{-1/2}\tilde{\mathbf{v}}. \tag{4}$$

It follows from (4) that the received SNR for symbol s_j is $(\rho/n_T)/[\boldsymbol{\Omega}^{-1}]_{j,j}$, $j = 1, 2, \ldots, n_T$. Hence the average received SNR under linear zero-forcing BLAST detection is given by

$$\overline{\mathrm{SNR}}_{\text{BLAST-LZF}} = \rho\left(\frac{1}{n_T^2}\sum_{j=1}^{n_T}\frac{1}{[\boldsymbol{\Omega}^{-1}]_{j,j}}\right). \tag{5}$$

2.1.2. Linear precoding and decoding

When the channel $\mathbf{H}$ is known to the transmitter, a linear precoder can be employed at the transmitter and a corresponding linear decoder can be used at the receiver [7]. Specifically, suppose $m \le n_T$ symbols $\mathbf{s} = [s_1 \;\; s_2 \;\; \cdots \;\; s_m]^T$ are transmitted per transmission, where $m = \mathrm{rank}(\mathbf{H})$. Then the linear precoder is an $n_T \times m$ matrix $\mathbf{F}$ such that the transmitted signal is $\mathbf{Fs}$. The $n_R \times 1$ received signal vector is then

$$\mathbf{y} = \mathbf{HFs} + \mathbf{n}, \tag{6}$$

where $\mathbf{n} \sim \mathcal{N}_c(\mathbf{0}, \mathbf{I}_{n_R})$. At the receiver, $\mathbf{y}$ is first matched filtered, and then an $m \times n_T$ linear decoder $\mathbf{G}$ is applied to the matched-filter output to obtain the decision statistics

$$\mathbf{x} = \mathbf{G}\mathbf{H}^H\mathbf{y} = \mathbf{G}\boldsymbol{\Omega}\mathbf{Fs} + \mathbf{G}\mathbf{H}^H\mathbf{n}. \tag{7}$$

The linear precoder $\mathbf{F}$ and decoder $\mathbf{G}$ are chosen to minimize a weighted combination of symbol estimation errors, that is, $\min_{\mathbf{F},\mathbf{G}} \mathbb{E}\{\|\mathbf{D}^{1/2}(\mathbf{s}-\mathbf{x})\|^2\}$, where $\mathbf{D}$ is a diagonal positive definite matrix subject to the total transmitter power constraint $\mathrm{tr}(\mathbf{FF}^H) \le \rho$. The weight matrix $\mathbf{D}$ is such that all decoded symbols have equal errors (equal error design). Denote the eigendecomposition of $\boldsymbol{\Omega}$ as $\boldsymbol{\Omega} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H + \tilde{\mathbf{V}}\tilde{\boldsymbol{\Lambda}}\tilde{\mathbf{V}}^H$, where $\boldsymbol{\Lambda}$ and $\mathbf{V}$ contain the m largest eigenvalues and the corresponding eigenvectors of $\boldsymbol{\Omega}$, respectively; and $\tilde{\boldsymbol{\Lambda}}$ and $\tilde{\mathbf{V}}$ contain the remaining $(n_T - m)$ eigenvalues and the corresponding eigenvectors, respectively. Denote $\gamma = \rho/\mathrm{tr}(\boldsymbol{\Lambda}^{-1})$. Then the linear precoder and decoder are given by [7]

$$\begin{aligned}\mathbf{F} &= \gamma^{1/2}\mathbf{V}\boldsymbol{\Lambda}^{-1/2},\\ \mathbf{G} &= \frac{1}{\gamma^{-1/2} + \gamma^{1/2}}\boldsymbol{\Lambda}^{-1/2}\mathbf{V}^H.\end{aligned} \tag{8}$$

It can be verified that $\mathbf{GH}^H\mathbf{HF} = (1/(\gamma^{-1} + \gamma))\mathbf{I}_m$. Hence this precoding scheme transforms the MIMO channel into a scaled identity matrix. Furthermore, the received SNRs for all decoded symbols are equal, given by γ, that is,

$$\overline{\mathrm{SNR}}_{\text{equal-error precoding}} = \frac{\rho}{\mathrm{tr}\,(\boldsymbol{\Lambda}^{-1})} = \frac{\rho}{\mathrm{tr}\,(\boldsymbol{\Omega}^{-1})}. \tag{9}$$

Remark 1. The BLAST system can be viewed as a special case of linear precoding with the transmitter filter $\mathbf{F} = \sqrt{\rho/n_T}\mathbf{I}_{n_T}$. And the zero-forcing BLAST detection scheme corresponds to choosing the receiver filter $\mathbf{G} = \boldsymbol{\Omega}^{1/2}$.

Remark 2. An alternative precoding scheme is to choose $\mathbf{F} = \sqrt{\rho/n_T}\mathbf{V}$ and $\mathbf{G} = \mathbf{V}^H$. Then the output of the linear decoder can be written as

$$\mathbf{x} = \sqrt{\frac{\rho}{n_T}}\mathbf{V}^H\mathbf{H}^H\mathbf{HVs} + \mathbf{V}^H\mathbf{H}^H\mathbf{n} = \sqrt{\frac{\rho}{n_T}}\boldsymbol{\Lambda}\boldsymbol{s} + \mathbf{w}, \tag{10}$$

where $\mathbf{w} \sim \mathcal{N}_c(\mathbf{0}, \boldsymbol{\Lambda})$. Hence this scheme also transforms the MIMO channel into a set of independent channels, but with different SNRs. The received SNR for the jth symbol is $(\rho/n_T)\lambda_j$, where λ_j is the jth eigenvalue contained by $\boldsymbol{\Lambda}$. We call this method the whitening precoding. The average received SNR is given by

$$\overline{\mathrm{SNR}}_{\text{whitening precoding}} = \rho\left(\frac{1}{n_T^2}\sum_{j=1}^{n_T}\lambda_j\right) = \rho\left(\frac{\mathrm{tr}(\boldsymbol{\Omega})}{n_T^2}\right). \tag{11}$$

Note that the whitening precoding is different from the equal-error precoding in (8). In particular, different received SNRs are achieved over different subchannels for the whitening precoding, whereas the equal-error precoding provides the same SNR over all subchannels.

2.1.3. Comparisons

We have the following result on the relative SNR performance of the BLAST system and the two precoding schemes discussed above.

Proposition 1. *Suppose that an $n_T \times n_R$ MIMO system is employed to transmit n_T symbols per transmission, using either the BLAST system, the equal-error precoding scheme, or the whitening precoding scheme, then*

$$\begin{aligned}\overline{\mathrm{SNR}}_{\text{whitening precoding}} &\ge \overline{\mathrm{SNR}}_{\text{BLAST-LZF}}\\ &\ge \overline{\mathrm{SNR}}_{\text{equal-error precoding}}.\end{aligned} \tag{12}$$

Proof. We first show that

$$\overline{\mathrm{SNR}}_{\text{BLAST-LZF}} \ge \overline{\mathrm{SNR}}_{\text{equal-error precoding}}. \tag{13}$$

Since

$$\frac{1}{n_T}\sum_{j=1}^{n_T}\lambda_j^{-1} = \frac{1}{n_T}\sum_{j=1}^{n_T}[\boldsymbol{\Omega}^{-1}]_{j,j} \ge \frac{n_T}{\sum_{j=1}^{n_T}(1/[\boldsymbol{\Omega}^{-1}]_{j,j})}, \tag{14}$$

we have

$$\frac{1}{n_T^2}\sum_{j=1}^{n_T}\frac{1}{[\boldsymbol{\Omega}^{-1}]_{j,j}} \ge \frac{1}{\sum_{j=1}^{n_T}\lambda_j^{-1}}. \tag{15}$$

It follows from (5), (9), and (15) that $\overline{\mathrm{SNR}}_{\text{BLAST-LZF}} \ge \overline{\mathrm{SNR}}_{\text{equal-error precoding}}$.

We next show that $\overline{\text{SNR}}_{\text{BLAST-LZF}} \leq \overline{\text{SNR}}_{\text{whitening precoding}}$. First, we have the following.

Fact 1. Suppose that $\mathbf{A}$ is a $n \times n$ positive definite matrix, then

$$\frac{1}{[\mathbf{A}^{-1}]_{i,i}} = A_{i,i} - \tilde{\mathbf{a}}_i^H \tilde{\mathbf{A}}_i^{-1} \tilde{\mathbf{a}}_i, \tag{16}$$

where $\tilde{\mathbf{A}}_i$ is the $(n-1) \times (n-1)$ matrix obtained from $\mathbf{A}$ by removing the ith row and ith column; and $\tilde{\mathbf{a}}_i$ is the ith column of $\mathbf{A}$ with the ith entry $A_{i,i}$ removed. Note that $\tilde{\mathbf{A}}_i$ is a principal submatrix of $\mathbf{A}$; since $\mathbf{A}$ is positive definite, so is, $\tilde{\mathbf{A}}_i$, and $\tilde{\mathbf{A}}_i^{-1}$ exists. To see (16), denote the above-mentioned partitioning of the Hermitian matrix $\mathbf{A}$ with respect to the ith column and row by $\mathbf{A} = (\tilde{\mathbf{A}}_i, \tilde{\mathbf{a}}_i, A_{i,i})$. In the same way, we partition its inverse $\mathbf{B} \triangleq \mathbf{A}^{-1} = (\tilde{\mathbf{B}}_i, \tilde{\mathbf{b}}_i, B_{i,i})$. Now from the fact that $\mathbf{AB} = \mathbf{I}_n$, it follows that

$$A_{i,i} B_{i,i} + \tilde{\mathbf{a}}_i^H \tilde{\mathbf{b}}_i = 1, \qquad \tilde{\mathbf{a}}_i B_{i,i} + \tilde{\mathbf{A}}_i \tilde{\mathbf{b}}_i = \mathbf{0}. \tag{17}$$

Solving for $B_{i,i}$ from (17), we obtain (16).

Using (16), we have

$$\sum_{j=1}^{n_T} \frac{1}{[\mathbf{\Omega}^{-1}]_{j,j}} = \sum_{j=1}^{n_T} \left(\Omega_{i,i} - \tilde{\boldsymbol{\omega}}_i^H \tilde{\mathbf{\Omega}}_i^{-1} \tilde{\boldsymbol{\omega}}_i\right) \leq \sum_{j=1}^{n_T} \Omega_{i,i} = \text{tr}(\mathbf{\Omega}). \tag{18}$$

It then follows from (5), (11), and (18) that $\overline{\text{SNR}}_{\text{BLAST-LZF}} \leq \overline{\text{SNR}}_{\text{whitening precoding}}$. □

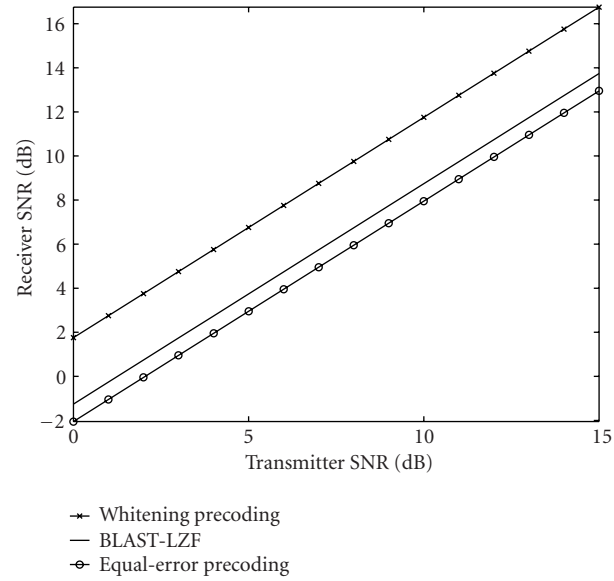
Figure 1 shows the comparisons between the BLAST and the linear precoding/decoding schemes in terms of the average receiver SNR as well as the BER for a system with $n_T = 4$ and $n_R = 6$. The rate is four symbols per transmission. The SNR curves in Figure 1a are plotted according to (9), (5), and (11). It is seen that the SNR curves confirm the conclusion of Proposition 1. Moreover, it is interesting to see that the SNR ordering given by (12) does not translate into the corresponding BER order. This can be roughly explained as follows. The BER for the ith symbol stream can be approximated as $Q(\gamma\sqrt{\overline{\text{SNR}}_i})$, where γ is a constant determined by the modulation scheme. The average BER is then

$$\overline{p} \cong \frac{1}{n_T} \sum_{i=1}^{n_T} Q\left(\gamma\sqrt{\overline{\text{SNR}}_i}\right). \tag{19}$$

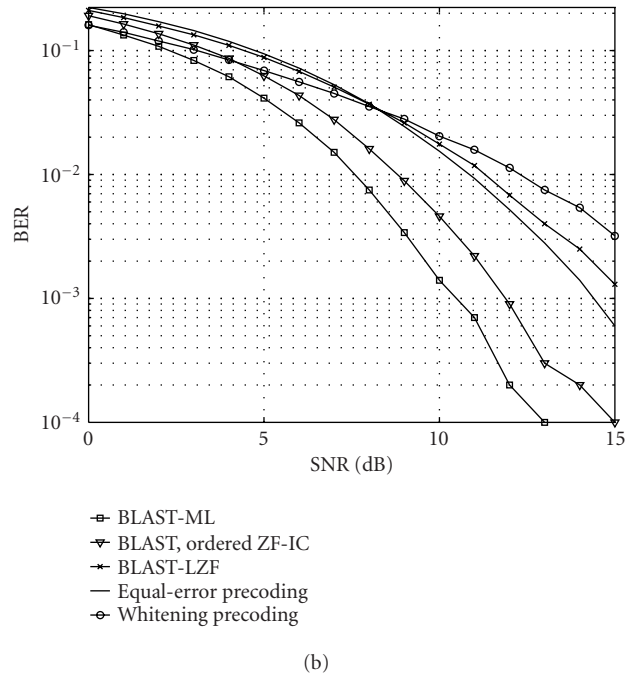
Since $Q(\cdot)$ is a concave function, we have

$$\overline{p} \leq Q\left(\gamma\sqrt{\overline{\text{SNR}}}\right). \tag{20}$$

Hence, the average SNR value does not directly translate into the average BER. Moreover, it is seen from the Figure 1b in Figure 1 that the interference cancellation with ordering [6] BLAST detection method offers a significant performance gain over the linear zero-forcing method, making the BLAST outperform the precoding schemes by a substantial margin.



(a)



(b)

Figure 1: Comparisons of the average receiver SNR and the BER between the BLAST and the linear precoding/decoding schemes: $n_T = 4$ and $n_R = 6$; the rate is four symbols/transmission.

2.2. Space-time block coding versus beamforming for diversity transmission

In contrast to the high data rate MIMO transmission scenario discussed in Section 2.1, an alternative approach to exploiting MIMO systems targets at achieving the full diversity. For example, with n_T transmit antennas and n_R receive

antennas, a maximum diversity order of $n_T n_R$ is possible when the transmission rate is one symbol per transmission. When the channel is unknown at the transmitter, this can be achieved using STBC (for $n_T = 2$); and when the channel is known at the transmitter, this can be achieved using beamforming.

2.2.1. Two transmit antennas case

Alamouti scheme

When $n_T = 2$, the elegant Alamouti transmission scheme can be used to achieve full diversity transmission at one symbol per transmission [8]. It transmits two symbols s_1 and s_2 over two consecutive transmissions as follows. During the first transmission, s_1 and s_2 are transmitted simultaneously from antennas 1 and 2, respectively; during the second transmission, $-s_2^*$ and s_1^* are transmitted simultaneously from transmit antennas 1 and 2, respectively. The received signals at receive antenna i corresponding to these two transmissions are given by

$$\begin{bmatrix} y_i(1) \\ y_i(2) \end{bmatrix} = \sqrt{\frac{\rho}{2}} \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix} \begin{bmatrix} h_{i,1} \\ h_{i,2} \end{bmatrix} + \begin{bmatrix} n_i(1) \\ n_i(2) \end{bmatrix}, \quad i = 1, 2, \ldots, n_R. \tag{21}$$

Note that (21) can be rewritten as follows:

$$\underbrace{\begin{bmatrix} y_i(1) \\ y_i(2)^* \end{bmatrix}}_{\mathbf{y}_i} = \sqrt{\frac{\rho}{2}} \underbrace{\begin{bmatrix} h_{i,1} & h_{i,2} \\ h_{i,2}^* & -h_{i,1}^* \end{bmatrix}}_{\tilde{\mathbf{H}}_i} \underbrace{\begin{bmatrix} s_1 \\ s_2 \end{bmatrix}}_{\mathbf{s}} + \underbrace{\begin{bmatrix} n_i(1) \\ n_i(2)^* \end{bmatrix}}_{\mathbf{n}_i}, \quad i = 1, 2, \ldots, n_R, \tag{22}$$

where $\mathbf{n}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}_c(\mathbf{0}, \mathbf{I}_2)$. Note that the channel matrix $\tilde{\mathbf{H}}_i$ is orthogonal, that is, $\tilde{\mathbf{H}}_i^H \tilde{\mathbf{H}}_i = (|h_{i,1}|^2 + |h_{i,2}|^2)\mathbf{I}_2$.

At each receive antenna, the received signal is matched filtered to obtain

$$\mathbf{z}_i = \tilde{\mathbf{H}}_i^H \mathbf{y}_i = \sqrt{\frac{\rho}{2}} \left(|h_{i,1}|^2 + |h_{i,2}|^2 \right) \mathbf{s} + \mathbf{w}_i, \quad i = 1, 2, \ldots, n_R, \tag{23}$$

where $\mathbf{w}_i \sim \mathcal{N}_c(\mathbf{0}, (|h_{i,1}|^2 + |h_{i,2}|^2)\mathbf{I}_2)$. The final decision on $\mathbf{s}$ is then made according to $\hat{\mathbf{s}} = \mathbb{Q}(\mathbf{z})$, where $\mathbb{Q}(\cdot)$ denotes the symbol slicing operation, and

$$\mathbf{z} = \sum_{i=1}^{n_R} \mathbf{z}_i = \sqrt{\frac{\rho}{2}} \left[\sum_{i=1}^{n_R} \left(|h_{i,1}|^2 + |h_{i,2}|^2 \right) \right] \mathbf{s} + \sum_{i=1}^{n_R} \mathbf{w}_i. \tag{24}$$

The received SNR is therefore given by

$$\begin{aligned} \text{SNR}_{\text{Alamouti}} &= \frac{(\rho/2)\left[\sum_{i=1}^{n_R} \left(|h_{i,1}|^2 + |h_{i,2}|^2 \right) \right]^2}{\sum_{i=1}^{n_R} \left(|h_{i,1}|^2 + |h_{i,2}|^2 \right)} \\ &= \frac{\rho}{2} \operatorname{tr}(\mathbf{H}_A^H \mathbf{H}_A) = \frac{\rho}{2} \operatorname{tr}(\mathbf{\Omega}_A) = \rho \left(\frac{\lambda_1 + \lambda_2}{2} \right), \end{aligned} \tag{25}$$

where $\mathbf{\Omega}_A \triangleq \mathbf{H}_A^H \mathbf{H}_A$, λ_1 and λ_2 are the two eigenvalues of $\mathbf{\Omega}_A$, and

$$\mathbf{H}_A = \begin{bmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & h_{2,2} \\ \vdots & \vdots \\ h_{n_R,1} & h_{n_R,2} \end{bmatrix}, \qquad \mathbf{\Omega}_A = \mathbf{H}_A^H \mathbf{H}_A. \tag{26}$$

Beamforming

Beamforming can be referred to as maximum ratio weighting [23], and it is a special case of the linear precoding/decoding discussed in Section 2.1.2, where

$$\begin{aligned} \mathbf{F} &= \sqrt{\rho}\mathbf{v}_1, \\ \mathbf{G} &= \mathbf{v}_1^H, \end{aligned} \tag{27}$$

and $\mathbf{v}_1$ is the eigenvector corresponding to the largest eigenvalue of $\mathbf{\Omega}$. Hence in the beamforming scheme, at each transmission, the transmitter transmits $\mathbf{v}_1 s$ from all transmit antennas, where s is a data symbol. The received signal is given by

$$\mathbf{y} = \mathbf{H}\mathbf{F}s + \mathbf{n} = \sqrt{\rho}\mathbf{H}\mathbf{v}_1 s + \mathbf{n}. \tag{28}$$

At the receiver, a decision on s is made according to $\hat{s} = \mathbb{Q}(u)$, where the decision statistic u is given by $u = \mathbf{v}_1^H \mathbf{H}^H \mathbf{y} = \sqrt{\rho} \underbrace{\mathbf{v}_1^H \mathbf{\Omega} \mathbf{v}_1}_{\lambda_1} s + \underbrace{\mathbf{v}_1^H \mathbf{H}^H \mathbf{n}}_{\mathcal{N}_c(0,\lambda_1)}$. The received SNR in this case is

$$\text{SNR}_{\text{beamforming}} = \rho\lambda_1. \tag{29}$$

Comparing (25) with (29), it is obvious that $\text{SNR}_{\text{beamforming}} \geq \text{SNR}_{\text{Alamouti}}$. Note that in this case, the SNR order indeed translates into the BER order; since in the Alamouti scheme, both symbols have the same SNR, then

$$p_{\text{beamforming}} = Q\left(\gamma\sqrt{\rho\lambda_1} \right) \leq Q\left(\gamma\sqrt{\frac{\rho}{2}(\lambda_1 + \lambda_2)} \right) = p_{\text{Alamouti}}. \tag{30}$$

2.2.2. Four transmit antennas case

One symbol per transmission

It is known that rate-one orthogonal STBC only exists for $n_T = 2$, that is, the Alamouti code. For the case of four transmit antennas ($n_T = 4$), we adopt a rate-one (almost orthogonal) transmission scheme with the following transmission matrix:

$$\mathbf{S} = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 \\ s_2^* & -s_1^* & s_4^* & -s_3^* \\ s_3 & -s_4 & -s_1 & s_2 \\ s_4^* & s_3^* & -s_2^* & -s_1^* \end{bmatrix}. \tag{31}$$

Such a transmission scheme was proposed in [24]. Hence four symbols s_1, s_2, s_3, and s_4 are transmitted across four transmit antennas over four transmissions. The received signals at the ith receive antenna corresponding to these four

transmissions are given by

$$\begin{bmatrix} y_i(1) \\ y_i(2) \\ y_i(3) \\ y_i(4) \end{bmatrix} = \sqrt{\frac{\rho}{4}}\mathbf{S}\begin{bmatrix} h_{i,1} \\ h_{i,2} \\ h_{i,3} \\ h_{i,4} \end{bmatrix} + \begin{bmatrix} n_i(1) \\ n_i(2) \\ n_i(3) \\ n_i(4) \end{bmatrix}, \quad i = 1, 2, \ldots, n_R. \tag{32}$$

Note that (32) can be rewritten as

$$\underbrace{\begin{bmatrix} y_i(1) \\ y_i(2)^* \\ y_i(3) \\ y_i(4)^* \end{bmatrix}}_{\mathbf{y}_i} = \sqrt{\frac{\rho}{4}}\underbrace{\begin{bmatrix} h_{i,1} & h_{i,2} & h_{i,3} & h_{i,4} \\ -h_{i,2}^* & h_{i,1}^* & -h_{i,4}^* & h_{i,3}^* \\ -h_{i,3} & h_{i,4} & h_{i,1} & -h_{i,2} \\ -h_{i,4}^* & -h_{i,3}^* & h_{i,2}^* & h_{i,1}^* \end{bmatrix}}_{\tilde{\mathbf{H}}_i}\underbrace{\begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{bmatrix}}_{\mathbf{s}} + \underbrace{\begin{bmatrix} n_i(1) \\ n_i(2)^* \\ n_i(3) \\ n_i(4)^* \end{bmatrix}}_{\mathbf{v}_i}, \quad i = 1, 2, \ldots, n_R. \tag{33}$$

The matched-filter output at the ith receive antenna is given by

$$\mathbf{z}_i = \tilde{\mathbf{H}}_i^H \mathbf{y}_i = \sqrt{\frac{\rho}{4}}\tilde{\mathbf{\Omega}}_i \mathbf{s} + \mathbf{w}_i, \tag{34}$$

where

$$\tilde{\mathbf{\Omega}}_i = \tilde{\mathbf{H}}_i^H \tilde{\mathbf{H}}_i = \begin{bmatrix} \gamma_i & 0 & \alpha_i & 0 \\ 0 & \gamma_i & 0 & -\alpha_i \\ -\alpha_i & 0 & \gamma_i & 0 \\ 0 & \alpha_i & 0 & \gamma_i \end{bmatrix}, \tag{35}$$

$\gamma_i = \sum_{j=1}^{n_T} |h_{i,j}|^2$, $\alpha_i = 2\jmath\Im(h_{i,1}^* h_{i,3} + h_{i,4}^* h_{i,2})$, and $\mathbf{w}_i = \tilde{\mathbf{H}}_i^H \mathbf{n}_i \sim \mathcal{N}_c(\mathbf{0}, \tilde{\mathbf{\Omega}}_i)$. By grouping the entries of $\mathbf{z}_i$ into two pairs, we can write

$$\begin{aligned} \underbrace{\begin{bmatrix} z_i(1) \\ z_i(3) \end{bmatrix}}_{\mathbf{z}_{i,1}} &= \sqrt{\frac{\rho}{4}}\mathbf{\Gamma}_i \underbrace{\begin{bmatrix} s_1 \\ s_3 \end{bmatrix}}_{\mathbf{s}_1} + \underbrace{\begin{bmatrix} w_i(1) \\ w_i(3) \end{bmatrix}}_{\mathbf{w}_{i,1}}, \\ \underbrace{\begin{bmatrix} z_i(4) \\ z_i(2) \end{bmatrix}}_{\mathbf{z}_{i,2}} &= \sqrt{\frac{\rho}{4}}\mathbf{\Gamma}_i \underbrace{\begin{bmatrix} s_4 \\ s_2 \end{bmatrix}}_{\mathbf{s}_2} + \underbrace{\begin{bmatrix} w_i(4) \\ w_i(2) \end{bmatrix}}_{\mathbf{w}_{i,2}}, \end{aligned} \tag{36}$$

where $\mathbf{\Gamma}_i = \begin{bmatrix} \gamma_i & \alpha_i \\ -\alpha_i & \gamma_i \end{bmatrix}$ and $\mathbf{w}_{i,\ell} \sim \mathcal{N}_c(\mathbf{0}, \mathbf{\Gamma}_i)$, $\ell = 1, 2$. Note that $\mathbf{\Gamma}_i^H = \mathbf{\Gamma}_i$. Note also that (36) are effectively 2×2 BLAST systems and they can be decoded using either linear detection or ML detection. For example, the linear decision rule is given by $\hat{\mathbf{s}}_\ell = \mathbb{Q}[\sum_{i=1}^{n_R} \mathbf{G}_{i,\ell}\mathbf{z}_{i,\ell}]$, $\ell = 1, 2$, where the linear detector can be either a zero-forcing detector, that is, $\mathbf{G}_{i,\ell} = \mathbf{\Gamma}_i^{-1}$, or an MMSE detector, that is, $\mathbf{G}_{i,\ell} = (\mathbf{\Gamma}_i + (4/\rho)\mathbf{I}_2)^{-1}$. On the other hand, the ML detection rule is given by

$$\begin{aligned} \hat{\mathbf{s}}_\ell &= \min_{\mathbf{s} \in \mathcal{A}^2} \sum_{i=1}^{n_R} \left(\mathbf{z}_{i,\ell} - \sqrt{\frac{\rho}{4}}\mathbf{\Gamma}_i \mathbf{s}\right)^H \mathbf{\Gamma}_i^{-1} \left(\mathbf{z}_{i,\ell} - \sqrt{\frac{\rho}{4}}\mathbf{\Gamma}_i \mathbf{s}\right) \\ &= \max_{\mathbf{s} \in \mathcal{A}^2} \left[\Re\left\{\mathbf{s}^H \sum_{i=1}^{n_R} \mathbf{z}_{i,\ell}\right\} - \sqrt{\frac{\rho}{4}}\mathbf{s}^H \left(\sum_{i=1}^{n_R} \mathbf{\Gamma}_i\right)\mathbf{s}\right], \quad \ell = 1, 2. \end{aligned} \tag{37}$$

When the channel state is known at the transmitter, the optimal transmission method to achieve one symbol per transmission is the beamforming scheme described by (27), (28), and (29).

Note that the received SNR of the above block coding scheme with linear zero-forcing detector is given by

$$\overline{\mathrm{SNR}} = \frac{\rho}{4} \cdot \frac{n_R^2}{\sum_{i=1}^{n_R} [\mathbf{\Gamma}_i^{-1}]_{1,1}}, \tag{38}$$

whereas the SNR of the beamforming scheme is given by $\overline{\mathrm{SNR}}_{\text{beamforming}} = \rho\lambda_1$.

Two symbols per transmission

Now suppose that a rate of two symbols per transmission is desired using four transmit antennas. When the channel is unknown at the transmitter, we can use one pair of the transmit antennas to transmit $\mathbf{s}_1 = [s_1 \;\; s_2]^T$ using the Alamouti scheme, and use the other pair to transmit $\mathbf{s}_2 = [s_3 \;\; s_4]^T$ also using Alamouti scheme. In this way, we transmit four symbols over two transmissions. At the ith receive antenna, the received signal $\mathbf{y}_i = [y_i(1) \;\; y_i(2)]^T$ corresponding to the two transmissions is given by

$$\mathbf{y}_i = \sqrt{\frac{\rho}{2}}\tilde{\mathbf{H}}_{i,1}\mathbf{s}_1 + \sqrt{\frac{\rho}{2}}\tilde{\mathbf{H}}_{i,2}\mathbf{s}_2, + \mathbf{n}, \quad i = 1, 2, \ldots, n_R, \tag{39}$$

where $\tilde{\mathbf{H}}_{i,1} = \begin{bmatrix} h_{i,1} & h_{i,2} \\ h_{i,2}^* & -h_{i,1}^* \end{bmatrix}$ and $\tilde{\mathbf{H}}_{i,2} = \begin{bmatrix} h_{i,3} & h_{i,4} \\ h_{i,4}^* & -h_{i,3}^* \end{bmatrix}$. Therefore, we have

$$\underbrace{\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_{n_R} \end{bmatrix}}_{\mathbf{y}} = \sqrt{\frac{\rho}{2}}\underbrace{\begin{bmatrix} \tilde{\mathbf{H}}_{1,1} & \tilde{\mathbf{H}}_{1,2} \\ \tilde{\mathbf{H}}_{2,1} & \tilde{\mathbf{H}}_{2,2} \\ \vdots & \vdots \\ \tilde{\mathbf{H}}_{n_R,1} & \tilde{\mathbf{H}}_{n_R,2} \end{bmatrix}}_{\tilde{\mathbf{H}}}\underbrace{\begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{bmatrix}}_{\mathbf{s}} + \mathbf{n}. \tag{40}$$

The received signal $\mathbf{y}$ is first matched filtered to obtain

$$\mathbf{z} = \tilde{\mathbf{H}}^H \mathbf{y} = \sqrt{\frac{\rho}{2}}\tilde{\mathbf{H}}^H \tilde{\mathbf{H}}\mathbf{s} + \tilde{\mathbf{H}}^H \mathbf{n}. \tag{41}$$

Denote

$$\tilde{\mathbf{\Omega}} \triangleq \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} = n_R \cdot \begin{bmatrix} \mathbf{I}_2 & \dfrac{1}{n_R}\displaystyle\sum_{j=1}^{n_R} \tilde{\mathbf{H}}_{j,1}^H \tilde{\mathbf{H}}_{j,2} \\ \dfrac{1}{n_R}\displaystyle\sum_{j=1}^{n_R} \tilde{\mathbf{H}}_{j,1}^H \tilde{\mathbf{H}}_{j,2} & \mathbf{I}_2 \end{bmatrix}. \tag{42}$$

Then the output of the whitening filter is given by $\mathbf{u} = \tilde{\mathbf{\Omega}}^{-1/2}\mathbf{z} = \sqrt{\rho/2}\tilde{\mathbf{\Omega}}^{1/2}\mathbf{s} + \mathbf{w}$, where $\mathbf{w} \sim \mathcal{N}_c(\mathbf{0}, \mathbf{I}_4)$. Now we can use any of the aforementioned BLAST decoding methods to decode $\mathbf{s}$.

When the channel is known at the transmitter, linear precoding/decoding can be used to transmit two symbols per transmission. For example, the equal-error precoding scheme is specified by (8) and (9) with $m = 2$. The received SNR of this method is given by $\overline{\mathrm{SNR}}_{\text{equal-error precoding}} =$

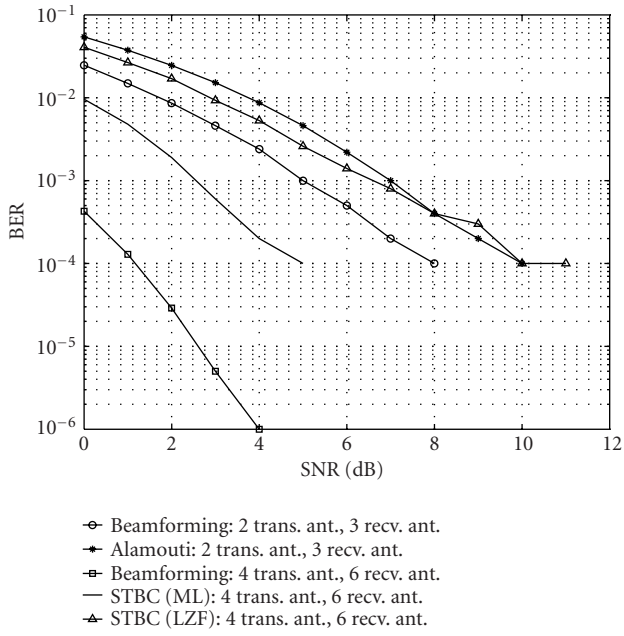


Figure 2: Comparisons of the BER performances among the MIMO techniques for one symbol/transmission: beamforming versus Alamouti with $n_T = 2$ and $n_R = 3$; beamforming versus rate-one STBC with $n_T = 4$ and $n_R = 6$.

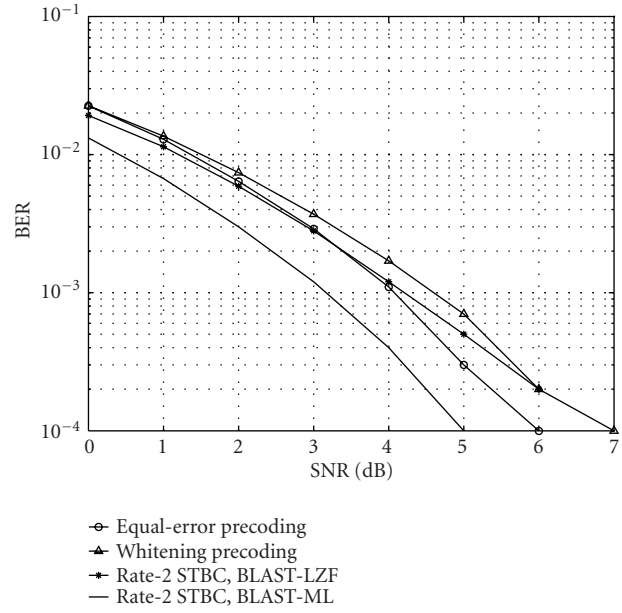


Figure 3: Comparisons of the BER performances between the linear precoding/decoding strategies and the rate-two STBC: $n_T = 4$ and $n_R = 6$; the rate is two symbols/transmission.

$\rho/(\lambda_1^{-1} + \lambda_2^{-1})$. The whitening precoding method, on the other hand, is specified by $\mathbf{F} = \sqrt{\rho/2}[\mathbf{v}_1\mathbf{v}_2]$ and $\mathbf{G} = [\mathbf{v}_1\mathbf{v}_2]^H$; and the average received SNR of this method is given by $\overline{\text{SNR}}_{\text{whitening precoding}} = \rho((\lambda_1 + \lambda_2)/4)$. Note that λ_1 and λ_2 are the two largest eigenvalues contained in $\mathbf{\Lambda}$.

2.2.3. Comparisons

Figure 2 shows the performance comparisons among the MIMO techniques to achieve one symbol per transmission. Specifically, the beamforming scheme is compared with the Alamouti code for a system with two transmit antennas, and the beamforming scheme is compared with the rate-one STBC for a system with four transmit antennas. It is observed from Figure 2 that the beamforming scheme achieves about 2 dB gain over the Alamouti code, and similarly, the beamforming can achieve much better performance than the rate-one STBC strategy.

Figure 3 shows the performance comparisons between the linear precoding/decoding schemes and the rate-two STBC strategy for a system with $n_T = 4$ and $n_R = 6$ to achieve two symbols per transmission. It is seen from Figure 3 that the rate-two STBC achieves a better performance than the linear precoding/decoding schemes, and the performance gap is not so large. In particular, the rate-two STBC with BLAST-LZF decoding has an approximate performance to the equal-error precoding scheme.

It is observed from Figures 1 and 3 that although the linear precoding/decoding schemes exploit the channel knowledge at the transmitter, they may not have performance gains compared to those MIMO techniques without channel knowledge requirement at the transmitter. And this phenomenon is evident especially in the high-data rate transmission scenario, that is, BLAST versus linear precoding/decoding schemes with $n_T = 4$. This can be explained as follows. Note that, for the linear precoding/decoding strategies discussed above, the adaptive modulation is not employed, and thus, the performance gain is limited for the fixed modulation.

3. WCDMA DOWNLINK SYSTEMS

In this section, a WCDMA downlink system based on the 3GPP standard, a subspace tracking algorithm, as well as a quantized feedback approach are specified. In Section 3.1, we describe the WCDMA system, including the structures of the transmitter and the receiver, the channelization and scrambling codes, the frame structures of the data and the pilot channels, the multipath fading channel model, as well as the channel estimation algorithm. In Section 3.2, we detail the subspace tracking method and the quantized feedback scheme.

3.1. System description

3.1.1. Transmitter and receiver structures

The system model of the downlink WCDMA system is shown in Figure 4. The left part of Figure 4 is the transmitter structure. The data sequences of the users are first spread by unique orthogonal variable spreading factor (OVSF) codes ($C_{ch,SF,1}$, $C_{ch,SF,2}, \ldots$), and then, the spread chip sequences of different users are multiplied by downlink scrambling codes ($C_{cs,1}, C_{cs,2}, \ldots$). After summing up the scrambled data

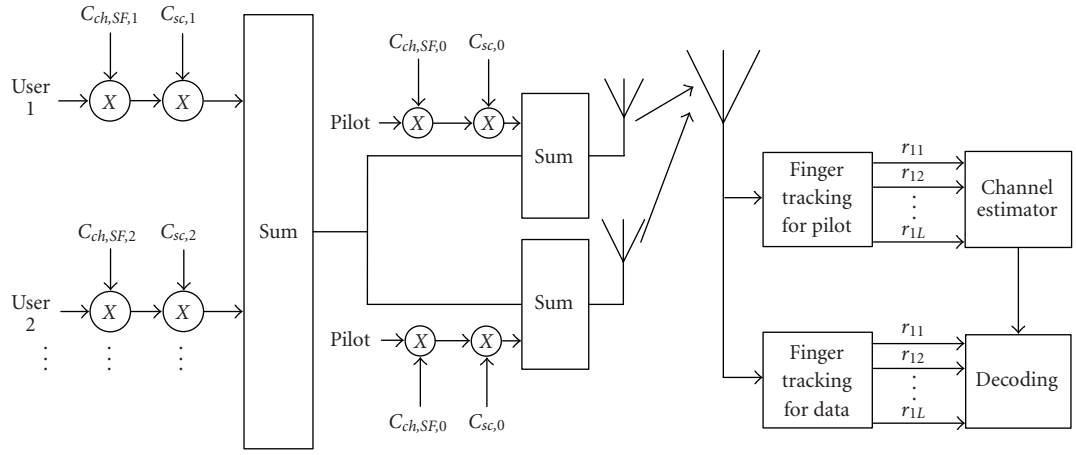


Figure 4: Transmitter and receiver structures of the downlink WCDMA system.

sequences from different users, the data sequences are combined with the pilot sequence, which is also spread and scrambled by the codes $(C_{ch,SF,0}, C_{cs,0})$ for the pilot channel sent to each antenna. The specifications of OVSF and scrambling codes can be referred to [15]. The right part of Figure 4 shows the receiver structure of this system with one receive antenna. We assume the number of multipaths in the WCDMA channel is L. Each receive antenna is followed by a bank of RAKE fingers. Each finger tracks the corresponding multipath component for the receiver antenna and performs descrambling and despreading for each of the L multipath components. Such a receiver structure is similar to the conventional RAKE receiver but without maximal ratio combining (MRC). Hence, there are L outputs for each receive antenna, and thus, each of the L antenna outputs can be viewed as a *virtual* receive antenna [14]. With the received pilot signals, the downlink channel is estimated accordingly. This channel estimate is provided to the detector to perform demodulation of the received users' signals.

It is shown in [14] that the above receiver scheme with *virtual* antennas essentially provides an interface between MIMO techniques and a WCDMA system. The outputs of the RAKE fingers are sent to a MIMO demodulator that operates at the symbol rate. The equivalent symbol-rate MIMO channel response matrix is given by

$$\mathbf{H} = \begin{bmatrix} h_{1,1,1} & h_{1,1,2} & \dots & h_{1,1,n_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_{1,L,1} & h_{1,L,2} & \dots & h_{1,L,n_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_R,1,1} & h_{n_R,1,2} & \dots & h_{n_R,1,n_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_R,L,1} & h_{n_R,L,2} & \dots & h_{n_R,L,n_T} \end{bmatrix}, \tag{43}$$

where $h_{i,l,j}$ denotes the complex channel gain between the jth transmit antenna and the lth finger of the ith receive antenna. Hence (43) is equivalent to a MIMO system with n_T transmit antennas and $(n_R \cdot L)$ receive antennas [14].

3.1.2. Multipath fading channel model and channel estimation

Each user's channel contains four paths, that is, $L = 4$. The channel multipath profile is chosen according to the 3GPP specifications. That is, the relative path delays are 0, 260, 521, and 781 nanoseconds, and the relative path power gains are $0, -3, -6$, and -9 dB, respectively.

There are two channels in the system, namely, common control physical channel (CCPCH) and common pilot channel (CPICH), whose rates are variable and fixed, respectively. For more details, see [15]. The CPICH is transmitted from all antennas using the same channelization and the scrambling code, and the different pilot symbol sequences are adopted on different antennas. Note that in the system, the pilot signal can be treated as the data of a special user. In other words, the pilot and the data of different users in the system are combined with code duplexing but not time duplexing.

Here we use orthogonal training sequences of length $T \geq n_T$ based on the Hadamard matrix to minimize the estimation error [25]. Note that, although the channel varies at the symbol rate, the channel estimator assumes it is fixed over at least n_T symbol intervals.

3.2. Subspace tracking with quantized feedback for beamforming

3.2.1. Tracking of the channel subspace

Recall that in the beamforming and general precoding transmission schemes, the value of the MIMO channel $\mathbf{H}$ has to be provided to the transmitter. Typically, in FDD systems, this can be done by feeding back to the transmitter the estimated channel value $\hat{\mathbf{H}}$. However, the feedback channel usually has a very low data rate. Here we propose to employ a subspace tracking algorithm, namely, projection approximation subspace tracking with deflation (PASTd) [20], with quantized feedback to track the MIMO eigen channels. Figure 5 shows the diagram of the MIMO system adopting a subspace tracking and the quantized feedback approach. In particular, the receiver employs the channel estimator to obtain the estimate of the channel $\hat{\mathbf{H}}$ and subsequently, PASTd algorithm

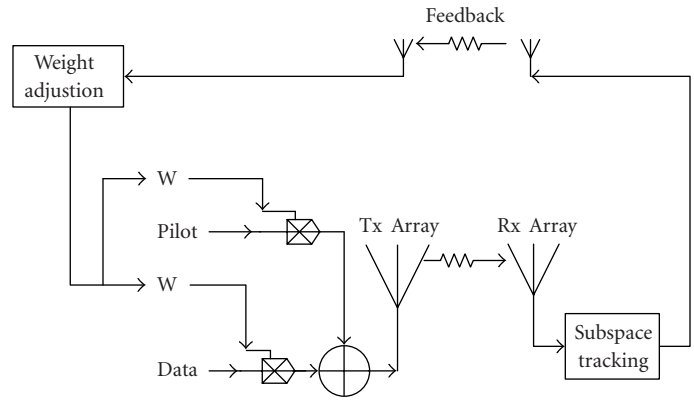


Figure 5: The MIMO linear precoding/decoding system with subspace tracking and quantized feedback schemes.

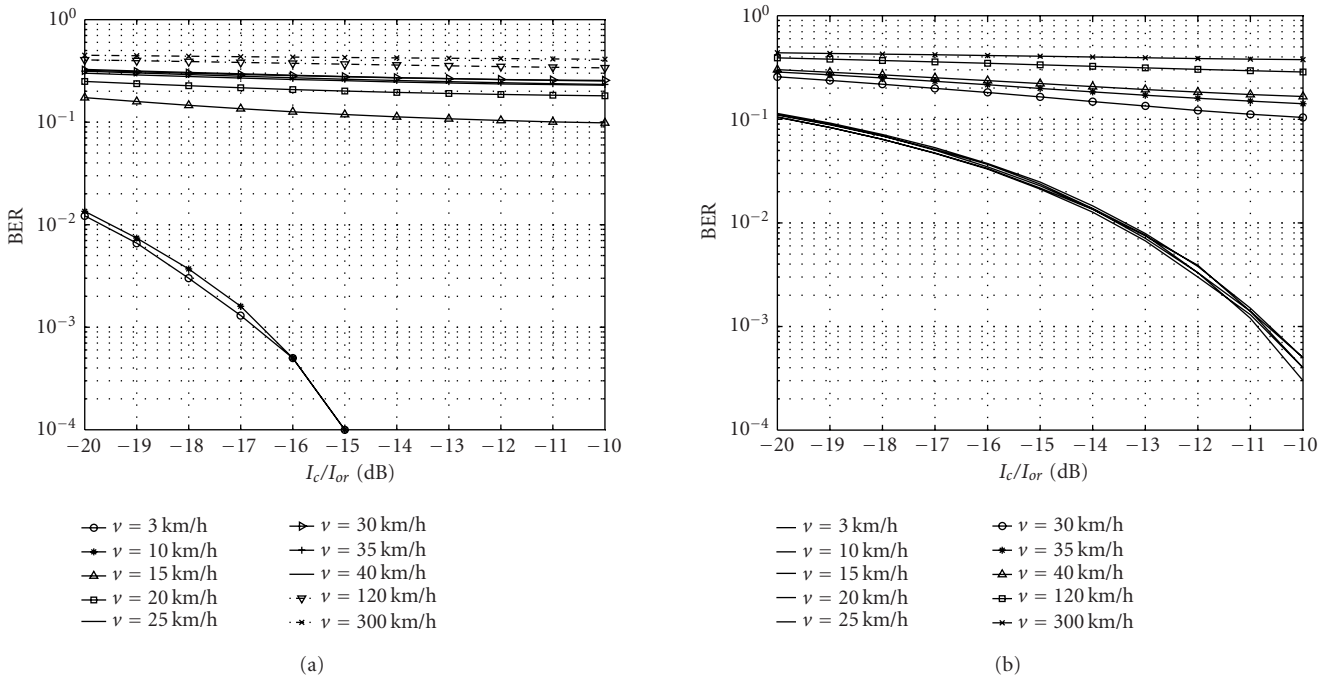


Figure 6: BER performance of beamforming under different doppler frequencies: (a) $n_T = 4$, $n_R = 1$ (beamforming, perfect known channel, lossless feedback (2 frames)), (b) $n_T = 2$, $n_R = 1$ (perfectly known channel, lossless feedback (1 frame)).

is adopted to get $\mathbf{F} = \mathbf{V} = [\mathbf{V}_1, \ldots, \mathbf{V}_m]$, which contains the principal eigenvectors of $\mathbf{\Omega} = \mathbf{H}^H\mathbf{H}$.

3.2.2. Frame-based feedback

Note that, for the uplink channel in the 3GPP standard [21], the bit rate is 1500 bits per second (bps), the frame rate is 100 frames per second (fps), and thus, there are fifteen bits in each uplink frame. On the other hand, the downlink WCDMA channel is a symbol-by-symbol varied channel. Thereby, it is necessary to consider an effective and efficient quantization and feed back scheme, so as to feed back $\mathbf{F}$ to the transmitter via the band-limited uplink channel.

For the beamforming scheme, we employ the feedback approach as follows. The average eigenvector of the channel over one frame or two frames is fed back instead of the eigenvectors of each symbol or slot duration. Note that such feedback approach assumes the downlink WCDMA channel as a block fading one, and actually, it is effective and efficient under low doppler frequencies. Figure 6 shows the BER performances of the MIMO system employing the beamforming scheme under different doppler frequencies. In Figure 6b, two transmit antennas are adopted, and the average eigenvectors over one frame duration are losslessly fed back. That is, the eigenvector information is precisely fed back without

Table 1: Frame structures for quantized feedback. Case 1: two transmit antennas and one receive antenna, (5, 5) quantization : 5 bits for the absolute value component and 5 bits for the phase component of each vector element; A_{ij}: jth bit for the absolute value of ith vector element; P_{ij}: jth bit for the phase of ith vector element. Case 2: two transmit antennas and 1 receive antenna, (4,7) quantization. Case 3: four transmit antennas and 1 receive antenna, (3,6) quantization.

Case 1

Slot	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Bits	A_{11}	A_{12}	A_{13}	A_{14}	A_{15}	A_{21}	A_{22}	A_{23}	A_{24}	A_{25}	P_{21}	P_{22}	P_{23}	P_{24}	P_{25}

Case 2

Slot	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Bits	A_{11}	A_{12}	A_{13}	A_{14}	A_{21}	A_{22}	A_{23}	A_{24}	P_{21}	P_{22}	P_{23}	P_{24}	P_{25}	P_{26}	P_{27}

Case 3

Slot	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Bits	A_{11}	A_{12}	A_{13}	A_{21}	A_{22}	A_{23}	P_{21}	P_{22}	P_{23}	P_{24}	P_{25}	P_{26}	A_{31}	A_{32}	A_{33}
Slot	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Bits	P_{31}	P_{32}	P_{33}	P_{34}	P_{35}	P_{36}	A_{41}	A_{42}	A_{43}	P_{41}	P_{42}	P_{43}	P_{44}	P_{45}	P_{46}

quantization. It is seen that the system achieves a good performance for the speeds lower than 30 km/h, and the BER curves are shown as "floors" when v is higher than 30 km/h. The appearance of such "floor" is due to the severe mismatch between the precoding and the downlink channel. Similarly, Figure 6a gives the BER performances of the system employing the beamforming with four transmit antennas, where the average eigenvectors over two frames are losslessly fed back. It is seen that the BER performances degrade to "floors" for the speeds higher than 15 km/h. It is observed from (6) that the frame-based feedback approach is feasible for the beamforming system under the low-speed cases. In particular, it is feasible for the system employing two transmit antennas and four transmit antennas, under the cases of $v \leq 25$ km/h and $v \leq 10$ km/h, respectively.

3.2.3. Quantization of the feedback

Table 1 shows the feedback frame structures for the MIMO system employing beamforming schemes, that is, the quantization of the elements of the eigenvector to be fed back. We consider three cases here. Case 1 and Case 2 are contrived for the beamforming system with two transmit antennas. These two bit allocation strategies of one feedback frame are, namely, (5, 5) and (4, 7) quantized feedback, respectively. In particular, (5, 5) quantized feedback allocates 5 bits each to the absolute value and the phase component of one eigenvector element; and (4, 7) quantized feedback allocates 4 bits and 7 bits to the absolute value and the phase component of one eigenvector element, respectively. Case 3, namely, (3, 6) quantized feedback, is contrived for the beamforming system with four transmit antennas. Two feedback frames are allocated for the average eigenvector over two frames. Note that relatively more bits should be allocated to the phase component, since the error caused by quantization is more sensitive to the preciseness of the phase components than that of the absolute value components moreover, our simulations show that the (5, 5) and (4, 7) quantized feedback approaches actually have very approximated performances.

4. SIMULATION RESULTS FOR WCDMA SYSTEMS

In the simulations, we adopt one receive antenna ($n_R = 1$), which is a realistic scenario for the WCDMA downlink receiver. For the multipath fading channel in the WCDMA system, the number of multipath is assumed to be four ($L = 4$), and the mobile speed is assumed to be three kilometers per hour ($v = 3$ km/h). QPSK is used as the modulation format. The performance metric is BER versus signal-to-interference-ratio (I_c/I_{or}). I_c/I_{or} is the power ratio between the signal of the desired user and the interference from all other simultaneous users in the WCDMA system. Subsequently, several cases with different transmission rates over two and four transmit antennas are studied.

BLAST versus linear precoding

Figure 7 shows the performance comparisons between the BLAST and the linear precoding/decoding schemes for a rate of four symbols per transmission over four transmit antennas ($n_T = 2$). In particular, the channel estimator given in Section 3.1.2 is adopted to acquire the channel knowledge. For the linear precoding/decoding schemes, lossless feedback is assumed. It is seen from Figure 7 that the BLAST scheme with ML detection achieves the best BER performance over all linear precoding/decoding schemes. Note that the reason that precoding does not offer performance advantage here is that we require the rate for different eigen channels to be the same, that is, no adaptive modulation scheme is allowed. Hence we conclude that to achieve high throughput, it suffices to employ the BLAST architecture and the knowledge of the channel at the transmitter offers no advantage.

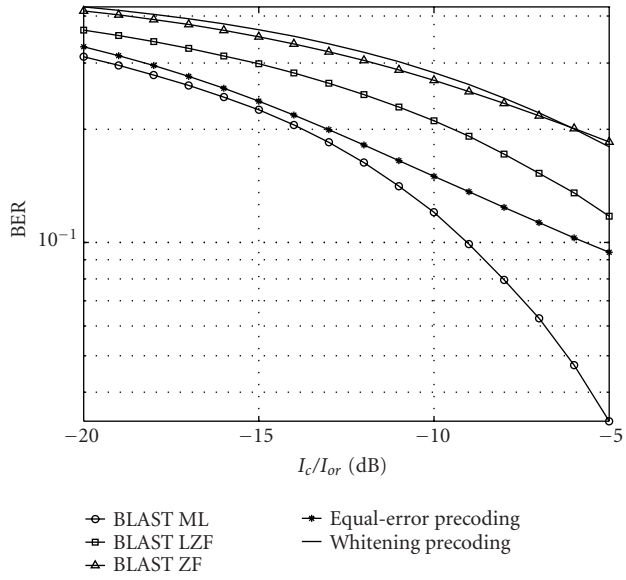


Figure 7: BER comparisons between the BLAST and the transmit precoding schemes: $n_T = 4$ and $n_R = 1$; four QPSK symbols/transmission; $v = 3\,km/h$, $L = 4$.

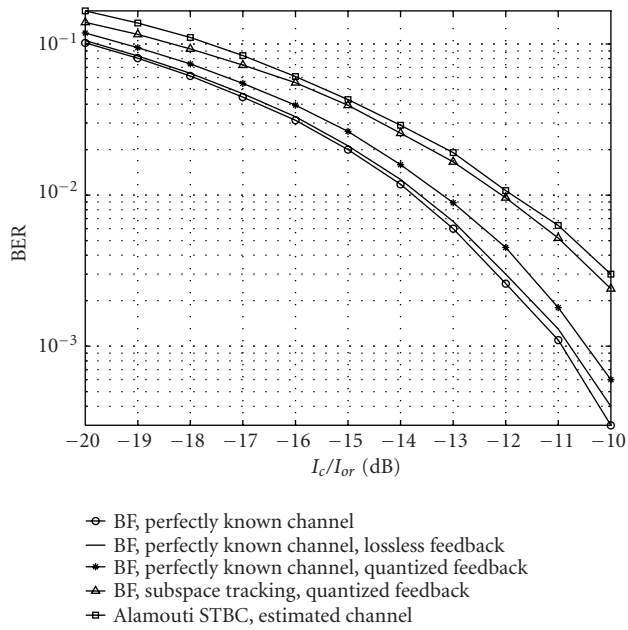


Figure 8: BER comparisons between Alamouti and beamforming with subspace tracking and quantized feedback schemes: $n_T = 2$ and $n_R = 1$; one QPSK symbol/transmission; $v = 3$ km/h; $(4, 7)$ quantized feedback.

STBC versus beamforming

Figure 8 gives the performance comparisons between the Alamouti STBC and the beamforming schemes for a rate of one symbol per transmission over two transmit antennas ($n_T = 2$). The effects of the quantized feedback approach is also shown in Figure 8. In particular, the cycled line is the BER performance when perfect channel knowledge is available at both the transmitter and the receiver. The solid line is the performance when perfect channel knowledge is available at the receiver and the frame-based feedback without quantization in Section 3.2.2 is adopted. It is seen that the frame-based feedback approach only causes very trivial performance degradation. The asteriated line is the performance when perfect channel knowledge is available at the receiver and the frame-based feedback with $(4, 7)$ quantized feedback approach in Section 3.2.3 is adopted. It is seen that the quantization of the feedback only generates about 0.5 dB performance loss. Moreover, the triangled line is the performance when the channel estimator in Section 3.1.2, the subspace tracking in Section 3.2.1, and the $(4, 7)$ quantized feedback approach are adopted. It is shown that the subspace tracking and the channel estimation cause about 1 to 1.5 dB performance degradation. Finally, the squared line is the performance of the Alamouti STBC, where the channel estimator is adopted at the receiver. It is observed from Figure 8 that the WCDMA system employing beamforming can have a better performance than that employing the Alamouti STBC scheme, though the performance gain is not very evident.

Figure 9 gives the comparison between the beamforming scheme and the rate-one STBC strategy discussed in Section 2.2.2 for a rate of one symbol per transmission over four transmit antennas ($n_T = 4$). Similarly, perfectly known channel knowledge, estimated channel knowledge, lossless feedback, and quantized feedback cases are shown. From bottom up, the first curve is the result of the beamforming scheme with perfectly known channel knowledge at both

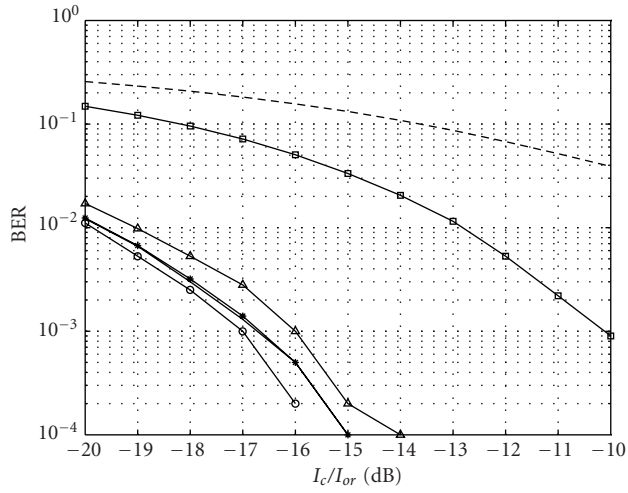


Figure 9: BER comparisons between rate-one STBC and beamforming with subspace tracking and quantized feedback schemes: $n_T = 4$ and $n_R = 1$; one QPSK symbol/transmission; $v = 3$ km/h; $(3, 6)$ quantized feedback.

Table 2: Summary of the performance comparisons of the MIMO techniques.

(a)

High-rate transmission	MIMO techniques	Channel information	BER performance
Four symbols/transmission over $n_T = 4$	BLAST	Receiver	Better
	Transmit precoding	Transmitter/receiver	Worse

(b)

Diversity transmission	MIMO techniques	channel Information	BER performance
One symbol/transmission over $n_T = 2$	Beamforming	Transmitter/receiver	Better
	Alamouti	Receiver	Worse
One symbol/transmission over $n_T = 4$	Beamforming	Transmitter/receiver	Better
	Rate-one STBC	Receiver	Worse
Two symbols/transmission over $n_T = 4$	Transmit precoding	Transmitter/receiver	Worse
	Rate-two STBC	Receiver	Better

the transmitter and the receiver; the second curve is the result of the beamforming scheme with perfectly known channel knowledge at the receiver and the frame-based feedback without quantization; the third curve is the result of the beamforming scheme with perfectly known channel knowledge at the receiver and the frame-based feedback with $(3, 6)$ quantization; the fourth curve is the result of channel estimator, subspace tracker, and the frame-based feedback with $(3, 6)$ quantization; the top two curves are the results of the rate-one STBC scheme with different detection methods. It is observed from Figure 9 that the beamforming can achieve a much better performance than the STBC for the case of four transmit antennas.

Moreover, it is also well confirmed that the subspace tracking algorithm discussed in Section 3.2.1, the frame-based feedback in Section 3.2.2, as well as the quantization approach discussed in Section 3.2.3 offer a practical way of realizing beamforming in MIMO WCDMA systems.

5. CONCLUSIONS

In this paper, we have analyzed and compared the performance of three MIMO techniques, namely, BLAST, STBC and linear precoding/decoding, and considered their applications in WCDMA downlink systems. For a certain transmission rate, we compared the different scenarios with different transmit antennas both analytically in terms of the average receiver SNR, as well as through simulations in terms of the BER performance. To cope with the channel feedback in WCDMA systems for beamforming, we adopted a subspace tracking method with a quantized feedback approach to make the principle eigenspace of the MIMO channel available to the transmitter.

Some instructive conclusions are drawn in this study. On the one hand, the optimal BLAST scheme can achieve the best performance in the high-rate transmission scenario, although with channel knowledge available at the transmitters, no performance gain is achievable by the linear precoding/decoding schemes without employing adaptive modulation. On the other hand, the beamforming scheme achieves better performances than the STBC schemes in the diversity transmission scenario. Table 2 gives a summary of the performance comparisons of the MIMO techniques in different scenarios. Moreover, it is well confirmed the effectiveness and feasibility of the combination of the subspace tracking algorithm and the quantized feedback approach for beamforming transmission in the MIMO WCDMA system. Finally, we note that in this paper, we only consider the *linear* precoding scheme. Significant performance improvement is expected when nonlinear precoder (e.g., adaptive modulation and bit loading) is employed [26, 27, 28].

ACKNOWLEDGMENT

This work was supported in part by the U.S. National Science Foundation (NSF) under Grants CCR-0225721 and CCR-0225826.

REFERENCES

[1] C.-N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Transactions on Information Theory*, vol. 48, no. 3, pp. 637–650, 2002.

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communication in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–355, 1998.

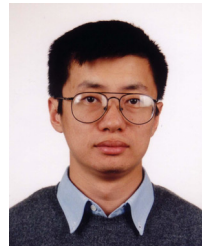
[3] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Communications*, vol. 48, no. 3, pp. 502–513, 2000.

[4] O. Damen, A. Chkeif, and J.-C. Belfiore, "Lattice code decoder for space-time codes," *IEEE Communications Letters*, vol. 4, no. 5, pp. 161–163, 2000.

[5] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[6] G. D. Golden, C. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electronics Letters*, vol. 35, no. 1, pp. 14–16, 1999.
[7] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Transactions on Communications*, vol. 49, no. 12, pp. 2198–2206, 2001.
[8] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.
[9] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.
[10] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: performance results," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 3, pp. 451–460, 1999.
[11] Z. Hong, K. Liu, R. W. Heath, and A. M. Sayeed Jr., "Spatial multiplexing in correlated fading via the virtual channel representation," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 856–866, 2003.
[12] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073–1096, 2003.
[13] K. Majonen and M. J. Heikkila, "Higher data rates with space-time block codes and puncturing in WCDMA systems," in *Proc. IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC '01)*, pp. 36–40, San Diego, Calif, USA, October 2001.
[14] D. Samardzija, P. Wolniansky, and J. Ling, "Performance evaluation of the VBLAST algorithm in W-CDMA systems," in *IEEE Vehicular Technology Conference (VTC '01)*, vol. 2, pp. 723–727, Atlantic City, NJ, USA, October 2001.
[15] Texas Instruments, "Space-time block coded transmit antenna diversity for WCDMA," proposal TDOC# 662/98 to ETSI SMG2 UMTS standards, December 1998.
[16] Lucent Technologies, "Downlink diversity improvements through space-time spreading," proposal 3GPP2-C30-19990817-014 to the CDMA-2000 standard, August 1999.
[17] B. C. Banister and J. R. Zeidler, "Tracking performance of a gradient sign algorithm for transmit antenna adaptation with feedback," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '01)*, Salt Lake City, Utah, USA, May 2001.
[18] B. C. Banister and J. R. Zeidler, "Transmission subspace tracking for MIMO communications systems," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM '01)*, vol. 1, pp. 161–165, San Antonio, Tex, USA, November 2001.
[19] W. Utschick, "Tracking of signal subspace projectors," *IEEE Transactions on Signal Processing*, vol. 50, no. 4, pp. 769–778, 2002.
[20] X. Wang and H. V. Poor, "Blind multiuser detection: a subspace approach," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 677–690, 1998.
[21] Third Generation Partnership Project, "Tx diversity solutions for multiple antennas (release 5)," proposal 3G TR25.869 V0.1.01 to Technical Specification Group Radio Access Network, November 2001.
[22] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 11, pp. 1841–1852, 1999.
[23] B. Hochwald, T. L. Marzetta, and C. B. Papadias, "A transmitter diversity scheme for wideband CDMA based on space-time spreading," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 1, pp. 48–60, 2001.
[24] C. B. Papadias and G. J. Foschini, "A space-time coding approach for systems employing four transmit antennas," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '01)*, pp. 2481–2484, Salt Lake City, Utah, USA, May 2001.
[25] T. L. Marzetta, "Blast training: estimating channel characteristics for high capacity space-time wireless," in *Proc. 37th Annual Allerton Conference on Communication, Control, and Computing*, pp. 958–966, Monticello, Ill, USA, September 1999.
[26] P. S. Chow, J. M. Cioffi, and J. A. C. Bingham, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Trans. Communications*, vol. 43, no. 2, pp. 773–775, 1995.
[27] R. F. H. Fischer and J. B. Huber, "A new loading algorithm for discrete multitone transmission," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM '96)*, pp. 724–728, London, England, November 1996.
[28] B. S. Krongold, K. Ramchandran, and D. L. Jones, "Computationally efficient optimal power allocation algorithms for multicarrier communication systems," *IEEE Trans. Communications*, vol. 48, no. 1, pp. 23–27, 2000.

Chuxiang Li received the B.S. and M.S. degrees from the Department of Electronics Engineering, Tsinghua University, Beijing, China, in 1999 and 2002, respectively. He is currently working toward the Ph.D. degree in the Department of Electrical Engineering, Columbia University, New York, NY. His research interests fall in the area of wireless communications and statistical signal processing.



Xiaodong Wang received the B.S. degree in electrical engineering and applied mathematics (with the highest honor) from Shanghai Jiao Tong University, Shanghai, China, in 1992; the M.S. degree in electrical and computer engineering from Purdue University in 1995; and the Ph.D. degree in electrical engineering from Princeton University in 1998. From July 1998 to December 2001, he was an Assistant Professor in the Department of Electrical Engineering, Texas A&M University. In January 2002, he joined the Department of Electrical Engineering, Columbia University, as an Assistant Professor. Dr. Wang's research interests fall in the general areas of computing, signal processing, and communications. He has worked in the areas of digital communications, digital signal processing, parallel and distributed computing, nanoelectronics, and bioinformatics, and has published extensively in these areas. His current research interests include wireless communications, Monte-Carlo-based statistical signal processing, and genomic signal processing. Dr. Wang received the 1999 National Science Foundation (NSF) Career Award, and the 2001 IEEE Communications Society and Information Theory Society Joint Paper Award. He currently serves as an Associate Editor for the IEEE Transactions on Communications, the IEEE Transactions on Wireless Communications, the IEEE Transactions on Signal Processing, and the IEEE Transactions on Information Theory.

EURASIP Journal on Applied Signal Processing 2004:5, 662–675

Generalized Alamouti Codes for Trading Quality of Service against Data Rate in MIMO UMTS

Christoph F. Mecklenbräuker
Forschungszentrum Telekommunikation Wien (ftw), Donau-City Straße 1, 1220 Vienna, Austria
Email: cfm@ftw.at

Markus Rupp
Institut für Nachrichtentechnik und Hochfrequenztechnik, Technische Universität Wien, Gusshausstraße 25-29, 1040 Vienna, Austria
Email: mrupp@nt.tuwien.ac.at

Received 17 December 2002; Revised 26 August 2003

New space-time block coding schemes for multiple transmit and receive antennas are proposed. First, the well-known Alamouti scheme is extended to $N_T = 2^m$ transmit antennas achieving high transmit diversity. Many receiver details are worked out for four and eight transmit antennas. Further, solutions for arbitrary, even numbers ($N_T = 2k$) of transmit antennas are presented achieving decoding advantages due to orthogonalization properties while preserving high diversity. In a final step, such extended Alamouti and BLAST schemes are combined, offering a continuous trade-off between quality of service (QoS) and data rate. Due to the simplicity of the coding schemes, they are very well suited to operate under UMTS with only very moderate modifications in the existing standard. The number of supported antennas at transmitter alone is a sufficient knowledge to select the most appropriate scheme. While the proposed schemes are motivated by utilization in UMTS, they are not restricted to this standard.

Keywords and phrases: mobile communications, space-time block codes, spatial multiplexing.

1. INTRODUCTION

One of the salient features of UMTS is the provisioning of moderately high data rates for packet switched data services. In order to maximize the number of satisfied users, an efficient resource assignment to the subscribers is desired allowing flexible sharing of the radio resources. Such schemes must address the extreme variations of the link quality. Standardization of UMTS is progressing steadily, and various schemes for transmit diversity [1] and high-speed downlink packet access (HSDPA) with multiple transmit and receive antennas (MIMO) schemes [2] are currently under debate within the Third Generation Partnership Project (http://www.3gpp.org/).

Recently, much attention has been paid to wireless MIMO systems, (cf. [3, 4, 5]). In [6, 7], it was shown that the wireless MIMO channel potentially has a much higher capacity than was anticipated previously. In [8, 9, 10], space-time coding (STC) schemes were proposed that efficiently utilize such channels. Alamouti [11] introduced a very simple scheme allowing transmissions from two antennas with the same data rate as on a single antenna but increasing the diversity at the receiver from one to two in a flat-fading channel. While the scheme works for BPSK even with four and eight antennas, it was proven that for QPSK, only the two-transmit-antenna scheme offers the full diversity gain [8, 12].

In order to evaluate the (single-) symbol error probability for a random channel $\mathbf{H}$ with N_T statistically independent transmission paths with zero-mean channel coefficients h_k ($k = 1, \ldots, N_T$) of equal variance,[1] known results from literature for maximum likelihood (ML) decoding of uncoded QPSK (with gray-code labelling) can be employed [13]:

$$\begin{aligned} \mathrm{BER}_{\mathrm{ML}} &= \frac{1}{2}\mathrm{E}_{\mathbf{H}}\left[\operatorname{erfc}\left(\sqrt{\frac{E_b}{2N_0}\sum_{k=1}^{N_T}|h_k|^2}\right)\right] \\ &= \frac{1}{2}\mathrm{E}_{\alpha_{\mathrm{ML}}}\left[\operatorname{erfc}\left(\sqrt{\frac{\alpha_{\mathrm{ML}}}{\sigma_V^2}}\right)\right]. \end{aligned} \tag{1}$$

Here the fading factor α_{ML} is introduced as a random variable with $\chi^2_{2N_T}$ density, the index indicating $2N_T$ degrees of freedom, that is, a diversity order of N_T. In case of independent complex Gaussian distributed variables h_k, the follow-

[1] We normalize $\sum_{k=1}^{N_T} \mathrm{E}[|h_k|^2] = 1$.

ing explicit result for QPSK modulation is obtained according to [13, Section 14.4, equations (15)]:

$$\begin{aligned} \mathrm{BER}_{\mathrm{ML}} &= \frac{1}{2}\int_0^\infty \mathrm{erfc}\left(\sqrt{x\frac{E_b}{2N_0}}\right)\frac{x^{N_T-1}}{\Gamma(N_T)}e^{-x}dx \\ &= \left[\frac{1-\mu}{2}\right]^{N_T}\sum_{k=0}^{N_T-1}\binom{N_T-1+k}{k}\left(\frac{1+\mu}{2}\right)^k, \end{aligned} \tag{2}$$

$$\mu = \sqrt{\frac{E_b/N_0}{N_T + E_b/N_0}}. \tag{3}$$

In contrast to this behavior, the performance for a linear zeroforeing (ZF) receiver is different. The bit error rate (BER) for a ZF receiver with N_T transmit and N_R receive antennas is given by [14]

$$\mathrm{BER}_{\mathrm{ZF}} = \frac{1}{2}\mathrm{E}_{\alpha_{\mathrm{ZF}}}\left[\mathrm{erfc}\left(\sqrt{\frac{\alpha_{\mathrm{ZF}}}{\sigma_V^2}}\right)\right], \tag{4}$$

with α_{ZF} being χ^2-distributed with $2(N_T - N_R + 1)$ degrees of freedom rather than $2N_T$. A good overview of the various single symbol error performances is given in [15] and some early results on multiple symbol errors in [16]. The proposed coding schemes of this paper will be compared with these results for uncoded transmissions. In particular, selecting space-time codes will result in different degrees of freedom for the resulting fading factor α when compared to (1) and (4).

The paper is composed as follows. In Section 2, the well-known Alamouti scheme is introduced setting the notation for the remaining of the paper. In Section 3, the Alamouti space-time codes for transmission diversity is extended recursively to $M = 2^m$ antenna elements at the transmitter. While it is well known that the resulting transmission matrix for flat-fading looses its orthogonality for $m \geq 2$, it is shown that the loss in orthogonality for the new schemes is not severe when utilizing gray-coded QPSK modulation. Starting with a four-antenna scheme in Section 3, it will be demonstrated that linear receivers perform close to the theoretical bound for four-path diversity offering significant gain over the two-antenna case proposed by Alamouti. Even more interestingly, linear interference suppression can be implemented at low-complexity because the channel matrix exhibits a high degree of structure, enabling factorization in closed-form. In Section 4, this observation is generalized to extended Alamouti schemes for an arbitrary number of transmit antennas $N_T = 2^m$ preserving as much orthogonality as possible. In particular, results will be presented for the case $N_T = 8$. Transmission schemes with more than one receive antenna will be considered in Section 5 and it will be shown that even in cases with $N_T \neq 2^m$ transmit antennas, preservation of orthogonality is possible. Variable bit rate services and bursty packet arrivals are handled flexibly in UMTS by dynamically changing the spreading factor in conjunction with the transmit power, thus preserving an average E_b/N_0, but without changing the diversity order and outage probability. A combination of BLAST and extended Alamouti schemes is proposed in Section 6 that makes use of the existing diversity in a flexible manner, trading diversity gain against data rate and thus augmenting the diversity order and outage probability for fulfilling the quality of service (QoS) requirements. Not considered in this paper is the impact of the modulation scheme on the achieved diversity. It is well known that a certain rank criterion [8] needs to be satisfied in order to utilize full channel diversity in MIMO systems.

2. ALAMOUTI SCHEME

A very simple but effective scheme for two ($N_T = 2$) antennas achieving a diversity gain of two was introduced by Alamouti [8, 11]. It works by sending the sequence $\{s_1, s_2^*\}$ on the first antenna and $\{s_2, -s_1^*\}$ on the other. Assuming a flat-fading channel and denoting the two channel coefficients by h_1 and h_2, the received vector $\mathbf{r}$ is formed by stacking two consecutive data samples $[r_1, r_2]^T$ in time:

$$\mathbf{r} = \mathbf{S}\mathbf{h} + \tilde{\mathbf{v}}. \tag{5}$$

Here, the symbol block $\mathbf{S}$ and the channel vector $\mathbf{h}$ are defined as follows:

$$\mathbf{S} = \begin{bmatrix} s_1 & s_2 \\ s_2^* & -s_1^* \end{bmatrix}, \qquad \mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}. \tag{6}$$

This can be reformulated as

$$\begin{bmatrix} r_1 \\ r_2^* \end{bmatrix} = \begin{bmatrix} h_1 & h_2 \\ -h_2^* & h_1^* \end{bmatrix}\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2^* \end{bmatrix} \tag{7}$$

or in short notation:

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{v}, \tag{8}$$

where the vector $\mathbf{y} = [r_1, r_2^*]^T$ is introduced. The resulting channel matrix $\mathbf{H}$ is orthogonal, that is, $\mathbf{H}^H\mathbf{H} = \mathbf{H}\mathbf{H}^H = h^2\mathbf{I}_2$, where the 2×2 identity matrix $\mathbf{I}_2$ as well as the gain of the channel $h^2 = |h_1|^2 + |h_2|^2$ are introduced. The transmitted symbols can be computed by the ZF approach

$$\hat{\mathbf{s}} = \left[\mathbf{H}^H\mathbf{H}\right]^{-1}\mathbf{H}^H\mathbf{y} = \frac{1}{h^2}\mathbf{H}^H\mathbf{y} = \mathbf{s} + \left[\mathbf{H}^H\mathbf{H}\right]^{-1}\mathbf{H}^H\mathbf{v}, \tag{9}$$

revealing a noise filtering. Note that due to the particular structure of $\mathbf{H}$, the two noise components are orthogonal. For a fixed channel matrix $\mathbf{H}$ and complex-valued Gaussian noise $\mathbf{v}$, it can be concluded that they are both i.i.d. and thus are two decoupled noise components. The noise variance for each of the two symbols is given by $2\sigma_V^2/h^2$. Comparing to the optimal ML result for two-path diversity, the results are identical indicating that with a simple ZF receiver technique, the full two-path diversity of the transmission system can be obtained. Using complex-valued modulation, only for the two-antenna scheme such an improvement is possible. Only in the case of binary transmission, higher schemes with four and eight antennas exist [12]. In UMTS, QPSK is utilized on CDMA preventing perfectly orthogonal schemes with an improvement larger than a diversity of two.

3. FOUR-ANTENNA SCHEME

In UMTS with frequencies around 2 GHz, four or even eight antennas are quite possible at the base stations and two or four antennas at the mobile [17]. Since the number of antennas will vary among base stations and mobile devices, it is vital to design a flexible MIMO transmission scheme supporting various multielement antennas. As a minimum requirement, the mobile station might only be informed about the number of transmit antennas at the base station. Based on its own number of receive antennas, it can then decide which decoding algorithm to apply. Some codes offer complexity proportional to the number of receive antennas, for example, cyclic space-time codes [18]. Another example being Hadamard codes, retransmitting the symbols in a specific manner. For the case of four transmit antennas, the resulting matrix becomes $[s_1, s_2, s_3, s_4;\ s_1, s_2, -s_3, -s_4;\ s_1, -s_2, -s_3, s_4;\ -s_1, s_2, -s_3, s_4]$. In such schemes, the receiver can be built with very low-complexity, and higher diversity is achievable with more receiver antennas. However, by only utilizing multiple receiver antennas, the maximum possible diversity is not utilized in such systems unless transmit diversity is utilized as well.

In the following, simple block codes supporting much higher diversity in a four transmit antenna scheme for UMTS are proposed which do take advantage of additional transmit diversity.[2]

Proposition 1. *Starting with the* 2×2*-Alamouti scheme, the following recursive construction rule (similar to the construction of a complex Walsh-Hadamard code) is applied:*

$$\begin{bmatrix} h_1 & h_2 \\ -h_2^* & h_1^* \end{bmatrix} \longrightarrow \begin{bmatrix} h_1 & h_2 & h_3 & h_4 \\ -h_2^* & h_1^* & -h_4^* & h_3^* \\ -h_3^* & -h_4^* & h_1^* & h_2^* \\ h_4 & -h_3 & -h_2 & h_1 \end{bmatrix}. \tag{10}$$

That is, the complex scalars h_1 and h_2 appearing to the left of the arrow "$\rightarrow$" are replaced by the 2×2 matrices

$$\begin{aligned} \mathbf{H}_1 &= \begin{bmatrix} h_1 & h_2 \\ -h_2^* & h_1^* \end{bmatrix}, \\ \mathbf{H}_2 &= \begin{bmatrix} h_3 & h_4 \\ -h_4^* & h_3^* \end{bmatrix}, \end{aligned} \tag{11}$$

and then reinserted into the Alamouti space-time channel matrix

$$\begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ -\mathbf{H}_2^* & \mathbf{H}_1^* \end{bmatrix}, \tag{12}$$

where $*$ denotes complex conjugation without transposition.

[2]The outage capacity of this scheme was originally reported in [19].

This results in the following symbol block $\mathbf{S}$ for transmitting the four symbols $\mathbf{s} = [s_1, \ldots, s_4]^T$:

$$\mathbf{S} = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 \\ s_2^* & -s_1^* & s_4^* & -s_3^* \\ s_3^* & s_4^* & -s_1^* & -s_2^* \\ s_4 & -s_3 & -s_2 & s_1 \end{bmatrix}. \tag{13}$$

The received vector can be expressed in the same form as (5). Converting the received vector by complex conjugation

$$\begin{aligned} y_1 &= r_1, & v_1 &= \bar{v}_1, \\ y_2 &= r_2^*, & v_2 &= \bar{v}_2^*, \\ y_3 &= r_3^*, & v_3 &= \bar{v}_3^*, \\ y_4 &= r_4, & v_4 &= \bar{v}_4, \end{aligned} \tag{14}$$

results in the following equivalent transmission scheme:

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{v}, \tag{15}$$

in which $\mathbf{H}$ appears again as channel transmission matrix. If $\bar{\mathbf{v}}$ is a complex-valued Gaussian vector with i.i.d. elements, then so is $\mathbf{v}$.

3.1. ML receiver performance

While a standard ML approach is possible with correspondingly high complexity, an alternative ML approach applying matched filtering is first possible with much less complexity. After the matched filtering operation, the resulting matrix $\mathbf{H}^H$ is

$$\mathbf{G} = \mathbf{H}^H\mathbf{H} = \mathbf{H}\mathbf{H}^H = h^2 \begin{bmatrix} \mathbf{I}_2 & X\mathbf{J}_2 \\ -X\mathbf{J}_2 & \mathbf{I}_2 \end{bmatrix}, \tag{16}$$

where the 2×2 matrix

$$\mathbf{J}_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{17}$$

as well as the Grammian $\mathbf{G}$ have been introduced. The gain of the channel is

$$h^2 = |h_1|^2 + |h_2|^2 + |h_3|^2 + |h_4|^2, \tag{18}$$

and the channel dependent real-valued random variable X is defined as follows:

$$X = \frac{2\,\mathrm{Re}\,(h_1 h_4^* - h_2 h_3^*)}{h^2}. \tag{19}$$

By applying the matched filter $\mathbf{H}^H$, this results in the reception of the following vector:

$$\mathbf{z} = \mathbf{H}^H\mathbf{y} = \mathbf{H}^H\mathbf{H}\mathbf{s} + \mathbf{H}^H\mathbf{v} = h^2 \begin{bmatrix} s_1 + Xs_4 \\ s_2 - Xs_3 \\ s_3 - Xs_2 \\ s_4 + Xs_1 \end{bmatrix} + \mathbf{H}^H\mathbf{v} \tag{20}$$

in which the pair $\{s_1, s_4\}$ is decoupled from $\{s_2, s_3\}$ allowing for a low-complexity solution based on the newly formed receiver vector $\mathbf{z}$.

The ML decoder selects $\mathbf{s}$ minimizing

$$\Lambda_1(\mathbf{s}) = \|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2 = \mathbf{s}^H\mathbf{G}\mathbf{s} - 2\mathrm{Re}(\mathbf{y}^H\mathbf{H}\mathbf{s}) + \|\mathbf{y}\|^2 \quad (21)$$

for all permissible symbol vectors $\mathbf{s}$ from the transmitter alphabet and spatially white interference plus noise was assumed. Alternatively, the matched filter can be applied to $\mathbf{y}$ and the ML estimator can be implemented on its output $\mathbf{z}$ given in (20) leading to

$$\Lambda_2(\mathbf{s}) = (\mathbf{z} - \mathbf{G}\mathbf{s})^H\mathbf{G}^{-1}(\mathbf{z} - \mathbf{G}\mathbf{s}). \quad (22)$$

Note that it needs to be taken into account that the noise plus interference is spatially correlated after filtering. Assuming the elements v_k of $\mathbf{v}$ to be zero mean and spatially white with variance σ_V^2 results in $\mathbf{w} = \mathbf{H}^H\mathbf{v}$ with covariance matrix

$$\mathrm{E}\left[\mathbf{w}\mathbf{w}^H\right] = \sigma_V^2\mathbf{H}^H\mathbf{H} = \sigma_V^2\mathbf{G}. \quad (23)$$

The advantage of this approach is that this partly decouples the symbols. The pair $\{s_1, s_4\}$ is decoupled from $\{s_2, s_3\}$ allowing for a low-complexity ML receiver using the partial metrics

$$\begin{aligned}
&\Lambda_{2a}(s_1, s_4) \\
&\quad = \left|z_1 - h^2(s_1 + Xs_4)\right|^2 + \left|z_4 - h^2(s_4 + Xs_1)\right|^2 \\
&\qquad - 2X\,\mathrm{Re}\left\{\left[z_1 - h^2(s_1 + Xs_4)\right]\left[z_4^* - h^2(s_4^* + Xs_1^*)\right]\right\}, \\
&\Lambda_{2b}(s_2, s_3) \\
&\quad = \left|z_2 - h^2(s_2 - Xs_3)\right|^2 + \left|z_3 - h^2(s_3 - Xs_2)\right|^2 \\
&\qquad + 2X\,\mathrm{Re}\left\{\left[z_2 - h^2(s_2 - Xs_3)\right]\left[z_3^* - h^2(s_3^* - Xs_2^*)\right]\right\}.
\end{aligned} \quad (24)$$

Note that the two metrics Λ_{2a} and Λ_{2b} are positive definite when $|X| < 1$. They become semidefinite for $|X| = 1$. In UMTS with QPSK modulation, this requires a search over 2×16 vector symbols rather than over 256.

3.2. Performance of linear receivers

Linear receivers typically suffer from noise enhancement. In this section, the increased noise caused by ZF and minimum mean squared error (MMSE) detectors is investigated. Both receivers can be described by the following detection principle:

$$\hat{\mathbf{s}} = \left(\mathbf{H}^H\mathbf{H} + \mu\mathbf{I}_4\right)^{-1}\mathbf{z}, \quad (25)$$

where $\mu = 0$ for ZF and $\mu = \sigma_V^2$ for MMSE. It turns out that both detection principles have essentially the same receiver complexity. The following lemmas can be stated.

Lemma 1. *Given the 4×4 Alamouti scheme as described in (10), the eigenvalues of $\mathbf{H}^H\mathbf{H}/h^2$ are given by*

$$\lambda_1 = \lambda_2 = 1 + X, \qquad \lambda_3 = \lambda_4 = 1 - X, \quad (26)$$

where h^2 and X are defined in (18) and (19).

Proof. The Grammian $\mathbf{H}^H\mathbf{H}$ is diagonalized by $\mathbf{V}_4^T\mathbf{H}^H\mathbf{H}\mathbf{V}_4$ with the orthogonal matrix

$$\mathbf{V}_4 = \frac{1}{\sqrt{2}}\begin{bmatrix}\mathbf{I}_2 & \mathbf{J}_2 \\ \mathbf{J}_2 & \mathbf{I}_2\end{bmatrix}. \quad (27)$$

□

Some favorable properties are worth mentioning. The eigenvectors of $\mathbf{H}^H\mathbf{H}$ which are stacked in the columns of $\mathbf{V}_4$ do not depend on the channel; they are constant. The scaled matrix $\sqrt{2}\mathbf{V}_4$ is sparse, that is, half of its elements vanish and the nonzero entries are ± 1.

Lemma 2. *If the channel coefficients h_i $(i = 1, \ldots, 4)$ are i.i.d. complex Gaussian variates with zero mean and variance 1/4, then the following properties hold:*

(1) X and h^2 are independent;

(2) let λ_i be an eigenvalue of $\mathbf{H}^H\mathbf{H}/h^2$. The probability density of λ_i is $f_{\lambda,4}(\lambda) = (3/4)\lambda(2 - \lambda)$ for $0 < \lambda < 2$ and zero elsewhere. Likewise, $\lambda_i/2$ is beta(2,2)-distributed;

(3) let ξ_i be an eigenvalue of $\mathbf{H}^H\mathbf{H}$. The probability density of ξ_i is $f_\xi(\xi) = 4\xi e^{-2\xi}$ for $\xi > 0$.

Proof. The joint distribution of X and h^2 is derived in Appendix A. The eigenvalues ξ_i of $\mathbf{H}^H\mathbf{H}$ and λ_i of $\mathbf{H}^H\mathbf{H}/h^2$ are proportional to each other, that is, $\xi_i = h^2\lambda_i$ for $i = 1, \ldots, 4$. □

It can be concluded that $\mathrm{E}[\lambda_i] = 1$ and $\mathrm{Var}(\lambda_i) = 0.2$ for all i, indicating that the normalized channel matrix $\mathbf{H}^H\mathbf{H}/h^2$ is close to a unitary matrix with high probability.

Let $\gamma \geq 1$ be the following random variable which depends on the channel gain if $\mu > 0$:

$$\gamma = \frac{h^2 + \mu}{h^2} = \begin{cases} 1 & \text{for ZF,} \\ 1 + \dfrac{\sigma_V^2}{h^2} & \text{for MMSE.} \end{cases} \quad (28)$$

For evaluating the BER of the linear receiver for general $\mu \neq 0$,

$$\begin{aligned}
&\mathrm{tr}\left[\left(\mathbf{H}^H\mathbf{H} + \mu\mathbf{I}_4\right)^{-1}\mathbf{H}^H\mathbf{H}\left(\mathbf{H}^H\mathbf{H} + \mu\mathbf{I}_4\right)^{-1}\right] \\
&\qquad = \left(\frac{4}{h^2}\right)\frac{\gamma^2 + X^2(1 - 2\gamma)}{\left(\gamma^2 - X^2\right)^2}
\end{aligned} \quad (29)$$

needs to be evaluated which is obtained via

$$\left[\mathbf{H}^H\mathbf{H} + \mu\mathbf{I}_4\right]^{-1} = \frac{1}{h^2(\gamma^2 - X^2)}\begin{pmatrix}\gamma\mathbf{I}_2 & -X\mathbf{J}_2 \\ X\mathbf{J}_2 & \gamma\mathbf{I}_2\end{pmatrix}. \quad (30)$$

When replacing the arguments of the complementary error function with (29), two interpretations can be discussed.

Comparing the arguments of the complementary error function with the standard ML solution for multiple diversity, one recognizes the beneficial diversity term h^2 indicating four times diversity together with an additional term, say

$$\delta_4 \triangleq \frac{\gamma^2 + X^2(1-2\gamma)}{(\gamma^2 - X^2)^2} = \frac{1}{\gamma^2 - X^2} - 2(\gamma - 1)\frac{X^2}{(\gamma^2 - X^2)^2}. \tag{31}$$

In Appendix A, it is shown that X and h^2 are statistically independent variates. Therefore, δ_4 can be interpreted as an increase in noise while h^2 causes full fourth-order diversity. Alternatively, one can interpret the whole expression $\alpha_{\mathrm{ZF},4} = h^2\delta_4$ as defining a new fading factor with the true diversity order without noise increase. Both interpretations can be used to describe the scheme's performance.

3.2.1. Noise enhancement

If the first interpretation is favoured, the following result is obtained.

Lemma 3. *Given the* 4×4 *Alamouti scheme in independent flat Rayleigh fading as described in* (10)*, a four-times diversity is obtained at the expense of a noise enhancement of*

$$\mathrm{E}[\delta_4] = \frac{3}{2} - 2\mu^2 + 2\mu e^{2\mu}\mathrm{E}_1(2\mu)(2\mu^2 + \mu - 2), \tag{32}$$

where $\mathrm{E}_n(x)$ *denotes the exponential integral defined for* $\mathrm{Re}(x) > 0$ *as follows:*

$$\mathrm{E}_n(x) \triangleq \int_1^\infty \frac{e^{-xt}}{t^n} dt. \tag{33}$$

Proof. The expectation $\mathrm{E}[\delta_4]$ in (A.10) needs to be evaluated. Note that δ_4 depends on X and h^2. It is shown in Appendix A that X and h^2 are independent if $h_1, \ldots, h_4$ are i.i.d. complex-valued zero-mean Gaussian variates. Therefore, we can evaluate $\mathrm{E}[\delta_4]$ via (A.11) which leads to the result (32). □

In case of a ZF receiver, the noise is increased by a factor of 3/2 which corresponds to 1.76 dB, a value for which the four-times diversity scheme gives much better results as long as E_b/N_0 is larger than about 3 dB. Therefore, the noise enhancement $\mathrm{E}[\delta_4]$ is maximum for ZF receivers ($\mu = 0$) and it does not exceed 1.76 dB for MMSE. The formula

$$\mathrm{E}[\delta_{N_T}] \triangleq \left(\frac{1}{N_T}\right)^2 \mathrm{E}\left[\mathrm{tr}\left((\mathbf{H}^H\mathbf{H})^{-1}\right)\mathrm{tr}\left(\mathbf{H}^H\mathbf{H}\right)\right] = 2\frac{N_T - 1}{N_T} \tag{34}$$

seems to describe the noise enhancement for ZF receivers for the general case of N_T transmit antennas. Note that $\mathrm{tr}(\mathbf{H}^H\mathbf{H})$ is the squared Frobenius norm of $\mathbf{H}$. The argument of the expectation operator is closely related to the numerical condition number κ of $\mathbf{H}$. Let ξ_{N_T} and ξ_1 be the largest and the smallest eigenvalue of $\mathbf{H}^H\mathbf{H}$, respectively. Then $\mathrm{tr}((\mathbf{H}^H\mathbf{H})^{-1})\,\mathrm{tr}(\mathbf{H}^H\mathbf{H}) \geq \xi_{N_T}/\xi_1 = \kappa^2$. The noise enhancement can be lower bounded by the squared numerical condition number, that is, $\mathrm{E}[\delta_{N_T}] \geq \mathrm{E}[\kappa^2]$.

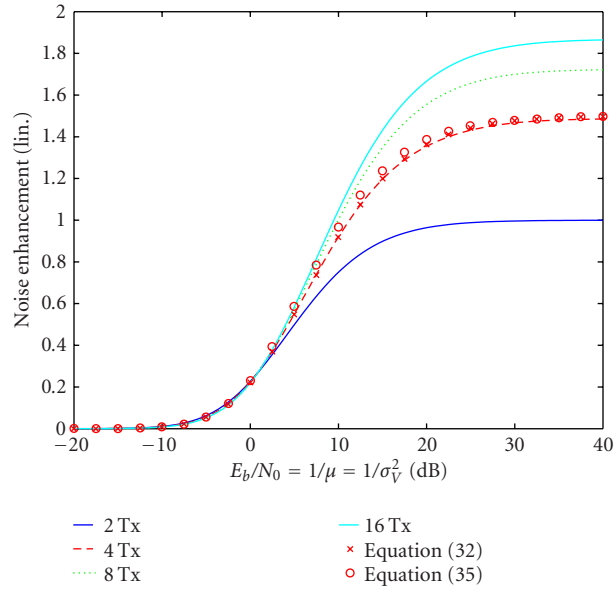


Figure 1: Comparison of the noise enhancement versus $E_b/N_0 = 1/\sigma_V^2$ for ZF and MMSE receivers.

The formula was explicitly validated for $N_T = 2, 4$, and 8 and with Monte Carlo simulations for larger values of N_T. Although no formal proof exists, the upper limit for the noise enhancement was found at 3 dB. The behavior of (32) versus $1/\mu$ (which equals E_b/N_0 for the MMSE) is shown in Figure 1 indicated by crosses labeled "x."

Additional insight into the behavior of (32) is gained by regarding the channel gain h^2 as approximately constant, an assumption that holds asymptotically true for $N_T \to \infty$. This assumption enables us to replace the joint expectation over X and γ in (32) by a conditional one, that is, conditioned on h^2,

$$\mathrm{E}[\delta_4|h^2] = \frac{9}{2}\gamma - 3 + \left(\frac{9}{4}\gamma^2 - \frac{3}{2}\gamma - \frac{3}{4}\right)\log\frac{\gamma - 1}{\gamma + 1}. \tag{35}$$

This approximation is compared with the exact expression of (32) in Figure 1 where the approximation obtained from (35) is plotted versus $\mathrm{E}[1/(\gamma - 1)] = E_b/N_0$. The values are indicated by circles labeled "∘." The horizontal shift in E_b/N_0 between (32) and (35) is generally less than 1 dB. This approximation becomes exact for the case of ZF receiver where $\mu \to 0$, that is, the limit for $\gamma \to 1$ of (35) is 3/2.

3.2.2. True diversity

The second interpretation of (29) leads to a refined diversity order. In this case, the term in γ and X purely modifies the diversity but leaves the noise part unchanged. The BER performance can be computed explicitly. We restrict ourselves to the ZF case for which $\gamma = 1$ and $\delta_4 = 1/[1 - X^2]$. In this case, δ_4 and h^2 are statistically independent. We obtain

$$\mathrm{BER}_{\mathrm{ZF}} = \int_h \int_\delta \mathrm{erfc}\left(\frac{h^2\delta}{2\sigma_V^2}\right)\frac{h^3 e^{-h}}{2\Gamma(4)} f_\delta(\delta) dh\, d\delta. \tag{36}$$

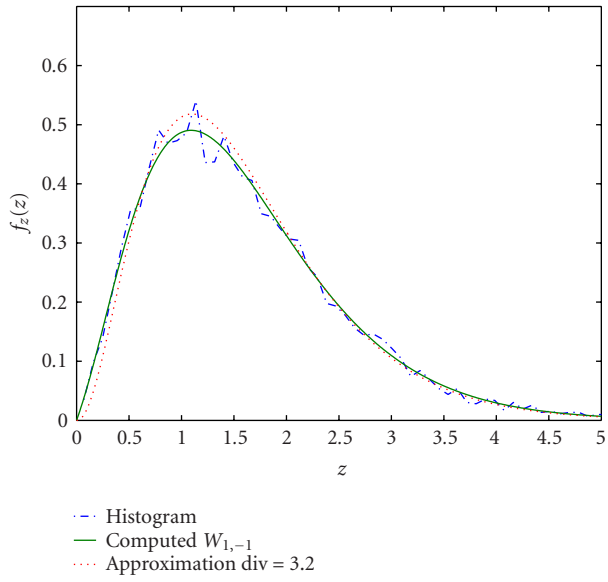


FIGURE 2: Histogram of a sample of z defined in (40) and its density $f_z(z)$ in (43).

Using the result from [13], (2) is obtained correspondingly, however, with a different solution for a random variable μ:

$$\mu(X) = \sqrt{\frac{E_b/N_0}{2(1-X^2)+E_b/N_0}}, \tag{37}$$

leading to rather involved terms. A much simpler method is to interpret the term $h^2\delta$ as a new fading factor $\alpha_{\mathrm{ZF},4}$ with χ-statistics. Since δ is a fractional number, the new factor $\alpha_{\mathrm{ZF},4} = h^2\delta$ cannot be expected to have an integer number of freedoms. Comparing with a Nakagami-m density, the mean value of $h^2\delta$ corresponds to the number of degrees of freedom m for this density. Computing $\mathrm{E}[h^2\delta_4] = m = 3.2$ is obtained. Figure 2 displays a histogram of $\alpha_{\mathrm{ZF},4}$ from 5,000 runs. Furthermore, the exact density function is shown and a close fit obtained by the squared Nakagami-m distribution with $m = 3.2$, or equivalent χ^2 with 6.4 degrees of freedom. This result contradicts the general belief that ZF receivers obtain only $2(N_R - N_T + 1) = 2$ degrees of freedom. The result is different here due to the channel structuring.

An exact derivation of the probability density for this random variable is lengthy and is only sketched here. The random variable $(1 - X^2)h^2$ can be constructed from two independent variables $\mathbf{u}^H\mathbf{u}$ and $\mathbf{v}^H\mathbf{v}$ which are each χ^2-distributed with four degrees of freedom (diversity order two). Substitute

$$\mathbf{x}^T = [h_1, h_2], \qquad \mathbf{y}^T = [h_4, -h_3]. \tag{38}$$

Then $X = (\mathbf{x}^H\mathbf{y} + \mathbf{y}^H\mathbf{x})/(\mathbf{x}^H\mathbf{x} + \mathbf{y}^H\mathbf{y})$. Using $\mathbf{u} = [\mathbf{x} - \mathbf{y}]/\sqrt{2}$ and $\mathbf{v} = [\mathbf{x} + \mathbf{y}]/\sqrt{2}$, the following result is obtained:

$$(1-X^2)h^2 = 4\frac{\mathbf{u}^H\mathbf{u}\mathbf{v}^H\mathbf{v}}{\mathbf{u}^H\mathbf{u}+\mathbf{v}^H\mathbf{v}} = \frac{4}{1/\mathbf{u}^H\mathbf{u}+1/\mathbf{v}^H\mathbf{v}}. \tag{39}$$

The joint density $p_{w,z}(w,z)$ of this expression can be computed via the transformation

$$z = \frac{1}{1/\mathbf{u}^H\mathbf{u}+1/\mathbf{v}^H\mathbf{v}}, \qquad w = \mathbf{v}^H\mathbf{v}, \tag{40}$$

achieving

$$p_{w,z}(w,z) = \frac{w^3 z}{(w-z)^3}\exp\left(-\frac{w^2}{w-z}\right). \tag{41}$$

The density of z is found by marginalizing the joint density $p(w,z)$. The density can be expressed using a Whittaker function (see [20]):

$$f_z(z) = 2^9 z\int_u^\infty \frac{t^{3/2}\exp(-4t)}{\sqrt{t-z}}dt \tag{42}$$

$$= 2^6 z^{3/2}\Gamma\left(\frac{1}{2}\right)\exp(-2z)W_{1,-1}(4z) \approx \frac{4^{3.2}z^{2.2}e^{4z}}{\Gamma(3.2)}. \tag{43}$$

This last approximation is also shown in Figure 2, obviously a good fit.

3.3. Simulation results

Figure 3 displays the simulated behavior of the uncoded BER transmitting QPSK (gray coded) of the linear MMSE receiver and zero fading correlation between the four transmit paths. The BER results were averaged over 16,000 symbols and 3,200 selections of channel matrices $\mathbf{H}$ for each simulated E_b/N_0. For comparison, the BER from the ZF receiver and the cases of ideal two- and four-path diversity are also shown. The values marked by circles "$\circ$" labeled "expected theory" are the same as for four-path diversity, but shifted by the noise enhancement (n.e.) of 1.76 dB. Compared to the ZF receiver performance, there is just a little improvement for MMSE.

For practical considerations, it is of interest to investigate the performance when the four paths are correlated, as can be expected in a typical transmission environment. Figure 4 displays the situation when the antenna elements are correlated by a factor of $\{0.5, 0.75, 0.95\}$. As the figure reveals, no further loss is shown until the value exceeds 0.5. Only with very strong correlation (0.95), a degradation of 4 dB was noticed.

3.4. Diversity cumulating property of receive antennas

An interesting property is worth mentioning coming with the 4×1 extended Alamouti scheme when using more than one receive antenna. Typically adding more receive antennas gives rise to expect a higher diversity order in the transmission system, however, available only at the expense of more complexity in the receiver algorithms. In the extended Alamouti scheme, the behavior is slightly different as stated in the following lemma.

Lemma 4. *When utilizing an arbitrary number N_R of receive antennas, the extended Alamouti scheme can obtain an N_R-fold*

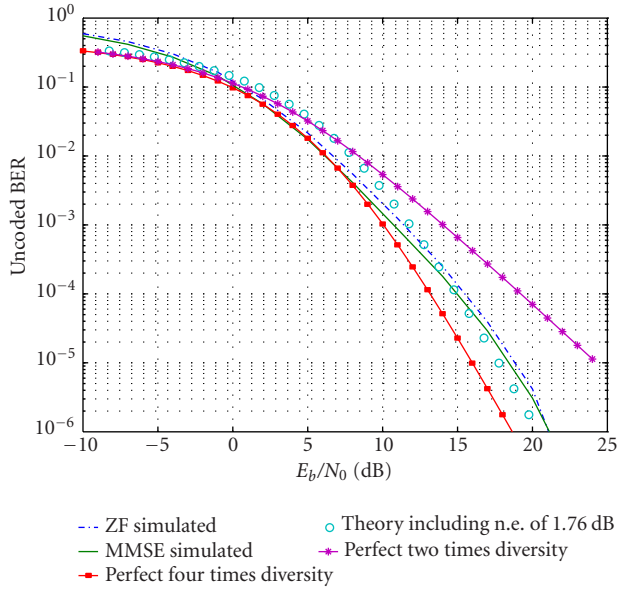


FIGURE 3: BER for four-antenna scheme with linear MMSE receiver and zero correlation between antennas.

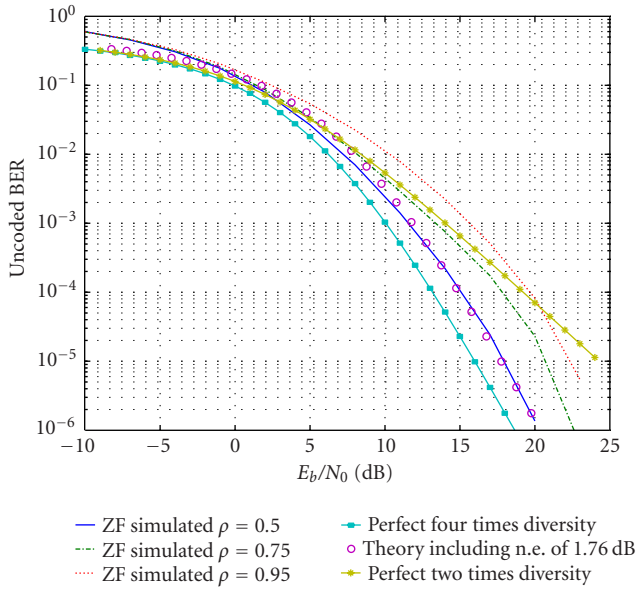


FIGURE 4: BER for four-antenna scheme with ZF receiver, fading correlation between adjacent antenna elements is $\{0.5, 0.75, 0.95\}$.

diversity compared to the single receive antenna case requiring only an asymptotically linear complexity $O(N_R)$ for ML as well as linear receivers.

Proof. The proof will be shown for two receive antennas. Extending it to more than two is a straight forward exercise:

$$\mathbf{r}_1 = \mathbf{H}_1\mathbf{s} + \mathbf{v}_1; \qquad \mathbf{r}_2 = \mathbf{H}_2\mathbf{s} + \mathbf{v}_2. \tag{44}$$

Matched filtering can be applied and the corresponding terms are summed up to obtain

$$\begin{aligned}\mathbf{H}_1^H\mathbf{r}_1 + \mathbf{H}_2^H\mathbf{r}_2 &= [\mathbf{H}_1^H\mathbf{H}_1 + \mathbf{H}_2^H\mathbf{H}_2]\mathbf{s} + \mathbf{H}_1^H\mathbf{v}_1 + \mathbf{H}_2^H\mathbf{v}_2 \\ &= [\mathbf{H}_1^H\mathbf{H}_1 + \mathbf{H}_2^H\mathbf{H}_2]\mathbf{s} + \tilde{\mathbf{v}}.\end{aligned} \tag{45}$$

Note that the new matrix $[\mathbf{H}_1^H\mathbf{H}_1 + \mathbf{H}_2^H\mathbf{H}_2]$ preserves the form (16):

$$\mathbf{H}_1^H\mathbf{H}_1 + \mathbf{H}_2^H\mathbf{H}_2 = (h_1^2 + h_2^2)\begin{bmatrix} 1 & 0 & 0 & X \\ 0 & 1 & -X & 0 \\ 0 & -X & 1 & 0 \\ X & 0 & 0 & 1 \end{bmatrix}, \tag{46}$$

with $X = [X_1h_1^2 + X_2h_2^2]/[h_1^2 + h_2^2]$. Thus, the matrix maintains its form and therefore, complexity of ML or a linear receiver remains identical to the one antenna case. Only the matched filtering needs to be performed additionally for as many receive antennas are present. The leading term $h_1^2 + h_2^2$ describes the diversity order, being twice as high as before. For N_R receiver antennas, a sum of all terms h_k^2, $k = 1, \ldots, N_R$, will appear in this position indicating an N_R-fold increase in capacity. □

Note that N_R receiver antennas can be purely virtual and do not necessarily require a larger RF front end effort. For example, UMTS's WCDMA scheme enables RAKE techniques to be utilized. Thus, at tap delays τ_k where large energies occur, a finger of the RAKE receiver is positioned. Correspondingly, the channel matrix $\mathbf{H}$ consists in this case of several components, all located at K different delay times. The received values can be structured in one vector as well and $\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{v}$ is obtained again, however now with $\mathbf{y}$ is of dimension $4K \times 1$ and $\mathbf{H}$ of dimension $4K \times 4$, while $\mathbf{s}$ remains of dimension 4×1 as before. The previously discussed schemes can be applied as well and each term h^2 now consists of K times as many components as before, thus increasing diversity by a factor of K. In conclusion, such techniques work as well in a scenario with interchip interference as in flat Rayleigh fading with the additional benefit of having even more diversity and thus a better QoS, provided the cross-correlation between different users remains limited.

4. EIGHT AND MORE ANTENNA SCHEMES

Applying (10) several times ($m - 1$ times), solutions for $N_T = 2^m \times 1$ antenna schemes can be obtained. The obtained matrices exhibit certain properties that will be utilized in the following. They are listed in the following lemma and proven in Appendix B.

Lemma 5. *Applying rule* (10) $m-1$ *times results in matrices* $\mathbf{H}$ *of dimension* $N_T \times N_T$, $N_T = 2^m$, *with the following properties:*

(1) all entries of $\mathbf{H}^H\mathbf{H}$ *are real-valued;*
(2) the matrix $\mathbf{H}^H\mathbf{H}$ *is of the form*

$$\mathbf{H}^H\mathbf{H} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B} & \mathbf{A} \end{bmatrix} \tag{47}$$

and the inverse of $\mathbf{H}^H\mathbf{H}$ *is of block matrix form*

$$\left[\mathbf{H}^H\mathbf{H}\right]^{-1} = \begin{bmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{bmatrix} \begin{bmatrix} \left(\mathbf{A}^2+\mathbf{B}^2\right)^{-1} & \varnothing \\ \varnothing & \left(\mathbf{A}^2+\mathbf{B}^2\right)^{-1} \end{bmatrix}. \quad (48)$$

Due to the form (47)*, all eigenvalues are double;*[3]

(3) *each nondiagonal entry* X_i *of* $\mathbf{H}^H\mathbf{H}/\mathrm{tr}[\mathbf{H}^H\mathbf{H}]$ *is either zero, or* X_i *follows the distribution*

$$f_X(\xi) = \frac{1}{2^{N_T-2}\mathrm{B}(N_T/2, N_T/2)}\left(1-\xi^2\right)^{N_T/2-1}, \quad |\xi| \le 1. \quad (49)$$

Applying rule (10) two times in succession results in the 8×8 scheme. It can immediately be verified that the matrix $\mathbf{H}^H\mathbf{H}$ is given by

$$\mathbf{H}^H\mathbf{H} = h^2 \begin{bmatrix} \mathbf{I}_2 & X\mathbf{J}_2 & -Z\mathbf{J}_2 & Y\mathbf{I}_2 \\ -X\mathbf{J}_2 & \mathbf{I}_2 & -Y\mathbf{I}_2 & -Z\mathbf{J}_2 \\ Z\mathbf{J}_2 & -Y\mathbf{I}_2 & \mathbf{I}_2 & X\mathbf{J}_2 \\ Y\mathbf{I}_2 & Z\mathbf{J}_2 & -X\mathbf{J}_2 & \mathbf{I}_2 \end{bmatrix}, \quad (50)$$

with

$$\begin{aligned} h^2 &= \sum_{k=1}^{8} |h_k|^2, \\ X &= \frac{2\,\mathrm{Re}\left(h_1h_4^* - h_2h_3^* + h_5h_8^* - h_6h_7^*\right)}{h^2}, \\ Y &= \frac{2\,\mathrm{Re}\left(h_1h_7^* - h_3h_5^* + h_2h_8^* - h_4h_6^*\right)}{h^2}, \\ Z &= \frac{2\,\mathrm{Re}\left(h_2h_5^* - h_1h_6^* + h_4h_7^* - h_3h_8^*\right)}{h^2}. \end{aligned} \quad (51)$$

According to property (2), the block structure of this matrix can be recognized. Note that $\mathbf{A}^2+\mathbf{B}^2 = \alpha\mathbf{I}_4 + \beta\mathbf{J}_4$, with

$$\begin{aligned} \mathbf{J}_4 &= \begin{bmatrix} \varnothing & \mathbf{J}_2 \\ -\mathbf{J}_2 & \varnothing \end{bmatrix}, \\ \alpha = X^2 - Y^2 - Z^2 + 1, &\qquad \beta = 2(X - YZ), \end{aligned} \quad (52)$$

and the inverse can also be expressed by a combination of $\mathbf{I}_4$ and $\mathbf{J}_4$:

$$\left[\mathbf{A}^2+\mathbf{B}^2\right]^{-1} = \frac{1}{\alpha^2-\beta^2}\left(\alpha\mathbf{I}_4 - \beta\mathbf{J}_4\right) \quad (53)$$

if $|\alpha| \neq |\beta|$ which enables a computationally efficient implementation.

The ML receiver decouples into two 4×4 schemes by exploiting the structure of these matrices, (cf. Section 3.1). For UMTS with QPSK modulation, this leads to a search over 2×256 vector symbols rather than $4^8 = 65\,536$.

[3]The proof of the latter statement is simple: if an eigenvector $[\mathbf{x}, \mathbf{y}]$ exists for an eigenvalue λ, then also $[\mathbf{y}, -\mathbf{x}]$ must be an eigenvector, linear independent of the first one, and thus the eigenvalues must be double.

4.1. Performance of linear receivers

Proceeding analogously to Section 3.2, the noise enhancement $\mathrm{E}[\delta_8]$ for the eight-antenna scheme is governed by $\mathrm{tr}[(\mathbf{H}^H\mathbf{H} + \mu\mathbf{I}_8)^{-1}\mathbf{H}^H\mathbf{H}(\mathbf{H}^H\mathbf{H} + \mu\mathbf{I}_8)^{-1}] = 8\delta_8/h^2$, where $\gamma = 1 + \mu/h^2$ and

$$\delta_8 \triangleq \frac{1}{8}\sum_{i=1}^{8}\frac{\lambda_i}{(\gamma+\lambda_i-1)^2}. \quad (54)$$

Lemma 6. *All eigenvalues* λ_i *of* $\mathbf{H}^H\mathbf{H}/h^2$ *in* (50) *are given by*

$$\begin{aligned} \lambda_1 &= \lambda_2 = (1-X) + (Y-Z), \\ \lambda_3 &= \lambda_4 = (1+X) - (Y+Z), \\ \lambda_5 &= \lambda_6 = (1+X) + (Y+Z), \\ \lambda_7 &= \lambda_8 = (1-X) - (Y-Z). \end{aligned} \quad (55)$$

Proof. The Grammian $\mathbf{H}^H\mathbf{H}$ is diagonalized by $\mathbf{V}_8^T\mathbf{H}^H\mathbf{H}\mathbf{V}_8$ with the orthogonal matrix

$$\mathbf{V}_8 = \frac{1}{2}\begin{bmatrix} \mathbf{I}_2 & \mathbf{J}_2 & \mathbf{J}_2 & \mathbf{I}_2 \\ \mathbf{J}_2 & \mathbf{I}_2 & -\mathbf{I}_2 & -\mathbf{J}_2 \\ \mathbf{J}_2 & \mathbf{I}_2 & \mathbf{I}_2 & \mathbf{J}_2 \\ -\mathbf{I}_2 & -\mathbf{J}_2 & \mathbf{J}_2 & \mathbf{I}_2 \end{bmatrix} \quad (56)$$

resulting in the above given eigenvalues. □

Lemma 7. *If the channel coefficients* h_i $(i = 1,\ldots,8)$ *are i.i.d. complex-valued Gaussian variates with zero mean and variance* 1/8*, then the following properties hold:*

(1) *let* λ_i *be an eigenvalue of* $\mathbf{H}^H\mathbf{H}/h^2$*. The probability density of* λ_i *is* $f_{\lambda,8}(\lambda) = (21/8192)\lambda(4-\lambda)^5$ *for* $0 < \lambda < 4$ *and zero elsewhere. Likewise,* $\lambda_i/4$ *is beta(2,6)-distributed;*

(2) *let* ξ_i *be an eigenvalue of* $\mathbf{H}^H\mathbf{H}$*. The probability density of* ξ_i *is* $f_\xi(\xi) = 4\xi e^{-2\xi}$ *for* $\xi > 0$ *and zero elsewhere.*

Proof. It is sufficient to give the proof for one eigenvalue, say λ_5. The proof for the remaining eigenvalues proceeds similarly. By completing the squares (as in Appendix A), $h^2\lambda_5/4$ can be regarded as the sum of two χ_n^2-distributed variables with $n = 2$ degrees of freedom each, that is,

$$\left|\frac{h_1+h_4-h_6+h_7}{2}\right|^2 + \left|\frac{h_2-h_3+h_5+h_8}{2}\right|^2. \quad (57)$$

By introducing an orthogonal transformation via the matrix $\mathbf{V}_8^T$ from (56), the proof is completed following the procedure in Appendices A and B. □

The noise enhancement for the eight-antenna case and a ZF receiver ($\mu = 0$) is evaluated by using the eigenvalue statistics from Lemma 7:

$$\mathrm{E}[\delta_8] = \int_0^4 \lambda^{-1} f_{\lambda,8}(\lambda)\,d\lambda = \frac{7}{4} = 1.75 \quad (58)$$

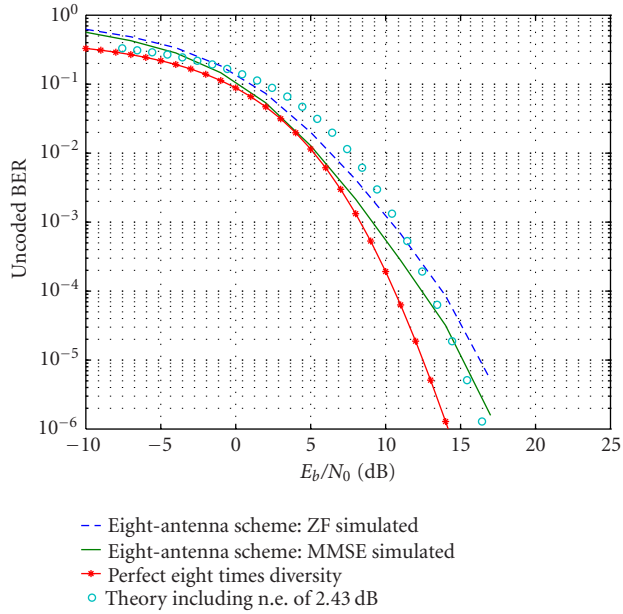


Figure 5: BER for eight-antenna scheme for ZF and MMSE receivers compared to theory.

or around 2.43 dB. The noise enhancement for the general linear receiver ($\mu \geq 0$) is obtained similarly to the four-antenna scheme; the result is

$$\mathrm{E}\left[\delta_8\right] = \frac{7}{4} + 2\mu - \mu^2 + \mu e^{2\mu}\mathrm{E}_1(2\mu)\left(2\mu^2 - 3\mu - 6\right). \quad (59)$$

Thus, the noise enhancement of the MMSE receiver is always smaller than 2.43 dB. Figure 1 compares the noise enhancement versus SNR for the ZF and MMSE receivers and for Alamouti's two-, and the proposed four-, and eight-antenna schemes. The noise enhancement for each scheme is calculated numerically by averaging over 4000 realizations of the channel matrix $\mathbf{H}$. For each realization, the eigenvalues λ_i of $\mathbf{H}^H\mathbf{H}$ are numerically computed and subsequently averaged over $(h^2/N_T)\sum_{i=1}^{N_T}\lambda_i/(\lambda_i+\mu)^2$, where $N_T = 2, 4, 8$, or 16. The resulting averaged curves are shown in Figure 1 labeled "2 Tx," "4 Tx," and so forth.

The theoretical values marked by small crosses, labeled "x," are calculated according to (32) versus $E_b/N_0 = 1/\sigma_V^2 = 1/\mu$ for the MMSE case. The values marked by small circles, labeled "∘," are calculated according to the approximation in (35) versus $E_b/N_0 = \mathrm{E}\left[1/(\gamma - 1)\right]$.

4.2. Simulation results

Figure 5 displays the simulated behavior of the uncoded BER for QPSK modulation and zero-fading correlation between the eight transmit paths. The BER results were averaged over 12,800 symbols and 4,000 selections of channel matrices $\mathbf{H}$ for each simulated E_b/N_0. The results are shown for a significance level of 99.7%. In other words, the scheme assumes a tolerated outage probability of 0.3%. Outage is assumed to occur if the numerical condition of $\mathbf{H}^H\mathbf{H}$ which is the ratio of the largest to the smallest eigenvalue exceeds $100 \approx 2^7$. Inverting these rare but adverse (nearly singular) channel matrices $\mathbf{H}^H\mathbf{H}$ lead to the loss of at least seven bits of numerical accuracy in the receiver. The values marked by little circles "∘" labeled "expected theory" are the same as for eight-path diversity, but shifted by the noise variance increase of 2.43 dB.

5. ALAMOUTIZATION

So far, mostly $N_T \times 1$ antenna schemes have been considered. However, in the future several antennas are likely to occur at the receiver as well. A cellular phone can carry two and a laptop as many as four antennas [17]. The proposed schemes can be applied, however, it remains unclear how to combine the received signals in an optimal fashion. In the following, an interesting approach is presented allowing an increase in diversity when the number of receiver antennas is more than one but typically less than the number of transmit antennas. The proposed STC schemes preserve a large part of the orthogonality so that the receivers can be implemented with low-complexity. The diversity is exploited in full and the noise enhancement remains small.

Proposition 2. *Assume that a block matrix form of the channel matrix* $\mathbf{H}$ *is given by*

$$\mathbf{H} = [\mathbf{H}_1\mathbf{H}_2], \quad (60)$$

where the matrices $\{\mathbf{H}_1, \mathbf{H}_2\}$ *are not necessarily quadratic. Then, the scheme can be* Alamouted by performing the following operation:

$$\mathbf{G} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ -\mathbf{H}_2^* & \mathbf{H}_1^* \\ \mathbf{H}_2^* & \mathbf{H}_1^* \\ \mathbf{H}_1 & -\mathbf{H}_2 \end{bmatrix}. \quad (61)$$

At the receiver, a ZF operation is performed, obtaining the corresponding term $\mathbf{G}^H\mathbf{G}$ with the property

$$\mathbf{G}^H\mathbf{G} = 2\begin{bmatrix} \mathbf{H}_1^H\mathbf{H}_1 + \mathbf{H}_2^T\mathbf{H}_2^* & \varnothing \\ \varnothing & \mathbf{H}_1^T\mathbf{H}_1^* + \mathbf{H}_2^H\mathbf{H}_2 \end{bmatrix}. \quad (62)$$

Thus perfect orthogonality on the nondiagonal block entries is achieved indicating little noise enhancement while the diagonal block terms indicate high diversity values.[4]

Example 1. A two-transmit-two-receive antenna system is considered:

$$\mathbf{H}_1 = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}, \qquad \mathbf{H}_2 = \begin{bmatrix} h_3 \\ h_4 \end{bmatrix}. \quad (63)$$

The matrix $\mathbf{G}^H\mathbf{G}$ becomes

$$\mathbf{G}^H\mathbf{G} = 2\left(|h_1|^2 + |h_2|^2 + |h_3|^2 + |h_4|^2\right)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (64)$$

[4] This was proposed in [4] in a simpler form.

Thus, the full four times diversity can be explored, without a matrix inverse computation. Note that in this case, the transmit sequence at the two antennas reads

$$\begin{Bmatrix} s_1 & s_2 & -s_3^* & -s_4^* & s_3^* & s_4^* & s_1 & s_2 \\ s_3 & s_4 & s_1^* & s_2^* & s_1^* & s_2^* & -s_3 & -s_4 \end{Bmatrix}. \tag{65}$$

Note also that during eight time periods, only four symbols are transmitted, that is, this particular scheme has the drawback of offering only half the symbol rate!

Example 2. Consider a 4×2 transmission scheme. The matrices are identified to

$$\mathbf{H}_1 = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}, \qquad \mathbf{H}_2 = \begin{bmatrix} h_{13} & h_{14} \\ h_{23} & h_{24} \end{bmatrix}. \tag{66}$$

The matrix $\mathbf{G}^H\mathbf{G}$ consists of two block matrices of size 2×2 on the diagonal. Thus, the scheme is still rather simple since only a 2×2 matrix has to be inverted although a four-path diversity is achieved. A comparison of the noise enhancement shows that for this 4×2 antenna system, 3 dB is gained compared to the 4×1 antenna system. Note that now the data rate is at full speed!

Example 3. The previously discussed 4×1 antenna system can be obtained when setting

$$\mathbf{H}_1 = \begin{bmatrix} h_1 & h_2 \end{bmatrix}, \qquad \mathbf{H}_2 = \begin{bmatrix} h_3 & h_4 \end{bmatrix}. \tag{67}$$

The reader may also try schemes in which the number of receive antennas is not given by $N_R = 2^n$. As long as N_R is even, the scheme can be separated in two matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ of same size allowing the Alamoutization rule (Proposition 2) to be applied.

6. COMBINING BLAST AND ALAMOUTI SCHEMES

Although the proposed extended Alamouti schemes allow for utilizing the channel diversity without sacrificing the receiver complexity, not much has been said on data rates yet. In the case of $N_T \times 1$ antenna schemes, the N_T symbols were repeated N_T times in a different and specific order guaranteeing a data rate of one. Thus, the data rates in the proposed schemes typically remain constant (equal to one) when the schemes are quadratic and can be lower when the receive antenna number is smaller than the transmit antennas as pointed out in the previous section. In BLAST transmissions, this is different. In its simplest form, the V-BLAST coding [21], N_T new symbols are offered to the N_T transmit antennas at every symbol time instant thus achieving data rates N_T times higher than in the Alamouti schemes. A combination of schemes can be achieved by simply transmitting more or less of the different repetitive transmissions. By utilizing the obtained transmission matrix structures, the diversity inherent in the transmission scheme can be exploited differently offering a trade-off between data rate and diversity order. In order to clarify this statement, an example is presented.

Table 1

Antenna	$n = 1$	$n = 2$
1	s_1	s_2^*
2	s_2	$-s_1^*$
3	s_3	s_4^*
4	s_4	$-s_3^*$

Example 4. A 4×2 antenna scheme is considered for transmission. In a flat-fading channel system, eight Rayleigh coefficients are available describing the transmissions from the four transmit to the two receive antennas, the transmission matrix being

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \end{bmatrix}. \tag{68}$$

It should thus be possible to transmit either four times the symbol data rate with diversity gain two, or two times the data rate with diversity four, or only at the symbol data rate but with diversity gain eight. In the first case, the 4×1 scheme as proposed in Section 3 will be used, repeating the four symbols four times, resulting in the reception of eight symbols. When assigning two paths each to one 2×2 matrix $\mathbf{H}_i$, $i = 1, \ldots, 4$, the following transmission matrix is obtained:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ -\mathbf{H}_2^* & \mathbf{H}_1^* \\ \mathbf{H}_3 & \mathbf{H}_4 \\ -\mathbf{H}_4^* & \mathbf{H}_3^* \end{bmatrix}. \tag{69}$$

Computing $\mathbf{H}^H\mathbf{H}$, a 4×4 matrix is obtained in a similar way to the 4×1 antenna case, however with twice the diversity. Thus in this case, a diversity of eight is achieved with a data rate of one.

On the other hand, by transmitting the sequences only twice, according to Table 1, the received signals at the two antennas can be formed to

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ -h_{12}^* & h_{11}^* & -h_{14}^* & h_{13}^* \\ h_{21} & h_{22} & h_{23} & h_{24} \\ -h_{22}^* & h_{21}^* & -h_{24}^* & h_{23}^* \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{bmatrix} = \mathbf{H}\mathbf{s}. \tag{70}$$

Thus, computing $\mathbf{H}^H\mathbf{H}$ results simply in the following block matrix:

$$\mathbf{H}^H\mathbf{H} = \begin{bmatrix} \gamma_1\mathbf{I} & \mathbf{B} \\ \mathbf{B}^H & \gamma_2\mathbf{I} \end{bmatrix} \tag{71}$$

with $\gamma_1 = |h_{11}|^2 + |h_{12}|^2 + |h_{13}|^2 + |h_{14}|^2$ and $\gamma_2 = |h_{21}|^2 + |h_{22}|^2 + |h_{23}|^2 + |h_{24}|^2$. Due to the condition $\mathbf{B}^H\mathbf{B} = \mathbf{B}\mathbf{B}^H$, such matrices can be inverted with a 2×2 matrix inversion rather than a 4×4:

$$\left[\mathbf{H}^H\mathbf{H}\right]^{-1} = \begin{bmatrix} \gamma_2\mathbf{I} & -\mathbf{B} \\ -\mathbf{B}^H & \gamma_1\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{C} & \varnothing \\ \varnothing & \mathbf{C} \end{bmatrix} \tag{72}$$

with $\mathbf{C} = [\gamma_1\gamma_2\mathbf{I} - \mathbf{B}\mathbf{B}^H]^{-1}$. Thus, the underlying Alamouti scheme gives us the advantage of lower complexity while the

BLAST scheme offers higher data rate. This specific scheme was investigated in [22, 23], where a diversity factor of six was found to closely match the diversity gain and the corresponding unitary matrices to diagonalize the scheme are presented.

Finally, the third transmission mode would send only one set of four symbols to the four transmit antennas. The corresponding matrix $\mathbf{H}^H\mathbf{H}$ is not of full rank and therefore, cannot be inverted. The entries on its diagonal consist of two times diversity terms like $|h_{11}|^2 + |h_{12}|^2$. The decoding can be performed either in MMSE mode or with an ML decoder [24] allowing only for diversity of two but with a data rate of four. Gaining such insight, the following conjecture can be made.

Conjecture 1. *Given a wireless communications system with N_T transmit and N_R receive antennas in a flat Rayleigh fading environment with maximum diversity $N_R N_T$ (see also [25] for definition), an Alamoutization scheme can be found with diversity order D and data rate R, if $D \in \mathbb{N}$ and $R \in \mathbb{N}$ approximately factorizing the maximum diversity, that is, $DR \approx N_T N_R$.*

Note that this statement was not formulated in terms of a lemma since it may not be exactly true in the sense that exactly a diversity of say eight is obtained when actually only 6.4 is achieved. It is thus to apply with some care. On non-flat-fading channels, the UMTS transmission allows the diversity to increase by assigning a number of fingers to each major energy contribution in the impulse response. In this case, all finger values are combined in a correspondingly larger matrix $\mathbf{H}$. However, $\mathbf{H}^H\mathbf{H}$ remains of the same size as before. The various fingers only contribute to higher diversity gain allowing to utilize BLAST schemes in which $\mathbf{H}^H\mathbf{H}$ would not be of full rank in a flat Rayleigh scenario.

7. CONCLUSION

In this paper, several extensions to the Alamouti space-time block code supporting very high transmit and receiver diversity have been proposed and their performance is evaluated. By combining conventional BLAST and new extended Alamouti schemes, a trade-off between diversity gain (and thus QoS) and supported data rate is offered allowing very high flexibility while the receiver complexity is kept approximately proportional to the transmitted data rate.

Not considered in this paper is the influence of the modulation scheme on the diversity. It is well known [8] that a rank criterion on the modulation scheme needs to be satisfied in order achieve the full diversity. In QPSK transmission, this rank criterion is, for example, not satisfied in the four- and eight-antenna transmission schemes. In other words, for some transmitted symbols, the full diversity will not be achieved. One can exclude such symbols or use different modulation schemes. In [26, 27], the possibility to use offset QPSK was proposed. This can be implemented in UMTS without sacrificing much of the existent hardware solutions. Another possibility very suitable for UMTS is to work with feedback schemes. In [28, 29], it is shown for the 4×1 as well as the 8×1 antenna scheme that a very simple feedback scheme retransmitting only one or two bits can reinstall the full diversity.

APPENDICES

A. PROOFS FOR LEMMAS 2 AND 3

Starting from the definition of X in (19), it is observed that the squares in the denominator can be appended:

$$X + 1 = \frac{|h_1 + h_4|^2 + |h_2 - h_3|^2}{|h_1|^2 + |h_2|^2 + |h_3|^2 + |h_4|^2}. \tag{A.1}$$

In the case of i.i.d. complex-valued Gaussian distributed variables h_i, the two variates $h_1 + h_4$ and $h_2 + h_3$ become complex Gaussian distributed and independent of each other. They depend, however, on the variates in the nominator. Now, a linear orthogonal coordinate transformation is defined:

$$\begin{aligned} u &= \frac{(h_1 + h_4)}{\sqrt{2}}, & v &= \frac{(h_2 - h_3)}{\sqrt{2}}, \\ u' &= \frac{(h_1 - h_4)}{\sqrt{2}}, & v' &= \frac{(h_2 + h_3)}{\sqrt{2}}, \end{aligned} \tag{A.2}$$

such that $\sum_{i=1}^{4} |h_i|^2 = |u|^2 + |v|^2 + |u'|^2 + |v'|^2$ and

$$\frac{X+1}{2} = \frac{|u|^2 + |v|^2}{|u|^2 + |v|^2 + |u'|^2 + |v'|^2} = \frac{X_1^2}{X_1^2 + X_2^2}. \tag{A.3}$$

Generally, if X_1^2 and X_2^2 are independent random variables following chi-square distributions with ν_1 and ν_2 degrees of freedom, respectively, then $X_1^2/(X_1^2 + X_2^2)$ is said to follow a beta(p, q) distribution with $\nu_1 = 2p$ and $\nu_2 = 2q$ degrees of freedom and the probability density is given by $(1/\mathrm{B}(p,q))\xi^{p-1}(1-\xi)^{q-1}$, with $p = \nu_1/2$, $q = \nu_2/2$. This matches our case (A.3) for $\nu_1 = \nu_2 = 4$ and the probability density specializes to $6\xi(1-\xi)$. Transforming this back to X gives the probability density

$$f_X(x) = \begin{cases} \dfrac{3}{4}(1 - x^2) & \text{for } |x| < 1, \\ 0 & \text{elsewhere.} \end{cases} \tag{A.4}$$

The independency of X and $\eta = h^2$ can be established by transformation of variables starting from the two independent variates $Z_i = X_i^2$ defined above which are (up to a scaling) χ_4^2 distributed, that is, their joint probability density is given by

$$f_{Z_1 Z_2}(z_1, z_2) = \frac{1}{16} z_1 z_2 e^{-(z_1+z_2)/2} \quad \text{for } z_1 > 0,\ z_2 > 0. \tag{A.5}$$

The 2×2 transformation between the variates X, η and Z_1, Z_2 is derived from (A.3):

$$\left\{ \begin{array}{l} X = \dfrac{Z_1 - Z_2}{Z_1 + Z_2} \\ \eta = Z_1 + Z_2 \end{array} \right\}, \qquad \left\{ \begin{array}{l} Z_1 = \dfrac{1+X}{2}\eta \\ Z_2 = \dfrac{1-X}{2}\eta \end{array} \right\}. \tag{A.6}$$

The rules for transformation of variates result in the follow-

ing joint probability density for X, η:

$$f_{X,\eta}(x,\eta) = \frac{1}{64}(1-x^2)\eta^3 e^{-\eta/2} = f_X(x)f_\eta(\eta), \quad \text{for } |x|<1,\ \eta>0, \tag{A.7}$$

where f_X is given above and f_η is the χ^2_8-density rescaled to unit mean, that is,

$$f_\eta(\eta) = \begin{cases} \frac{128}{3}\eta^3 \mathrm{e}^{-4\eta} & \text{for } \eta>0, \\ 0 & \text{elsewhere.} \end{cases} \tag{A.8}$$

For the ZF receiver (where $\mu = 0$), it follows that the noise is increased by a factor of

$$\mathrm{E}[\delta_4] = \mathrm{E}\left[\frac{1}{1-X^2}\right] = \int_0^2 \lambda^{-1} f_X(\lambda-1)d\lambda = \frac{3}{2}. \tag{A.9}$$

For the general linear receiver with $\mu > 0$ (including the MMSE),

$$\mathrm{E}[\delta_4] = \frac{1}{2}\mathrm{E}\left[\frac{1+X}{(1+\mu/h^2+X)^2} + \frac{1-X}{(1+\mu/h^2-X)^2}\right] \tag{A.10}$$

is evaluated by exploiting independency of X and $\eta = h^2$:

$$\mathrm{E}[\delta_4] = \int_{\eta=0}^{\infty}\int_{x=-1}^{1} \frac{1+x}{(1+\mu/\eta+x)^2} f_X(x)f_\eta(\eta)dx\,d\eta. \tag{A.11}$$

The integration over x is straightforward. The remaining integral

$$\mathrm{E}[\delta_4] = 32\int_0^\infty \Big(2\eta^2 + 6\mu\eta - \mu(3\mu+4\eta)\log(2\eta+\mu) + \mu(3\mu+4\eta)\log\mu\Big)\eta e^{-4\eta}d\eta \tag{A.12}$$

is evaluated in terms of the exponential integral which leads to (32).

B. SOME PROPERTIES OF $\mathbf{H}\mathbf{H}^H$

In the following it will be shown that all entries of $\mathbf{H}^H\mathbf{H}$ are real valued. The proof is performed by induction. Using block matrix notation $\mathbf{H}^H\mathbf{H}$ at a certain level m equals

$$\mathbf{H}^H\mathbf{H} = \begin{bmatrix} \mathbf{H}_1^H\mathbf{H}_1 + \mathbf{H}_2^H\mathbf{H}_2 & \mathbf{H}_1^H\mathbf{H}_2 - \mathbf{H}_2^T\mathbf{H}_1^* \\ \mathbf{H}_2^H\mathbf{H}_1 - \mathbf{H}_1^T\mathbf{H}_2^* & \mathbf{H}_1^H\mathbf{H}_1 + \mathbf{H}_2^H\mathbf{H}_2 \end{bmatrix}. \tag{B.1}$$

Thus, if the property is given at the lower level scheme, $\mathbf{H}_1^H\mathbf{H}_1$ and $\mathbf{H}_2^H\mathbf{H}_2$ are real valued and so are the diagonal block matrices. In the next step, the recursion for the nondiagonal block matrix $\mathbf{H}_1^H\mathbf{H}_2 - \mathbf{H}_2^T\mathbf{H}_1^*$ is investigated. Assuming $\mathbf{H}_1$ is constructed by $\mathbf{H}_{11}$ and $\mathbf{H}_{12}$ and $\mathbf{H}_2$ in a similar manner by the matrices $\mathbf{H}_{21}$ and $\mathbf{H}_{22}$, then the term $\mathbf{H}_1^H\mathbf{H}_2$ is given by

$$\mathbf{H}_1^H\mathbf{H}_2 = \begin{bmatrix} \mathbf{H}_{11}^H\mathbf{H}_{21} + \mathbf{H}_{12}^T\mathbf{H}_{22}^* & \mathbf{H}_{11}^H\mathbf{H}_{22} - \mathbf{H}_{12}^T\mathbf{H}_{21}^* \\ \mathbf{H}_{12}^H\mathbf{H}_{21} - \mathbf{H}_{11}^T\mathbf{H}_{22}^* & \mathbf{H}_{12}^H\mathbf{H}_{22} + \mathbf{H}_{11}^T\mathbf{H}_{21}^* \end{bmatrix} \tag{B.2}$$

and the nondiagonal block matrix is obtained by such value minus its transposed form $\mathbf{H}_2^T\mathbf{H}_1^*$. Thus, every term $\mathbf{H}_1^H\mathbf{H}_2 - \mathbf{H}_2^T\mathbf{H}_1^*$ is replaced by a sum of terms of the form $\mathbf{H}_{kl}^H\mathbf{H}_{mn} - \mathbf{H}_{mn}^T\mathbf{H}_{kl}^*$. If the property holds for the level below, it also holds for the current level. To complete the induction argument, it has to be shown that the property also holds for the first level ($m = 1$). In this case, the diagonal elements are $|h_1|^2 + |h_2|^2$ and the nondiagonal values are $h_1^*h_2 \pm h_2h_1^*$, that is, either zero or $2\Re\{h_1^*h_2\}$. Thus, all entries are real valued. Note that due to the different signs occurring, it cannot be concluded that the terms become zero.

The second property is shown in [30]. For the third property, it is observed that every nondiagonal term X consists of either elements $(h_ih_k^* + h_i^*h_k)/h^2$ or $-(h_lh_m^* + h_l^*h_m)/h^2$. Thus building $X+1$ allows to consider $(h^2+h_ih_k^* + h_i^*h_k)/h^2$ and $(h^2 - h_lh_m^* + h_l^*h_m)/h^2$, further allowing to reorganize the terms into $(h_i + h_k)(h_i + h_k)^*/h^2$ and $(h_l + h_m)(h_l - h_m)^*/h^2$. By applying the same transformation as in Appendix A:

$$u = \frac{(h_i+h_k)}{\sqrt{2}}, \qquad v = \frac{(h_l-h_m)}{\sqrt{2}}, \qquad u' = \frac{(h_i-h_k)}{\sqrt{2}}, \qquad v' = \frac{(h_l+h_m)}{\sqrt{2}}, \tag{B.3}$$

the terms $X+1$ can be written in terms of independent Gaussian variables and the same rules as before apply. The resulting term then reads $(X+1)/2 = X_1^2/(X_1^2+X_2^2)$ with X_1 and X_2 being χ^2-distributed with $\nu = 2^m = N_T$ degrees of freedom each and

$$\frac{1}{\mathrm{B}(N_T/2, N_T/2)}\xi^{N_T/2-1}(1-\xi)^{N_T/2-1} \tag{B.4}$$

is obtained, resulting in the density (49) for X.

ACKNOWLEDGMENTS

The authors like to thank Maja Lončar, Lund University of Technology, Ralf Müller, ftw., as well as Gerhard Gritsch, Vienna University of Technology, for their helpful comments. This work was carried out with funding from **K** *plus* in the ftw. project C3 "Smart Antennas for UMTS Frequency Division Duplex" together with Infineon Technologies and Austrian Research Centers, Seibersdorf (ARCS).

REFERENCES

[1] Third Generation Partnership Project, "Transmitter diversity solutions for multiple antennas, version 1.0.2," 3GPP Technical Report 25.869, Technical Specification Group Radio Access Network, March 2002.

[2] Third Generation Partnership Project, "Multiple Input Multiple Output (mimo) antennae processing for HSDPA, version 1.1.0," Tech. Rep. 25.876, Technical Specification Group Radio Access Network, February 2002.

[3] A. F. Naguib, N. Seshadri, and A. R. Calderbank, "Increasing data rate over wireless channels," *IEEE Signal Processing Magazine*, vol. 17, no. 3, pp. 76–92, 2000.
[4] T. H. Liew and L. Hanzo, "Space-time codes and concatenated channel codes for wireless communications," *Proceedings of the IEEE*, vol. 90, no. 2, pp. 187–219, 2002.
[5] T. Matsumoto, J. Jlitaro, and M. Juntti, "Overview and recent challenges towards multiple-input multiple-output communications systems," *IEEE Vehicular Technology Society News*, vol. 50, no. 2, pp. 4–9, 2003.
[6] G. J. Foschini and M. J. Gans, "On limits of wireless communication in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.
[7] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999, Technical Memorandum, Bell Laboratories, Lucent Technologies, October 1998.
[8] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.
[9] B. Hassibi and B. M. Hochwald, "Linear dispersion codes," in *Proc. IEEE International Symposium on Information Theory*, p. 325, Wash, DC, USA, 2001.
[10] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1804–1824, 2002.
[11] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.
[12] B. Hochwald, T. L. Marzetta, and C. B. Papadias, "A transmitter diversity scheme for wideband CDMA systems based on space-time spreading," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 1, pp. 48–60, 2001.
[13] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 4th edition, 2001.
[14] J. H. Winters, J. Salz, and R. D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Trans. Communications*, vol. 42, no. 2–4, pp. 1740–1751, 1994.
[15] M. Rupp, C. F. Mecklenbräuker, and G. Gritsch, "High diversity with simple space time block-codes and linear receivers," in *Proc. IEEE Global Telecommunications Conference*, pp. 302–306, San Francisco, Calif, USA, December 2003.
[16] G. Gritsch, H. Weinrichter, and M. Rupp, "Understanding the BER performance of space-time block codes," in *Proc. IEEE Signal Processing Advances in Wireless Communications*, pp. 400–404, Rome, Italy, June 2003.
[17] C. C. Martin, J. H. Winters, and N. R. Sollenberger, "MIMO radio channel measurements: performance comparison of antenna configurations," in *Proc. IEEE 54th Vehicular Technology Conference*, vol. 2, pp. 1225–1229, Atlantic City, NJ, USA, October 2001.
[18] M. Rupp and C. F. Mecklenbräuker, "Improving transmission by MIMO channel structuring," in *Proc. IEEE International Conference on Communications*, vol. 5, pp. 3066–3070, Anchorage, Alaska, USA, May 2003.
[19] C. B. Papadias and G. J. Foschini, "A space-time coding approach for systems employing four transmit antennas," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 4, pp. 2481–2484, Salt Lake City, Utah, USA, 2001.
[20] I. S. Gradsteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, Academic Press, New York, NY, USA, 1980.
[21] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. International Symposium Signals, Systems, and Electronics*, pp. 295–300, Pisa, Italy, September 1998.
[22] G. Gritsch and H. Weinrichter, "Adaptive subspace modulation in spatially correlated MIMO channels," in *Proc. IEEE 13th International Symposium on Personal, Indoor, and Mobile Radio Communications*, vol. 4, pp. 1772–1776, Lisboa, Portugal, September 2002.
[23] G. Gritsch, H. Weinrichter, and M. Rupp, "Two adaptive coded MIMO systems exploiting partial channel knowledge at the transmitter," in *Proc. of SSC 04*, pp. 31–38, Erlangen, Germany, January 2004.
[24] R. van Nee, A. van Zelst, and G. Awater, "Maximum likelihood decoding in a space division multiplexing system," in *Proc. IEEE 51st Vehicular Technology Conference*, vol. 1, pp. 6–10, Tokyo, Japan, May 2000.
[25] J. B. Andersen, "Array gain and capacity for known random channels with multiple element arrays at both ends," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 11, pp. 2172–2178, 2000.
[26] M. Rupp and C. F. Mecklenbräuker, "On extended Alamouti schemes for space-time coding," in *Proc. 5th International Symposium on Wireless Personal Multimedia Communications*, pp. 115–119, Honolulu, Hawaii, USA, October 2002.
[27] M. M. da Silva and A. Correia, "Space time block coding for 4 antennas with coding rate 1," in *Proc. IEEE 17th International Symposium on Spread Spectrum Techniques and Applications*, pp. 318–322, Prague, Czech Republic, September 2002.
[28] B. Badic, M. Rupp, and H. Weinrichter, "Quasi-orthogonal space-time block codes for data transmission over four and eight transmit antennas with very low feedback rate," in *Proc. of SCC 04*, pp. 157–164, Erlangen, Germany, January 2004.
[29] B. Badic, M. Rupp, and H. Weinrichter, "Adaptive channel matched extended alamouti space-time code exploiting partial feedback," in *8th Conference on CIC*, p. 350, Seoul, October 2003.
[30] A. Stamoulis, Z. Liu, and G. B. Giannakis, "Space-time block-coded OFDMA with linear precoding for multirate services," *IEEE Trans. Signal Processing*, vol. 50, no. 1, pp. 119–129, 2002.

Christoph F. Mecklenbräuker was born in Darmstadt, Germany, in 1967. He received his Dipl.Ing. degree in electrical engineering from Vienna University of Technology in 1992 and the Dr.Ing. degree from Ruhr University of Bochum in 1998, respectively. His doctoral thesis on matched field processing was awarded with the Gert Massenberg Prize. He worked for the Mobile Networks Radio Department of Siemens AG, where he participated in the European framework of ACTS 90 "Frames." He was a delegate to the Third Generation Partnership Project (3GPP) and engaged in the standardization of the radio access network for UMTS. Since 2000, he holds a Senior Research position at the Telecommunications Research Center Vienna (ftw.) in the field of mobile communications. Currently, he gives a course at the Vienna Technical University on 3G mobile networks. He has authored around 60 papers in international journals and conferences, for which he has also served as a reviewer and holds 8 patents in the field of mobile cellular networks. His current research interests include antenna array and MIMO signal processing for mobile communications.

Markus Rupp received his Diploma degree in electrical engineering from FHS Saarbrücken, Germany, and the University of Saarbrücken in 1984 and 1988, respectively, and the Doctoral degree (summa cum laude) in acoustical echo compensation from the Technische Universität (TH) Darmstadt, Germany, in 1993. He was awarded a DAAD postdoctoral fellowship and spent the time from November 1993 until September 1995 in the Department of Electrical and Computer Engineering, University of California, Santa Barbara, working on a robustness theory for adaptive filters. From October 1995 until August 2001, he has been with Bell Labs, Lucent Technologies (before AT&T), Wireless Research Lab, Holmdel, NJ, working on wireless phones and implementation issues of wireless modems. In May 1999, he moved to The Netherlands, joining Bell Labs efforts in wireless LANs in Europe. In October 2001, he joined the Faculty of Electrical Engineering and Information Technology as a Full Professor at the TU Wien, Vienna, Austria. He is currently involved in rapid prototyping methods for wireless systems as well as MIMO coding and equalization techniques. He has more than 100 publications and patents in the fields of adaptive filters and wireless systems.

EURASIP Journal on Applied Signal Processing 2004:5, 676–684

Maximum MIMO System Mutual Information with Antenna Selection and Interference

Rick S. Blum

Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA 18015-3084, USA
Email: rblum@eecs.lehigh.edu

Received 31 December 2002; Revised 13 August 2003

Maximum system mutual information is considered for a group of interfering users employing single user detection and antenna selection of multiple transmit and receive antennas for flat Rayleigh fading channels with independent fading coefficients for each path. In the case considered, the only feedback of channel state information to the transmitter is that required for antenna selection, but channel state information is assumed at the receiver. The focus is on extreme cases with very weak interference or very strong interference. It is shown that the optimum signaling covariance matrix is sometimes different from the standard scaled identity matrix. In fact, this is true even for cases without interference if SNR is sufficiently weak. Further, the scaled identity matrix is actually that covariance matrix that yields worst performance if the interference is sufficiently strong.

Keywords and phrases: MIMO, antenna selection, interference, capacity.

1. INTRODUCTION

Multiple-input multiple-output (MIMO) channels formed using transmit and receive antenna arrays are capable of providing very high data rates [1, 2]. Implementation of such systems can require additional hardware to implement the multiple RF chains used in a standard multiple transmit and receive antenna array MIMO system. Employing antenna selection [3, 4] is one promising approach for reducing complexity while retaining a reasonably large fraction of the high potential data rate of a MIMO approach. One antenna is selected for each available RF chain. In this case, only the best set of antennas is used, while the remaining antennas are not employed, thus reducing the number of required RF chains. For cases with only a single transmit antenna where standard diversity reception is to be employed, this approach, known as "hybrid selection/maximum ratio combining," has been shown to lead to relatively small reductions in performance, as compared with using all receive antennas, for considerable complexity reduction [3, 4]. Clearly, antenna selection can be simultaneously employed at the transmitter and at the receiver in a MIMO system leading to larger reductions in complexity.

Employing antenna selection both at the transmitter and the receiver in a MIMO system has been studied very recently [5, 6, 7]. Cases with full and limited feedback of information from the receiver to the transmitter have been considered. The cases with limited feedback are especially attractive in that they allow antenna selection at the transmitter without requiring a full description of the channel or its eigenvector decomposition to be fed back. In particular, the only information fed back is the selected subset of transmit antennas to be employed. While cases with this limited feedback of information from the receiver to the transmitter have been studied in these papers, each assume that the transmitter sends a different (independent) equal power signal out of each selected antenna. Transmitting a different equal power signal out of each antenna is the optimum approach for the case where selection is not employed [8] but it is not optimum if antenna selection is used. The purpose of this paper is to find the optimum signaling. This problem is still unsolved to date. For simplicity, we ignore any delay or error that might actually be present in the feedback signal. We assume the feedback signal is accurate and instantly follows any changes in the environment.

Consider a system where cochannel interference is present from $L-1$ other users. We focus on the Lth user and assume each user employs n_t transmit antennas and n_r receive antennas. In this case, the vector of received complex baseband samples after matched filtering becomes

$$\mathbf{y}_L = \sqrt{\rho_L}\,\mathbf{H}_{L,L}\mathbf{x}_L + \sum_{j=1}^{L-1} \sqrt{\eta_{L,j}}\,\mathbf{H}_{L,j}\mathbf{x}_j + \mathbf{n}, \tag{1}$$

where $\mathbf{H}_{L,j}$ and $\mathbf{x}_j$ represent the normalized channel matrix and the normalized transmitted signal of user j, respectively. The signal-to-noise ratio (SNR) of user L is ρ_L and the interference-to-noise ratio (INR) for user L due to interference from user j is $\eta_{L,j}$. For simplicity, we assume all

of the interfering signals $\mathbf{x}_j$, $j = 1,\ldots,L-1$, are unknown to the receiver and we model each of them as being complex Gaussian distributed, the usual form of the optimum signal in MIMO problems. Then if we condition on $\mathbf{H}_{L,1},\ldots,\mathbf{H}_{L,L}$, the interference-plus-noise from (1), $\sum_{j=1}^{L-1}\sqrt{\eta_{L,j}}\mathbf{H}_{L,j}\mathbf{x}_j + \mathbf{n}$, is complex Gaussian distributed with the covariance matrix $\mathbf{R}_L = \sum_{j=1}^{L-1}\eta_{L,j}\mathbf{H}_{L,j}\mathbf{S}_j\mathbf{H}_{L,j}^H + \mathbf{I}_{n_r}$, where $\mathbf{S}_j$ denotes the covariance matrix of $\mathbf{x}_j$ and $\mathbf{I}_{n_r}$ is the covariance matrix of $\mathbf{n}$. Under this conditioning, the interference-plus-noise is whitened by multiplying $\mathbf{y}_L$ by $\mathbf{R}_L^{-1/2}$. After performing this multiplication, we can use results from [2, 8, 9] (see also [10, pp. 12–23, pp. 250,256]) to express the ergodic mutual information between the input and output for the user of interest as in the following:

$$\begin{aligned} &I(\mathbf{x}_L;(\mathbf{y}_L,\mathcal{H}))\\ &= E\left\{\log_2\left[\det\left(\mathbf{I}_{n_r}+\rho_L(\mathbf{R}_L^{-1/2}\mathbf{H}_{L,L})\mathbf{S}_L(\mathbf{R}_L^{-1/2}\mathbf{H}_{L,L})^H\right)\right]\right\} \quad (2)\\ &= E\{\log_2[\det(\mathbf{I}_{n_r}+\rho_L\mathbf{H}_{L,L}\mathbf{S}_L\mathbf{H}_{L,L}^H\mathbf{R}_L^{-1})]\} \end{aligned}$$

($\mathcal{H}$ reminds us of the assumed model for $\mathbf{H}_{L,1},\ldots,\mathbf{H}_{L,L}$). In (2), the identity $\det(\mathbf{I}+\mathbf{AB}) = \det(\mathbf{I}+\mathbf{BA})$ was used. If we wish to compute total system mutual information, we should find $\mathbf{S}_1,\ldots,\mathbf{S}_L$ to maximize

$$\begin{aligned} &\Psi(\mathbf{S}_1,\ldots,\mathbf{S}_L)\\ &= \sum_{i=1}^{L} I(\mathbf{x}_i;(\mathbf{y}_i,\mathcal{H}))\\ &= \sum_{i=1}^{L} E\left\{\log_2\left[\det\left(\mathbf{I}_{n_r}+\rho_i\mathbf{H}_{i,i}\mathbf{S}_i\mathbf{H}_{i,i}^H\right.\right.\right.\\ &\qquad\left.\left.\left.\times\left(\mathbf{I}_{n_r}+\sum_{j=1,\,j\neq i}^{L}\eta_{i,j}\mathbf{H}_{i,j}\mathbf{S}_j\mathbf{H}_{i,j}^H\right)^{-1}\right)\right]\right\}. \end{aligned} \quad (3)$$

Now, assume that each receiver selects $n_{sr} < n_r$ receive antennas and $n_{st} < n_t$ transmit antennas based on the channel conditions and feeds back the information to the transmitter.[1] Then the observations from the selected antennas follow the model in (1) with n_t and n_r replaced by n_{st} and n_{sr}, respectively, and $\mathbf{H}_{i,j}$ replaced by $\tilde{\mathbf{H}}_{i,j}$. The matrix $\tilde{\mathbf{H}}_{i,j}$ is obtained by eliminating those columns and rows of $\mathbf{H}_{i,j}$ corresponding to unselected transmit and receive antennas. Thus we can write $\tilde{\mathbf{H}}_{i,j} = g(\mathbf{H}_{i,j})$, where the function g will choose $\tilde{\mathbf{H}}_{i,j}$ to maximize the instantaneous (and thus also the ergodic) mutual information (or some related quantity for the signaling approach employed). In order to promote brevity, we will restrict attention in the rest of this paper to the case where $n_{st} = n_{sr}$ so we will only use the notation for n_{st}. We note that the majority of the results given carry over immediately for the case of $n_{st} \neq n_{sr}$, and since this will be obvious in these cases, we will not discuss this further.

It is important to note that we restrict attention to narrowband systems using single user detection, equal power (constant over time) for each user, and fixed definitions of the transmitting and receiving users. Future extensions which remove some assumptions are of great interest. However, as we will show, these assumptions lead to interesting closed form results which we believe give insight into the fundamental properties of MIMO with antenna selection.

In Section 2, we give a general discussion and some useful relationships used to study the convexity and concavity properties of the system mutual information. In Section 3, we study cases with weak interference. We follow this, in Section 4, with our results for strong interference. The results in Sections 3 and 4 are general for any $n_{st} = n_{sr}, n_t, n_r$, and L. Section 5 is devoted to numerical studies for the particular case of $n_r = n_t = 8$, $n_{sr} = n_{st} = L = 2$ to illustrate the agreement with the theory from Sections 3 and 4. The results in Section 5 also show that our asymptotic results give useful information for nonasymptotic cases as well. The paper concludes with Section 6.

2. GENERAL ANALYSIS OF SYSTEM MUTUAL INFORMATION

Clearly, the nature of the functional[2] $\Psi(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ will depend on the SNRs ρ_i, $i = 1,\ldots,L$, and the INRs $\eta_{i,j}$, $i,j = 1,\ldots,L$, $i \neq j$. This can be seen by considering the convexity and the concavity of $\Psi(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ as a function of $\mathbf{S}_1,\ldots,\mathbf{S}_L$. Towards this goal, we define a general convex combination of two possible solutions $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ and $(\hat{\mathbf{S}}_1,\ldots,\hat{\mathbf{S}}_L)$ as follows:

$$\begin{aligned} (\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L) &= (1-t)(\mathbf{S}_1,\ldots,\mathbf{S}_L) + t(\hat{\mathbf{S}}_1,\ldots,\hat{\mathbf{S}}_L)\\ &= (\mathbf{S}_1,\ldots,\mathbf{S}_L) + t((\hat{\mathbf{S}}_1,\ldots,\hat{\mathbf{S}}_L) - (\mathbf{S}_1,\ldots,\mathbf{S}_L))\\ &= (\mathbf{S}_1,\ldots,\mathbf{S}_L) + t(\mathbf{S}_1',\ldots,\mathbf{S}_L') \end{aligned} \quad (4)$$

for $0 \le t \le 1$ a scalar. Then $\Psi(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ is a convex function of $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ if [12]

$$\frac{d^2}{dt^2}\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L) \ge 0 \quad \forall\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L. \quad (5)$$

Similarly, $\Psi(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ is a concave function of $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ if

$$\frac{d^2}{dt^2}\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L) \le 0 \quad \forall\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L. \quad (6)$$

There are several useful known relationships for the derivative of a function of a matrix $\mathbf{\Phi}$ with respect to a scalar parameter t. In particular, we note that [13, Appendix A, pp. 1342, 1345, 1349, 1351, 1359, 1401]

$$\begin{aligned} \frac{d}{dt}\ln[\det(\mathbf{\Phi})] &= \operatorname{trace}\left[\mathbf{\Phi}^{-1}\left(\frac{d}{dt}\mathbf{\Phi}\right)\right],\\ \frac{d}{dt}\mathbf{\Phi}^{-1} &= -\mathbf{\Phi}^{-1}\left(\frac{d}{dt}\mathbf{\Phi}\right)\mathbf{\Phi}^{-1}. \end{aligned} \quad (7)$$

[1] The case where each user employs a different n_{st} and n_{sr} is also easy to handle.

[2] In the case without antenna selection [11], it is possible to argue that each $\mathbf{S}_j$ can be taken as diagonal. These arguments are based on the joint Gaussianity of the $\mathbf{H}_{ij}$ which does not hold after selection.

Assuming selection is employed, we can use (3) and (7) to find (interchanging a derivative and an expected value)

$$\frac{d}{dt}\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L) = \frac{1}{\ln(2)}\sum_{i=1}^{L} E\left\{\operatorname{trace}\left[\mathbf{Q}_i^{-1}\frac{d}{dt}\mathbf{Q}_i\right]\right\}, \quad (8)$$

where

$$\begin{aligned}\mathbf{Q}_i &= \mathbf{I}_{n_{st}} + \rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H\left(\mathbf{I}_{n_{st}} + \sum_{j=1,\,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\bar{\mathbf{S}}_j\tilde{\mathbf{H}}_{i,j}^H\right)^{-1} \quad (9)\\ &= \mathbf{I}_{n_{st}} + \rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H\tilde{\mathbf{Q}}_i^{-1},\end{aligned}$$

$$\frac{d}{dt}\mathbf{Q}_i = \rho_i\tilde{\mathbf{H}}_{i,i}\mathbf{S}_i'\tilde{\mathbf{H}}_{i,i}^H\tilde{\mathbf{Q}}_i^{-1} - \rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H\tilde{\mathbf{Q}}_i^{-1}\left(\frac{d}{dt}\tilde{\mathbf{Q}}_i\right)\tilde{\mathbf{Q}}_i^{-1}, \quad (10)$$

$$\frac{d}{dt}\tilde{\mathbf{Q}}_i = \sum_{j=1,\,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\mathbf{S}_j'\tilde{\mathbf{H}}_{i,j}^H. \quad (11)$$

A second derivative yields

$$\begin{aligned}\frac{d^2}{dt^2}\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L) &= \frac{1}{\ln(2)}\sum_{i=1}^{L} E\left\{\operatorname{trace}\left[\mathbf{Q}_i^{-1}\left(\frac{d^2}{dt^2}\mathbf{Q}_i\right)\right.\right. \quad (12)\\ &\qquad \left.\left. - \mathbf{Q}_i^{-1}\left(\frac{d}{dt}\mathbf{Q}_i\right)\mathbf{Q}_i^{-1}\left(\frac{d}{dt}\mathbf{Q}_i\right)\right]\right\}\end{aligned}$$

with

$$\begin{aligned}\frac{d^2}{dt^2}\mathbf{Q}_i = &-2\rho_i\tilde{\mathbf{H}}_{i,i}\mathbf{S}_i'\tilde{\mathbf{H}}_{i,i}^H\tilde{\mathbf{Q}}_i^{-1}\left(\frac{d}{dt}\tilde{\mathbf{Q}}_i\right)\tilde{\mathbf{Q}}_i^{-1}\\ &+2\rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H\tilde{\mathbf{Q}}_i^{-1}\left(\frac{d}{dt}\tilde{\mathbf{Q}}_i\right)\tilde{\mathbf{Q}}_i^{-1}\left(\frac{d}{dt}\tilde{\mathbf{Q}}_i\right)\tilde{\mathbf{Q}}_i^{-1}.\end{aligned} \quad (13)$$

3. OPTIMUM SIGNALING FOR WEAK INTERFERENCE

We can use (12) to investigate convexity and concavity for any particular set of SNRs ρ_i, $i = 1,\ldots,L$, and INRs $\eta_{i,j}$, $i, j = 1,\ldots,L$, $i \neq j$. We investigate extreme cases, weak or strong interference, to gain insight. The following lemma considers the case of very weak interference.

Lemma 1. *Assuming sufficiently weak interference, the best* $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ *(that maximizes the ergodic system mutual information) must be of the form*

$$\begin{aligned}&(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L)\\ &\quad = \alpha(\gamma_1\mathbf{I}_{n_{st}} + (1-\gamma_1)\mathbf{O}_{n_{st}},\ldots,\gamma_L\mathbf{I}_{n_{st}} + (1-\gamma_L)\mathbf{O}_{n_{st}}),\end{aligned} \quad (14)$$

where $\mathbf{O}_{n_{st}}$ *is an* n_{st} *by* n_{st} *matrix of all ones,* $\alpha = 1/n_{st}$*, and* $0 \le \gamma_i \le 1$, $i = 1,\ldots,L$.

Outline of the proof. For the case of very weak interference, we ignore terms which are multiples of $\eta_{i,j}$ (essentially, we set $\eta_{i,j} \to 0$ for $i = 1,\ldots,L$, $j = 1,\ldots,L$, and $j \neq i$) and we find $(d/dt)\tilde{\mathbf{Q}}_i = 0$ so that $(d^2/dt^2)\mathbf{Q}_i = 0$ which leads to

$$\begin{aligned}&\frac{d^2}{dt^2}\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L)\\ &= -\frac{1}{\ln(2)}\sum_{i=1}^{L} E\left\{\operatorname{trace}\left[\left(\mathbf{I}_{n_{st}} + \rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H\right)^{-1}\rho_i\tilde{\mathbf{H}}_{i,i}\mathbf{S}_i'\tilde{\mathbf{H}}_{i,i}^H\right.\right.\\ &\qquad \left.\left.\times\left(\mathbf{I}_{n_{st}} + \rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H\right)^{-1}\rho_i\tilde{\mathbf{H}}_{i,i}\mathbf{S}_i'\tilde{\mathbf{H}}_{i,i}^H\right]\right\}.\end{aligned} \quad (15)$$

Since $\bar{\mathbf{S}}_i$ is a covariance matrix, $(\mathbf{I}_{n_{st}} + \rho_i\mathbf{H}_{i,i}\bar{\mathbf{S}}_i\mathbf{H}_{i,i}^H)^{-1} = (\mathbf{U}^H\mathbf{U} + \mathbf{U}^H\boldsymbol{\Lambda}\mathbf{U})^{-1} = (\mathbf{U}(\mathbf{I}_{n_{st}} + \boldsymbol{\Lambda})^{-1}\mathbf{U}^H) = \mathbf{U}(\boldsymbol{\Omega})^2\mathbf{U}^H = \mathbf{U}(\boldsymbol{\Omega})\mathbf{U}^H\mathbf{U}(\boldsymbol{\Omega})\mathbf{U}^H$, where $\mathbf{U}$ is unitary and $\boldsymbol{\Lambda}$ and $\boldsymbol{\Omega}$ are diagonal matrices with nonnegative entries. Define $\mathbf{A} = \rho_i\mathbf{H}_{i,i}\mathbf{S}_i'\mathbf{H}_{i,i}^H$ and note that $\mathbf{A}^H = \mathbf{A}$ due to $\mathbf{S}_i'$ being a difference of two covariance matrices (easy to see using $\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$ expansion for each covariance matrix). Thus the trace in (15) can be written as $\operatorname{trace}[\mathbf{U}(\boldsymbol{\Omega})^2\mathbf{U}^H\mathbf{A}\mathbf{U}(\boldsymbol{\Omega})^2\mathbf{U}^H\mathbf{A}] = \operatorname{trace}[\mathbf{U}\boldsymbol{\Omega}\mathbf{U}^H\mathbf{A}\mathbf{U}(\boldsymbol{\Omega})^2\mathbf{U}^H\mathbf{A}\mathbf{U}\boldsymbol{\Omega}\mathbf{U}^H] = \operatorname{trace}[\mathbf{B}\mathbf{B}^H]$ since $\operatorname{trace}[\mathbf{C}\mathbf{D}] = \operatorname{trace}[\mathbf{D}\mathbf{C}]$ [13]. We see $\operatorname{trace}[\mathbf{B}\mathbf{B}^H]$ must be nonnegative since the matrix inside the trace is nonnegative-definite so that (15) implies that $\Psi(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ is concave. This will be true for sufficiently small $\eta_{i,j}$, $i, j = 1,\ldots,L$, $i \neq j$, relative to ρ_i, $i = 1,\ldots,L$. To recognize the significance of the concavity, we note that given any permutation matrix $\boldsymbol{\Pi}$, we know [8] that $\tilde{\mathbf{H}}_{i,j}$ has the same distribution as $\tilde{\mathbf{H}}_{i,j}\boldsymbol{\Pi}$ (switching the ordering or names of selected antennas cannot change the physical problem), so $\Psi(\boldsymbol{\Pi}\mathbf{S}_1\boldsymbol{\Pi}^H,\ldots,\boldsymbol{\Pi}\mathbf{S}_L\boldsymbol{\Pi}^H) = \Psi(\mathbf{S}_1,\ldots,\mathbf{S}_L)$. Let $\sum_{\boldsymbol{\Pi}}$ denote the sum over all the different permutation matrices and let N denote the number of terms in the sum. From concavity, $\Psi((1/N)\sum_{\boldsymbol{\Pi}}\boldsymbol{\Pi}\mathbf{S}_1\boldsymbol{\Pi}^H,\ldots,(1/N)\sum_{\boldsymbol{\Pi}}\boldsymbol{\Pi}\mathbf{S}_L\boldsymbol{\Pi}^H) \ge \Psi(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ [8] which implies that the optimum $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ must be of the form such that it is invariant to transforms by permutation matrices. This implies that the best $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ must be of the form given in (14). We refer the interested reader to [14] for a rigorous proof of this (taken from a single user case). □

Before considering specific assumptions on the SNR, we note the similarity of (14) to (4) with $(\mathbf{S}_1,\ldots,\mathbf{S}_L) = (1/n_{st})(\mathbf{O}_{n_{st}},\ldots,\mathbf{O}_{n_{st}})$, $(\hat{\mathbf{S}}_1,\ldots,\hat{\mathbf{S}}_L) = (1/n_{st})(\mathbf{I}_{n_{st}},\ldots,\mathbf{I}_{n_{st}})$, and $t = \gamma_1 = \cdots = \gamma_L$.

Small SNR

Thus we have determined the best signaling except for the unknown scalar parameters $\gamma_1,\ldots,\gamma_L$ which we now investigate. Generally, the best approach will change with SNR. First, consider the case of weak SNR for which the following lemma applies (recall we have now already focused on very weak or no interference).

Lemma 2. *Let* $\tilde{h}(p,p)_{i,j}$ *denote the* (i,j)*th entry of the matrix* $\tilde{\mathbf{H}}_{p,p}$ *and define* $\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L$ *from* (14)*. Assuming sufficiently*

weak interference and sufficiently weak SNR,

$$\frac{d}{d\gamma_p}\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L) = -\frac{1}{n_{st}\ln(2)}\rho_p E\left\{\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}\sum_{j'=1,\,j'\neq j}^{n_{st}}\tilde{h}^*(p,p)_{i,j}\tilde{h}(p,p)_{i,j'}\right\} \quad \text{for } p=1,\ldots,L. \tag{16}$$

Outline of the proof. Using the similarity of (14) to (4), $(d/d\gamma_p)\Psi$ can be seen to be the pth component of the sum in (8) with $(\mathbf{S}_1,\ldots,\mathbf{S}_L) = (1/n_{st})(\mathbf{O}_{n_{st}},\ldots,\mathbf{O}_{n_{st}})$, $(\hat{\mathbf{S}}_1,\ldots,\hat{\mathbf{S}}_L) = (1/n_{st})(\mathbf{I}_{n_{st}},\ldots,\mathbf{I}_{n_{st}})$, and $t=\gamma_p$. To assert the weak signal and interference assumptions, we set $\eta_{i,j}\to 0$ for all i,j and $\rho_i\to 0$ for all i and in this case we find

$$\mathbf{Q}_i^{-1}\frac{d}{dt}\mathbf{Q}_i \longrightarrow \frac{d}{dt}\mathbf{Q}_i \longrightarrow \rho_i\tilde{\mathbf{H}}_{i,i}\mathbf{S}_i'\tilde{\mathbf{H}}_{i,i}^H \tag{17}$$

and using (8) gives

$$\frac{d}{d\gamma_p}\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L) = \frac{1}{n_{st}\ln(2)}E\left\{\operatorname{trace}\left[\tilde{\mathbf{H}}_{p,p}(\mathbf{I}_{n_{st}}-\mathbf{O}_{n_{st}})\tilde{\mathbf{H}}_{p,p}^H\right]\right\}, \tag{18}$$

where the $n_{st}\times n_{st}$ matrix can be explicitly written as

$$\mathbf{I}_{n_{st}}-\mathbf{O}_{n_{st}} = \begin{pmatrix} 0 & -1 & \cdots & -1 & -1 & -1\\ -1 & 0 & -1 & \cdots & -1 & -1\\ -1 & -1 & 0 & -1 & \cdots & -1\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ -1 & -1 & \cdots & -1 & -1 & 0 \end{pmatrix}. \tag{19}$$

Explicitly carrying out the operations in (18) gives (16). □

Notice that without selection (in this case $\tilde{\mathbf{H}}_{p,p}=\mathbf{H}_{p,p}$), the quantity in (16) becomes zero under the assumed model for $\mathbf{H}_{p,q}$ (i.i.d complex Gaussian entries). Thus selection turns out to be an important aspect in the analysis. The following lemmas will be used with the result in Lemma 2 to develop the main result of this section.

Lemma 3. *Let $\tilde{h}(p,p)_{i,j}$ denote the (i,j)th entry of the matrix $\tilde{\mathbf{H}}_{p,p}$ and define $\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L$ from* (14). *Assuming sufficiently weak interference and sufficiently weak SNR,*

$$\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L) = \frac{1}{n_{st}\ln(2)}\sum_{p=1}^{L}\rho_p \times E\left\{\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}|\tilde{h}(p,p)_{i,j}|^2 + (1-\gamma_p)\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}\sum_{j'=1,\,j'\neq j}^{n_{st}}\tilde{h}^*(p,p)_{i,j}\tilde{h}(p,p)_{i,j'}\right\}. \tag{20}$$

Outline of the proof. Consider an $n_{st}\times n_{st}$ nonnegative definite matrix $\mathbf{A}$ and let $\lambda_1(\mathbf{A}),\ldots,\lambda_{n_{st}}(\mathbf{A})$ denote the eigenvalues of $\mathbf{A}$. For sufficiently weak SNR ρ_i, we can approximate $\ln[\det(\mathbf{I}+\rho_i\mathbf{A})] = \ln[\prod_{j=1}^{n_{st}}(1+\rho_i\lambda_j(\mathbf{A}))] = \sum_{j=1}^{n_{st}}\ln[1+\rho_i\lambda_j(\mathbf{A})] \approx \rho_i\sum_{j=1}^{n_{st}}\lambda_j(\mathbf{A}) = \rho_i\operatorname{trace}(\mathbf{A})$. Now, consider Ψ itself, from (3), for the set of covariance matrices in (14) and assume that selection is employed. Thus we consider the resulting Ψ as a function of $(\gamma_1,\ldots,\gamma_L)$ and we see

$$\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L) = \frac{1}{n_{st}\ln(2)}\sum_{p=1}^{L}\rho_p \times E\left\{\operatorname{trace}\left[\tilde{\mathbf{H}}_{p,p}[\gamma_p\mathbf{I}_{n_{st}}+(1-\gamma_p)\mathbf{O}_{n_{st}}]\tilde{\mathbf{H}}_{p,p}^H\right]\right\}. \tag{21}$$

Note that the $n_{st}\times n_{st}$ matrix can be explicitly written as

$$[\gamma_p\mathbf{I}_{n_{st}}+(1-\gamma_p)\mathbf{O}_{n_{st}}] = \begin{pmatrix} 1 & 1-\gamma_p & \cdots & 1-\gamma_p & 1-\gamma_p & 1-\gamma_p\\ 1-\gamma_p & 1 & 1-\gamma_p & \cdots & 1-\gamma_p & 1-\gamma_p\\ 1-\gamma_p & 1-\gamma_p & 1 & 1-\gamma_p & \cdots & 1-\gamma_p\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 1-\gamma_p & 1-\gamma_p & \cdots & 1-\gamma_p & 1-\gamma_p & 1 \end{pmatrix}. \tag{22}$$

Using (22) in (21) with further simplification gives (20). □

Lemma 4. *Assuming sufficiently weak interference and sufficiently weak SNR, the antenna selection that maximizes the ergodic system mutual information will make*

$$E\left\{\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}\sum_{j'=1,\,j'\neq j}^{n_{st}}\tilde{h}^*(p,p)_{i,j}\tilde{h}(p,p)_{i,j'}\right\} \tag{23}$$

positive.

Outline of the proof. First, consider the antenna selection approach for the pth link which maximizes the ergodic system mutual information in (20) when $\gamma_p=1$ in (14). Thus the selection approach will maximize the quantity in the pth term in the first sum in (20) when $\gamma_p=1$ by selecting antennas for each set of instantaneous channel matrices to make the terms inside the expected value as large as possible. It is important to note that the choice (if $\gamma_p=1$) depends only on the squared magnitude of elements of the channel matrices.

If we use this selection approach when $\gamma_p\neq 1$, then the terms multiplied by $(1-\gamma_p)$ in (20) will be averaged to zero due to the symmetry in the selection criterion. To see this, first note that the contribution to the ergodic mutual information due to the pth term is

$$\int\cdots\int_{h(p,p)_{1,1},\ldots,h(p,p)_{n_t,n_r}}\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}\sum_{j'=1,\,j'\neq j}^{n_{st}}\tilde{h}^*(p,p)_{i,j}\tilde{h}(p,p)_{i,j'} \times f_{h(p,p)_{1,1},\ldots,h(p,p)_{n_t,n_r}}(h(p,p)_{1,1},\ldots,h(p,p)_{n_t,n_r}) \times dh(p,p)_{1,1}\cdots dh(p,p)_{n_t,n_r} \tag{24}$$

times the constant $\rho_p/n_{st}\ln(2)$. In (24),

$$f_{h(p,p)_{1,1},\ldots,h(p,p)_{n_t,n_r}}(h(p,p)_{1,1},\ldots,h(p,p)_{n_t,n_r}) \tag{25}$$

is the probability density function of the channel coefficients prior to selection, the integral is over all values of the arguments and the selection rule $\tilde{\mathbf{H}} = g(\mathbf{H})$ is important in determining the integrand. If the optimum selection rule for (20) with $\gamma_p = 1$ will select a particular set of transmit and receive antennas for a particular instance of $h(p,p)_{1,1},\ldots,h(p,p)_{n_t,n_r}$, then due to symmetry, this same selection will also occur several more times as we run through all the possible values of $h(p,p)_{1,1},\ldots,h(p,p)_{n_t,n_r}$. Thus assume that terms with $|h(p,p)_{ij}|^2 = a$ and $|h(p,p)_{ij'}|^2 = b$ in (20) with $\gamma_p = 1$ are large enough to cause the corresponding antennas to be selected by the selection criterion trying to maximize (20) with $\gamma_p = 1$ for some set of $h(p,p)_{11},\ldots,h(p,p)_{n_{st},n_{st}}$. Then due to the symmetry,

$$\begin{aligned}
(\tilde{h}(p,p)^*_{ij}, \tilde{h}(p,p)_{ij'}) &= (\sqrt{a}e^{j\phi_a}, \sqrt{b}e^{j\phi_b}),\\
(\tilde{h}(p,p)^*_{ij}, \tilde{h}(p,p)_{ij'}) &= (\sqrt{a}e^{j\phi_a}, -\sqrt{b}e^{j\phi_b}),\\
(\tilde{h}(p,p)^*_{ij}, \tilde{h}(p,p)_{ij'}) &= (-\sqrt{a}e^{j\phi_a}, \sqrt{b}e^{j\phi_b}),\\
(h(p,p)^*_{ij}, h(p,p)_{ij'}) &= (-\sqrt{a}e^{j\phi_a}, -\sqrt{b}e^{j\phi_b})
\end{aligned} \tag{26}$$

will all appear in (24). Since each of these four possible values appear for four equal area (actually probability) regions in channel coefficient space, a complete cancellation of these terms results in (24). In fact, this leads to (24) averaging to zero. Thus if we use the selection approach that will maximize (20) with $\gamma_p = 1$, this is the best we can do.

However, if $\gamma_p \neq 1$, we can do better. Due to the cross terms in (20) in the term multiplied by $(1-\gamma_p)$, we can use selection to do better by modifying the selection approach. To understand the basic idea, let $\tilde{\mathbf{H}}'$ denote the matrix $\tilde{\mathbf{H}}_{p,p}$ for a particular selection of antennas and $\tilde{\mathbf{H}}''$ denote the same quantity for a different selection of antennas. Now consider two selection approaches which are the same except the second approach will choose $\tilde{\mathbf{H}}''$ in cases where

$$\begin{aligned}
&\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}} |\tilde{H}''_{ij}|^2 = \sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}} |\tilde{H}'_{ij}|^2,\\
&\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}\sum_{j'=1,\, j'\neq j}^{n_{st}} \tilde{H}''_{ij}\tilde{H}''^{*}_{ij'} > 0,
\end{aligned} \tag{27}$$

and (in the sum, both a term and its conjugate appear, giving a real quantity)

$$\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}\sum_{j'=1,\, j'\neq j}^{n_{st}} \tilde{H}'_{ij}\tilde{H}'^{*}_{ij'} < 0. \tag{28}$$

Assume the first selection approach is the one trying to maximize (20) with $\gamma_p = 1$ so it will just select randomly if

$$\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}} |\tilde{H}''_{ij}|^2 = \sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}} |\tilde{H}'_{ij}|^2, \tag{29}$$

since it ignores the cross terms in its selection.

From (20), the second selection approach will give larger instantaneous mutual information for each event where the selection is different. Since the probability of the event that makes the two approaches different is greater than zero under our assumed model, then the second antenna selection approach will lead to improvement (if $\gamma_p \neq 1$) and it will do this by making the term multiplied by $(1-\gamma_p)$ in (20) positive. Clearly the optimum selection scheme will be at least as good or better, so it must also give improvement by making the term multiplied by $(1-\gamma_p)$ in (20) positive. □

We are now ready to give the main result of this section.

Theorem 1. *Assuming sufficiently weak interference, sufficiently weak SNRs, and optimum antenna selection, the best* $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ *(that maximizes the ergodic system mutual information) uses*

$$(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L) = \frac{1}{n_{st}}(\mathbf{O}_{n_{st}},\ldots,\mathbf{O}_{n_{st}}). \tag{30}$$

Outline of the proof. The assumption of weak SNRs implies that ρ_p is small for all $1 \le p \le L$. In this case, optimum selection will attempt to make $E\{\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}\sum_{j'=1,\, j'\neq j}^{n_{st}} \tilde{h}^*(p,p)_{i,j}\tilde{h}(p,p)_{i,j'}\}$ as large as possible as shown in Lemma 3. Lemma 4 builds on Lemma 3 to show that optimum selection can always make $E\{\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}\sum_{j'=1,\, j'\neq j}^{n_{st}} \tilde{h}^*(p,p)_{i,j}\tilde{h}(p,p)_{i,j'}\}$ positive. Lemma 2 shows that $(d/d\gamma_p)\Psi$ is directly proportional to the negative of $E\{\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}\sum_{j'=1,\, j'\neq j}^{n_{st}} \tilde{h}^*(p,p)_{i,j}\tilde{h}(p,p)_{i,j'}\}$ which the selection is making positive and large. Thus it follows that $(d/d\gamma_p)\Psi$ is always negative which implies that the best solution employs $\gamma_p = 0$ since any increase in γ_p away from $\gamma_p = 0$ causes a decrease in Ψ. Since ρ_p is small for all p, the theorem follows. □

Large SNR

Now consider the case of large SNR, where the following theorem applies.

Theorem 2. *Assuming sufficiently weak interference, sufficiently large SNRs, and optimum antenna selection, the best* $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ *(that maximizes the ergodic system mutual information) uses*

$$(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L) = \frac{1}{n_{st}}(\mathbf{I}_{n_{st}},\ldots,\mathbf{I}_{n_{st}}). \tag{31}$$

Outline of the proof. Asserting the weak interference, large SNR assumption in (8) gives

$$\mathbf{Q}_i^{-1}\frac{d}{dt}\mathbf{Q}_i \longrightarrow \left(\rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H\right)^{-1}\rho_i\tilde{\mathbf{H}}_{i,i}\mathbf{S}_i'\tilde{\mathbf{H}}_{i,i}^H, \tag{32}$$

so that

$$\begin{aligned}
&(d/d\gamma_p)\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L)\\
&= \frac{1}{\ln(2)}E\Big\{\operatorname{trace}\Big[\Big(\tilde{\mathbf{H}}_{p,p}[\gamma_p\mathbf{I}_{n_{st}}+(1-\gamma_p)\mathbf{O}_{n_{st}}]\tilde{\mathbf{H}}_{p,p}^H\Big)^{-1}\\
&\qquad\times\Big(\tilde{\mathbf{H}}_{p,p}[\mathbf{I}_{n_{st}}-\mathbf{O}_{n_{st}}]\tilde{\mathbf{H}}_{p,p}^H\Big)\Big]\Big\}\\
&= \frac{1}{\ln(2)}E\Big\{\operatorname{trace}\Big[\Big[\tilde{\mathbf{H}}_{p,p}^H\Big]^{-1}[\gamma_p\mathbf{I}_{n_{st}}+(1-\gamma_p)\mathbf{O}_{n_{st}}]^{-1}\\
&\qquad\times[\mathbf{I}_{n_{st}}-\mathbf{O}_{n_{st}}]\tilde{\mathbf{H}}_{p,p}^H\Big]\Big\}\\
&= \frac{1}{\ln(2)}E\Big\{\operatorname{trace}\Big[(\gamma_p\mathbf{I}_{n_{st}}+(1-\gamma_p)\mathbf{O}_{n_{st}})^{-1}(\mathbf{I}_{n_{st}}-\mathbf{O}_{n_{st}})\Big]\Big\}\\
&= \frac{n_{st}(n_{st}-1)(\gamma_p-1)}{\gamma_p((n_{st}-1)\gamma_p-n_{st})\ln(2)} \geq 0
\end{aligned} \tag{33}$$

which is positive for $0 < \gamma_p < 1$ (since $(n_{st}-1)\gamma_p < n_{st}$) and zero if $\gamma_p = 1$. In (33), we used trace [**CD**] = trace [**DC**] [13]. Thus for the large SNR case (large ρ_p for all p) when the interference is very weak, the best signaling uses (14) with $\gamma_p = 1$. Since this is true for all p, the theorem follows. □

As a further comment on Theorem 2, we note that the proof makes it clear that if ρ_p is large only for certain p, then $\gamma_p = 1$ for those p only. Likewise, it is clear from Theorem 1 that if ρ_p is small only for certain p, then $\gamma_p = 0$ for those p only. Of course, this assumes weak interference. Thus we can image a case where the best signaling uses $\gamma_p = 1$ for some p and $\gamma_{p'} = 0$ for some $p' \neq p$ with proper assumptions on the corresponding $\rho_p, \rho_{p'}$. One can construct similar cases where only some of the $\eta_{i,j}$ are small and easily extend the results given here in a straight forward way.

4. STRONG INTERFERENCE

Now consider the other extreme of dominating interference where $\eta_{i,j}$, $i = 1,\ldots,L$, $j = 1,\ldots,L$, is large (compared to $\rho_1,\ldots,\rho_L$). The following lemma addresses the worst signaling to use.

Lemma 5. *Assuming sufficiently strong interference, the worst* $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ *(that minimizes the ergodic system mutual information) must be of the form*

$$\begin{aligned}
&(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L)\\
&\quad= \alpha(\gamma_1\mathbf{I}_{n_{st}}+(1-\gamma_1)\mathbf{O}_{n_{st}},\ldots,\gamma_L\mathbf{I}_{n_{st}}+(1-\gamma_L)\mathbf{O}_{n_{st}}),
\end{aligned} \tag{34}$$

where $\mathbf{O}_{n_{st}}$ *is an* n_{st} *by* n_{st} *matrix of all ones,* $\alpha = 1/n_{st}$*, and* $0 \leq \gamma_i \leq 1$, $i = 1,\ldots,L$.

Outline of the proof. Provided $\eta_{i,j}$ is sufficiently large, we can approximate (9) as

$$\mathbf{Q}_i = \mathbf{I}_{n_{st}} + \rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H\left(\mathbf{I}_{n_{st}} + \sum_{j=1,\,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\bar{\mathbf{S}}_j\tilde{\mathbf{H}}_{i,j}^H\right)^{-1} \approx \mathbf{I}_{n_{st}}. \tag{35}$$

After applying this to (12) and using (13) for large $\eta_{i,j}$ so that $\tilde{\mathbf{Q}}_i^{-1} \approx (\sum_{j=1,\,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\bar{\mathbf{S}}_j\tilde{\mathbf{H}}_{i,j}^H)^{-1}$, we find the first term inside the trace in (12) depends inversely on $\eta_{i,j}$, while the second term inside the trace in (12) depends inversely on $\eta_{i,j}^2$ so that the first term dominates for large $\eta_{i,j}$. Further, we can interchange the expected value and the trace in (12) so we are concerned with the expected value of (13). Now note that the first term in (13) consists of the product of a term $\mathbf{A} = \tilde{\mathbf{H}}_{i,i}\mathbf{S}_i'\tilde{\mathbf{H}}_{i,i}^H$ and another term depending on $\tilde{\mathbf{H}}_{i,j}$ for $j \neq i$. Now consider the expected value of (13) computed first as an expected value conditioned on $\{\tilde{\mathbf{H}}_{i,j},\ j \neq i\}$ and then this expected value is averaged over $\{\tilde{\mathbf{H}}_{i,j},\ j \neq i\}$. Now note that the conditional expected value of **A** becomes the zero matrix.[3] Thus the contribution from the first term in (13) averages to zero so that

$$\begin{aligned}
&\frac{d^2}{dt^2}\Psi(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L)\\
&\approx \frac{1}{\ln(2)}\sum_{i=1}^{L}\operatorname{trace}\left[E\left\{\frac{d^2}{dt^2}\mathbf{Q}_i\right\}\right]\\
&\approx \frac{1}{\ln(2)}\sum_{i=1}^{L}E\Bigg\{\operatorname{trace}\Bigg[2\rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H\\
&\qquad\times\left(\sum_{j=1,\,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\bar{\mathbf{S}}_j\tilde{\mathbf{H}}_{i,j}^H\right)^{-1}\\
&\qquad\times\left(\sum_{j=1,\,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\mathbf{S}_j'\tilde{\mathbf{H}}_{i,j}^H\right)\\
&\qquad\times\left(\sum_{j=1,\,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\bar{\mathbf{S}}_j\tilde{\mathbf{H}}_{i,j}^H\right)^{-1}\\
&\qquad\times\left(\sum_{j=1,\,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\mathbf{S}_j'\tilde{\mathbf{H}}_{i,j}^H\right)\\
&\qquad\times\left(\sum_{j=1,\,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\bar{\mathbf{S}}_j\tilde{\mathbf{H}}_{i,j}^H\right)^{-1}\Bigg]\Bigg\}
\end{aligned} \tag{36}$$

which is nonnegative. To see this, we can use a few of the same simplifications used previously. Expand the nonnegative definite matrices $2\rho_i\tilde{\mathbf{H}}_{i,i}\bar{\mathbf{S}}_i\tilde{\mathbf{H}}_{i,i}^H$ and $(\sum_{j=1,\,j\neq i}^{L}\eta_{i,j}\tilde{\mathbf{H}}_{i,j}\bar{\mathbf{S}}_j\tilde{\mathbf{H}}_{i,j}^H)^{-1}$ using the unitary matrix/eigenvalue expansions as done after (15). Then the matrix inside the expected value in (36) can be factored

[3] Recall $\mathbf{S}_i' = \hat{\mathbf{S}}_i - \mathbf{S}_i$ and use the appropriate eigenvector expansions, problem symmetry, and constraints on trace [$\hat{\mathbf{S}}_i$], trace [$\mathbf{S}_i$].

into $\mathbf{B}\mathbf{B}^H$ after manipulations similar to those used after (15). Thus $\Psi(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ is convex. Thus using the same permutation argument as used for the weak interference case, the result stated in the theorem follows. □

The following theorem builds on Lemma 5 to specify the exact $\gamma_1,\ldots,\gamma_L$ giving worst performance.

Theorem 3. *Assuming sufficiently strong interference and optimum antenna selection, the worst* $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ *(that minimizes the ergodic system mutual information) uses*

$$(\bar{\mathbf{S}}_1,\ldots,\bar{\mathbf{S}}_L) = \frac{1}{n_{st}}(\mathbf{I}_{n_{st}},\ldots,\mathbf{I}_{n_{st}}). \tag{37}$$

Outline of the proof. Consider $\Psi(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ for $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ of the form given by Lemma 5 which is (from (2) and (3))

$$\begin{aligned}
&\Psi(\mathbf{S}_1,\ldots,\mathbf{S}_L)\\
&= \sum_{i=1}^{L} E\{\log_2[\det(\mathbf{I}_{n_{st}} + \rho_i \tilde{\mathbf{H}}_{i,i}\mathbf{S}_i\tilde{\mathbf{H}}_{i,i}^H\mathbf{R}_i^{-1})]\}\\
&\approx \sum_{i=1}^{L} \rho_i E\{\operatorname{trace}[\tilde{\mathbf{H}}_{i,i}\mathbf{S}_i\tilde{\mathbf{H}}_{i,i}^H\mathbf{R}_i^{-1}]\}\\
&= \frac{1}{n_{st}\ln(2)}\sum_{p=1}^{L}\rho_p\\
&\quad\times E\left\{\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}|\hat{h}(p,p)_{i,j}|^2\right.\\
&\qquad\left.+(1-\gamma_p)\sum_{i=1}^{n_{st}}\sum_{j=1}^{n_{st}}\sum_{j'=1,\,j'\neq j}^{n_{st}}\hat{h}^*(p,p)_{i,j}\hat{h}(p,p)_{i,j'}\right\},
\end{aligned} \tag{38}$$

where the first simplification follows from large $\eta_{i,j}$ and the same simplifications used in (21). The second simplification follows from those in (20) but now $\hat{h}(p,p)_{i,j}$ denotes the (i,j)th entry of the matrix $\mathbf{R}_p^{-1/2}\tilde{\mathbf{H}}_{p,p}$. Now note that antenna selection will attempt to make the second term in the last line of (38), which multiplies the positive constant $1-\gamma_p$, as large and positive as it possibly can. In fact, it is easy to argue that antenna selection can always make this term positive as done previously for (20). We skip this since the problems are so similar. Thus we see that the best performance for $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ of the form given by Lemma 5 must be obtained for $\gamma_p = 0$ and the worst performance must occur at $\gamma_p = 1$. Since this is true for all p, the result in the theorem follows. □

The result in Theorem 3 tells us that the best signaling for cases without interference and selection is the worst for strong interference and selection. It appears that the best signaling for $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ of the form given by Lemma 5 (see the discussion in Theorem 3) may be the best signaling overall. However, it appears difficult to show this generally.

The following intuitive discussion gives some further insight. Due to convexity, the best performance will occur at a point as far away from the point giving worst performance $(\mathbf{S}_1,\ldots,\mathbf{S}_L) = (1/n_{st})(\mathbf{I}_{n_{st}},\ldots,\mathbf{I}_{n_{st}})$ as possible (recall that the $\gamma_1 = \cdots = \gamma_L = 1$ point gives the worst performance). Thus the best performance occurs for a point on the boundary of our space of feasible $(\mathbf{S}_1,\ldots,\mathbf{S}_L)$ and this point must be as far away from the point giving the worst performance as possible. One such point is $(\mathbf{S}_1,\ldots,\mathbf{S}_L) = (1/n_{st})(\mathbf{O}_{n_{st}},\ldots,\mathbf{O}_{n_{st}})$. It can be shown generally (for any n_{st}) that this solution is the farthest from $(\mathbf{S}_1,\ldots,\mathbf{S}_L) = (1/n_{st})(\mathbf{I}_{n_{st}},\ldots,\mathbf{I}_{n_{st}})$ (Frobenius norm). This follows because $(1/n_{st})\mathbf{O}_{n_{st}}$ is the farthest from $(1/n_{st})\mathbf{I}_{n_{st}}$. Note that $\mathbf{S}$ with one entry of 1 and the rest zero is equally far from $(1/n_{st})\mathbf{I}_{n_{st}}$ but numerical results in some specific cases indicate that the rate of increase in this direction is not as great as the rate of increase experienced by moving along the line $(\mathbf{S}_1,\ldots,\mathbf{S}_L) = \gamma(1/n_{st})(\mathbf{I}_{n_{st}},\ldots,\mathbf{I}_{n_{st}}) + (1-\gamma)(1/n_{st})(\mathbf{O}_{n_{st}},\ldots,\mathbf{O}_{n_{st}})$ away from $\gamma = 1$ towards $\gamma = 0$.

5. NUMERICAL RESULTS FOR $n_{st} = n_{sr} = L = 2$, $n_t = n_r = 8$

Consider the case of $n_{st} = n_{sr} = L = 2$, $n_t = n_r = 8$, $\eta_{1,2} = \eta_{2,1} = \eta$, and $\rho_1 = \rho_2 = \rho$ and assume that the optimum antenna selection (to optimize system mutual information) is employed. First consider the case of no interference and assume a set of covariance matrices of the form $(\mathbf{S}_1,\mathbf{S}_2) = \gamma(1/2)(\mathbf{I}_2,\mathbf{I}_2) + (1-\gamma)(1/2)(\mathbf{O}_2,\mathbf{O}_2)$. Thus since $\rho_1 = \rho_2 = \rho$ and $\eta_{1,2} = \eta_{2,1} = \eta$, we set $\gamma_1 = \gamma_2 = \gamma$. Figure 1 shows a plot of the γ giving the largest mutual information versus SNR, for SNR (ρ) ranging from −10 dB to +10 dB. We see that the best performance for very small ρ is obtained for $\gamma = 0$ which is in agreement with our analytical results given previously. For large ρ, the best signaling uses $\gamma = 1$ which is also in agreement with our analytical results given previously. Figure 1 shows that the switch from where $\gamma = 0$ is optimum to where $\gamma = 1$ is optimum is very rapid and occurs near $\rho = -3$ dB.

Now consider cases with possible interference. Again consider the case of $n_{st} = n_{sr} = L = 2$, $n_t = n_r = 8$, $\eta_{1,2} = \eta_{2,1} = \eta$, and $\rho_1 = \rho_2 = \rho$ and assume that the optimum antenna selection (to optimize system mutual information) is employed. To simplify matters, we constrain $\mathbf{S}_1 = \mathbf{S}_2$ in all cases shown. First we considered three specific signaling covariance matrices which are

$$\begin{aligned}
\mathbf{S}_1 = \mathbf{S}_2 &= \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix},\\
\mathbf{S}_1 = \mathbf{S}_2 &= \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix},\\
\mathbf{S}_1 = \mathbf{S}_2 &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.
\end{aligned} \tag{39}$$

We tried each of these for SNRs and INRs between −10 dB and +10 dB. Then we recorded which of the approaches provided the smallest and the largest system mutual information. These results can be compared with the analytical

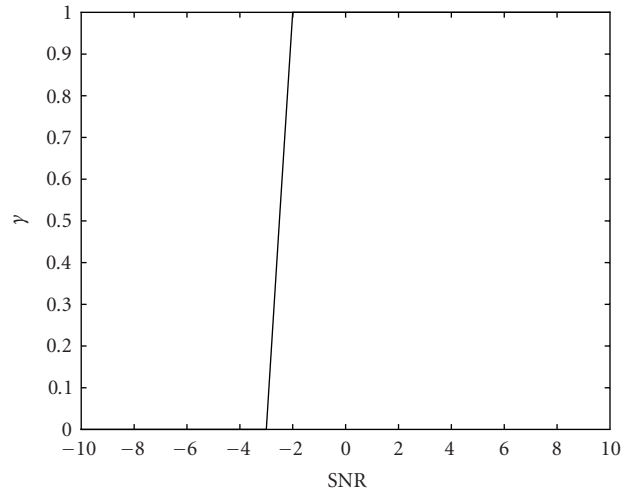


Figure 1: Optimum γ versus $\rho_1 = \rho_2 =$ SNR for cases with no interference and $n_{st} = n_{sr} = 2$, $n_t = n_r = 8$. Note that $\gamma = 0$ is the best for $-10\,\text{dB} < \text{SNR} < -3\,\text{dB}$ and $\gamma = 1$ is the best for $-2\,\text{dB} < \text{SNR} < 10\,\text{dB}$.

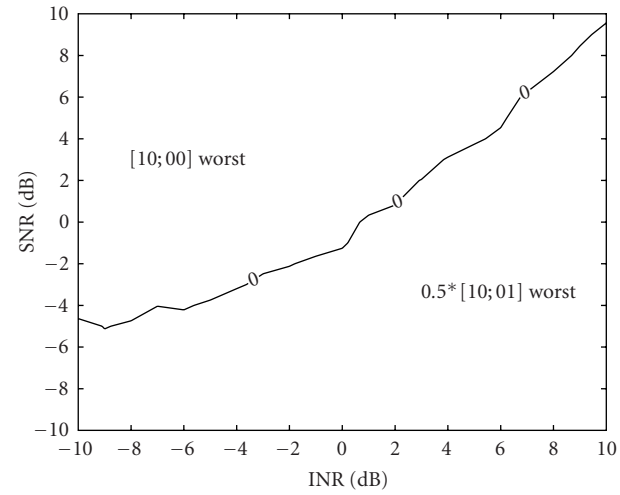


Figure 2: The worst signaling (of the three approaches) versus SNR and INR for $n_{st} = n_{sr} = L = 2$, $n_t = n_r = 8$, $\rho_1 = \rho_2 =$ SNR, and $\eta_{1,2} = \eta_{2,1} =$ INR.

results given in Sections 3 and 4 of this paper for weak and strong interference and SNR. Figure 2 shows the worst signaling we found versus SNR and INR for $\rho_1 = \rho_2 =$ SNR and $\eta_{1,2} = \eta_{2,1} =$ INR. For large INR, Figure 2 indicates that $\mathbf{S}_1 = \mathbf{S}_2 = (1/2)\mathbf{I}_2$ leads to worst performance which is in agreement with our analytical results given previously. Figure 2 also shows that either $\mathbf{S}_1 = \mathbf{S}_2 = (1/2)\mathbf{I}_2$ (for weak SNR) or (for large SNR) $\mathbf{S}_1 = \mathbf{S}_2$ with only one nonzero entry (a one which must be along the diagonal) will lead to worst performance for weak interference.

For weak interference, Figure 3 shows that the best performance is achieved by either $\mathbf{S}_1 = \mathbf{S}_2 = (1/2)\mathbf{O}_2$ (for weak

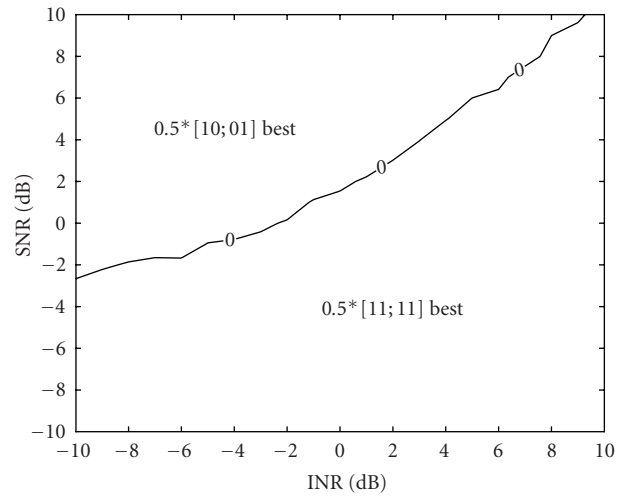


Figure 3: The best signaling (of the three choices) versus SNR and INR for $n_{st} = n_{sr} = L = 2$, $n_t = n_r = 8$, $\rho_1 = \rho_2 =$ SNR, and $\eta_{1,2} = \eta_{2,1} =$ INR.

SNR) or $\mathbf{S}_1 = \mathbf{S}_2 = (1/2)\mathbf{I}_2$ (for large SNR). This agrees with our analytical results presented previously. Figure 3 shows that the best performance is achieved by $\mathbf{S}_1 = \mathbf{S}_2 = (1/2)\mathbf{O}_2$ for large interference and this also agrees with our analytical results presented previously. We note that in the cases of interest (those for which we give analytical results), the difference in mutual information between the best and the worst approach in Figures 2 and 3 was about 1 to 3 bits/s/Hz.

We selected a few SNR-INR points sufficiently (greater than 2 dB) far from the dividing curves in Figures 2 and 3. For these points, we attempted to obtain further information on whether the approaches shown to be the best and worst in Figures 2 and 3 are actually the best and the worst of all valid approaches under the assumption that $\mathbf{S}_1 = \mathbf{S}_2$. We did this by evaluating the system mutual information for

$$\mathbf{S}_1 = \mathbf{S}_2 = \begin{pmatrix} a & b \\ b^* & \rho - a \end{pmatrix} \tag{40}$$

for various values of the real constant a and the complex constant b on a grid. When we evaluated (40) for all real a and b an a grid for a range of values consistent with the trace (power) and nonnegative definite enforcing constraints on $\mathbf{S}_1 = \mathbf{S}_2$, we did find the approaches in Figures 2 and 3 did indicate the overall best and worst approaches for the few cases we tried. Limited investigations involving complex b (here the extra dimension complicated matters, making strong conclusions difficult) indicated that these conclusions appeared to generalize to complex b also.

Partitioning the SNR-INR Plane

Based on Sections 3 and 4, we see that generally the space of all SNRs ρ_i, $i = 1, \ldots, L$, and INRs $\eta_{i,j}$, $i, j = 1, \ldots, L$, $i \neq j$, can be divided into three regions: one where the interference is considered weak (where Figure 1 and its generalization

apply), one where the interference is considered to dominate (where Figure 3 and its generalization apply), and a transition region between the two.

For the case with $n_{st} = n_{sr} = L = 2$, $n_t = n_r = 8$, $\eta_{1,2} = \eta_{2,1} = \eta$, and $\rho_1 = \rho_2 = \rho$, we have used (12) to study the three regions. We first evaluated (12) numerically using Monte Carlo simulations for a grid of points in SNR and INR space. The Monte Carlo simulations just described were calculated over a very fine grid over the region $-10\,\text{dB} \le \rho \le 10\,\text{dB}$ and $-10\,\text{dB} \le \eta \le 10\,\text{dB}$. For each given point in SNR and INR space, we evaluated (12) for many different choices of $(\mathbf{S}_1, \ldots, \mathbf{S}_L)$, $(\hat{\mathbf{S}}_1, \ldots, \hat{\mathbf{S}}_L)$, and the scalar t. We checked for a consistent positive or negative value for (12) for all $(\mathbf{S}_1, \ldots, \mathbf{S}_L)$, $(\hat{\mathbf{S}}_1, \ldots, \hat{\mathbf{S}}_L)$, and the scalar t on the discrete grid (quantize each scalar variable, including those in each entry of each matrix). In this way, we have viewed the approximate form of these three regions. We found that generally for points sufficiently far (more than 2 dB from closest curve) from the two dividing curves in Figures 2 and 3, the convexity and concavity follows that for the asymptotic case (strong or weak INR) in the given region. Thus the asymptotic results appear to give valuable conclusions about finite SNR and INR cases. Limited numerical investigations suggest this is true in other cases but the high dimensionality of the problem (especially for $n_{st}, n_{sr}, L > 2$) makes strong conclusions difficult.

6. CONCLUSIONS

We have analyzed the (mutual information) optimum signaling for cases where multiple users interfere while using single user detection and antenna selection. We concentrate on extreme cases with very weak interference or very strong interference. We have found that the best signaling is sometimes different from the scaled identity matrix that is best for no interference and no antenna selection. In fact, this is true even for cases without interference if SNR is sufficiently weak. Further, the scaled identity matrix is actually the covariance matrix that yields worst performance if the interference is sufficiently strong.

ACKNOWLEDGMENT

This material is based on research supported by the Air Force Research Laboratory under agreements no. F49620-01-1-0372 and no. F49620-03-1-0214 and by the National Science Foundation under Grant no. CCR-0112501.

REFERENCES

[1] J. H. Winters, "On the capacity of radio communication systems with diversity in a Rayleigh fading environment," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 5, pp. 871–878, 1987.

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[3] N. Kong and L. B. Milstein, "Combined average SNR of a generalized diversity selection combining scheme," in *Proc. IEEE International Conference on Communications*, vol. 3, pp. 1556–1560, Atlanta, Ga, USA, June 1998.

[4] M. Z. Win and J. H. Winters, "Analysis of hybrid selection/maximal-ratio combining in Rayleigh fading," *IEEE Trans. Communications*, vol. 47, no. 12, pp. 1773–1776, 1999.

[5] R. Nabar, D. Gore, and A. Paulraj, "Optimal selection and use of transmit antennas in wireless systems," in *Proc. International Conference on Telecommunications*, Acapulco, Mexico, May 2000.

[6] D. Gore, R. Nabar, and A. Paulraj, "Selection of an optimal set of transmit antennas for a low rank matrix channel," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 2785–2788, Istanbul, Turkey, June 2000.

[7] A. F. Molisch, M. Z. Win, and J. H. Winters, "Capacity of MIMO systems with antenna selection," in *Proc. IEEE International Conference on Communications*, vol. 2, pp. 570–574, Helsinki, Finland, June 2001.

[8] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[9] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Communications*, vol. 46, no. 3, pp. 357–366, 1998.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1991.

[11] R. S. Blum, "MIMO capacity with interference," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 793–801, 2003.

[12] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[13] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation and Modulation Theory*, John Wiley & Sons, New York, NY, USA, 2002.

[14] P. J. Voltz, "Characterization of the optimum transmitter correlation matrix for MIMO with antenna subset selection," submitted to IEEE Trans. Communications.

Rick S. Blum received his B.S. degree in electrical engineering from the Pennsylvania State University in 1984 and his M.S. and Ph.D. degrees in electrical engineering from the University of Pennsylvania in 1987 and 1991. From 1984 to 1991, he was a member of technical staff at General Electric Aerospace in Valley Forge, Pennsylvania, and he graduated from GE's Advanced Course in Engineering. Since 1991, he has been with the Electrical and Computer Engineering Department at Lehigh University in Bethlehem, Pennsylvania, where he is currently a Professor and holds the Robert W. Wieseman Chair in electrical engineering. His research interests include signal detection and estimation and related topics in the areas of signal processing and communications. He is currently an Associate Editor for the IEEE Transactions on Signal Processing and for IEEE Communications Letters. He was a member of the Signal Processing for Communications Technical Committee of the IEEE Signal Processing Society. Dr. Blum is a member of Eta Kappa Nu and Sigma Xi, and holds a patent for a parallel signal and image processor architecture. He was awarded an Office of Naval Research (ONR) Young Investigator Award in 1997 and a National Science Foundation (NSF) Research Initiation Award in 1992.

EURASIP Journal on Applied Signal Processing 2004:5, 685–695

Channel Estimation and Data Detection for MIMO Systems under Spatially and Temporally Colored Interference

Yi Song
Department of Electrical and Computer Engineering, Queen's University, Kingston, Ontario, Canada K7L 3N6
Email: songy@ee.queensu.ca

Steven D. Blostein
Department of Electrical and Computer Engineering, Queen's University, Kingston, Ontario, Canada K7L 3N6
Email: sdb@ee.queensu.ca

Received 20 December 2002; Revised 6 November 2003

The impact of interference on multiple-input multiple-output (MIMO) systems has recently attracted interest. Most studies of channel estimation and data detection for MIMO systems consider spatially and temporally white interference at the receiver. In this paper, we address channel estimation, interference correlation estimation, and data detection for MIMO systems under both spatially and temporally colored interference. We examine the case of one dominant interferer in which the data rate of the desired user could be the same as or a multiple of that of the interferer. Assuming known temporal interference correlation as a benchmark, we derive maximum likelihood (ML) estimates of the channel matrix and spatial interference correlation matrix, and apply these estimates to a generalized version of the Bell Labs Layered Space-Time (BLAST) ordered data detection algorithm. We then investigate the performance loss by not exploiting interference correlation. For a $(5, 5)$ MIMO system undergoing independent Rayleigh fading, we observe that exploiting both spatial and temporal interference correlation in channel estimation and data detection results in potential gains of 1.5 dB and 4 dB for an interferer operating at the same data rate and at half the data rate, respectively. Ignoring temporal correlation, it is found that spatial correlation accounts for about 1 dB of this gain.

Keywords and phrases: multiple-input multiple-output, interference, channel estimation, data detection.

1. INTRODUCTION

Wireless systems with multiple transmitting and receiving antennas have been shown to have a large Shannon channel capacity in a rich scattering environment [1, 2]. By transmitting parallel data streams over a multiple-input multi-output (MIMO) channel, it was shown that the Shannon capacity of the MIMO channel increases significantly with the number of transmitting and receiving antennas [2]. Layered space-time architectures were proposed for high-rate transmission in [3, 4]. Space-time coding techniques have also been investigated [5, 6].

While substantial research efforts have focussed on point-to-point MIMO link performance, the impact of interference on MIMO systems has received less interest. In a cellular environment, cochannel interference (CCI) from other cells exists due to channel reuse. In [7], channel capacities in the presence of spatially colored interference were derived under different assumptions of knowledge of the channel matrix and interference statistics at the transmitter. The impact of spatially colored interference on MIMO channel capacity was studied in [8, 9, 10]. The capacity of MIMO systems with interference in the limiting case of a large number of antennas was studied in [11]. The overall capacity of a group of users, each employing a MIMO link, was investigated in [12]. The output signal-to-interference power ratio (SIR) was analytically calculated in [13], when a single data stream is transmitted over independent Rayleigh MIMO channels. While the majority of the studies deals with channel capacity, in this paper we focus on the achievable symbol error rate performance of a MIMO link with interference.

Prior results on estimation of vector channels and spatial interference statistics for code division multiple access (CDMA) single-input multiple-output systems can be found in [14]. Most studies of channel estimation and data detection for MIMO systems assume spatially and temporally white interference. For example, in [15], maximum likelihood (ML) estimation of the channel matrix using training sequences was presented assuming temporally white interference. Assuming perfect knowledge of the channel matrix at

the receiver, ordered zero-forcing (ZF) and minimum mean-squared error (MMSE) detection were studied for both spatially and temporally white interference in [4, 16], respectively. However, in cellular systems, the interference is, in general, both spatially and temporally colored.

In this paper, we propose and study a new algorithm that jointly estimates the channel matrix and the spatial interference correlation matrix in an ML framework. We develop a multi-vector-symbol MMSE data detector that exploits interference correlation. In the case of a single dominant interferer and large signal-to-noise ratio (SNR), we show that spatial and temporal second-order interference statistics can be decoupled in the form of a matrix Kronecker product. In finite SNR, the decoupling of spatial and temporal statistics of interference-plus-noise is only an approximation. We also determine the conditions where this approximation breaks down.

Although temporal interference correlation is difficult to estimate in practice, our objectives are to determine the performance benchmark achieved if temporal correlation was known. As sources of temporal correlation, we consider cases in which the data rate of the desired user is either the same as or a multiple of that of the interferer. The new ML algorithm serves as a performance benchmark when temporal and spatial interference correlation are exploited in joint channel estimation and data detection. We also assess the performance improvement obtained in more practical cases where only part of the correlation information is exploited, including the performance obtained by assuming temporally white interference, that is, ignoring temporal correlation.

The paper is organized as follows. In Section 2, we present our system model of temporal and spatial interference. In Section 3, we derive ML estimates of channel and spatial interference correlation matrices assuming known temporal interference correlation. In Section 4, one-vector-symbol detection is extended to a multi-vector-symbol version which is used to exploit temporal interference correlation. In Section 5, we consider the case of one interferer and large SNR and assess the benefits of taking temporal and/or spatial interference correlation into account for channel estimation and data detection. We then examine the level of SNR at which the approximation of separate spatial and temporal interference-plus-noise statistics break down. In cases where the spatial and temporal correlation are not separable, the performance improvement obtained by exploiting the spatial correlation is evaluated. For reference, comparisons are made to the well-known direct matrix inversion (DMI) algorithm [17], generalized to multiple input signals, a batch method that does not require estimates of channel and spatial interference correlation matrices.

In this paper, the notation $(\cdot)^T$ refers to transpose, $(\cdot)^*$ refers to conjugate, $(\cdot)^\dagger$ refers to conjugate transpose, and $\mathbf{I}_N$ refers to an $N \times N$ identity matrix.

2. SYSTEM MODEL

We consider a single-user link consisting of N_t transmitting and N_r receiving antennas, denoted as (N_t, N_r). The desired user transmits data frame by frame. Each frame has M data vectors. The first N data vectors are used for training, so that the desired user's channel matrix and interference statistics can be estimated, and the remaining data vectors are for information transmission. In a slow flat fading environment, the received signal vector at time j is expressed as

$$\mathbf{y}_j = \mathbf{H}\mathbf{x}_j + \mathbf{n}_j, \quad j = 0, \ldots, M-1, \tag{1}$$

where $\mathbf{x}_j$ is the transmitted data vector, $\mathbf{H}$ is the $N_r \times N_t$ spatial channel gain matrix, and the interference vector $\mathbf{n}_j$ is zero-mean circularly symmetric complex Gaussian. We assume that the channel matrix $\mathbf{H}$ is fixed during one frame. This is a reasonable assumption since high-speed data services envisioned for MIMO systems are generally intended for low mobility users. By the same argument, it is also assumed that the interference statistics are fixed during one frame.

In practice, the interference may be both spatially and temporally correlated. We assume that the cross correlation between the interference vectors at time i and j is $E\{\mathbf{n}_i\mathbf{n}_j^\dagger\} = \mathbf{\Lambda}_M(i,j)\mathbf{R}$, where $\mathbf{\Lambda}_M(i,j)$ is the (i,j)th element of an $M \times M$ matrix $\mathbf{\Lambda}_M$. The (i,j)th element of matrix $\mathbf{R}$ is the correlation between the ith and jth elements of interference vector $\mathbf{n}_k$, $k \in 0, \ldots, M-1$. As a result, the covariance matrix of the concatenated interference vector $\bar{\mathbf{n}} = [\mathbf{n}_0^T \cdots \mathbf{n}_{M-1}^T]^T$ is

$$\begin{aligned} E\{\bar{\mathbf{n}}\bar{\mathbf{n}}^\dagger\} &= \begin{bmatrix} \mathbf{\Lambda}_M(0,0)\mathbf{R} & \cdots & \mathbf{\Lambda}_M(0,M-1)\mathbf{R} \\ \vdots & & \vdots \\ \mathbf{\Lambda}_M(M-1,0)\mathbf{R} & \cdots & \mathbf{\Lambda}_M(M-1,M-1)\mathbf{R} \end{bmatrix} \\ &= \mathbf{\Lambda}_M \otimes \mathbf{R}, \end{aligned} \tag{2}$$

where $\otimes$ denotes Kronecker product, and matrices $\mathbf{\Lambda}_M$ and $\mathbf{R}$ capture the temporal and spatial correlation of the interference, respectively. The above model implies that the spatial and temporal interference statistics are separable. The correlation matrices $\mathbf{\Lambda}_M$ and $\mathbf{R}$ are determined by the application-specific signal model. In Section 5, we provide an example in which the interference covariance matrix has the above Kronecker product form. When the interference statistics can only be approximated by (2), the conditions where this approximation breaks down are investigated in Section 5.4.3. In addition to interference correlation, we remark that a decoupled temporal and spatial correlation structure arises in the statistics of fading vector channels consisting of a mobile with one antenna and a base station with an antenna array [18].

3. JOINT ESTIMATION OF CHANNEL AND SPATIAL INTERFERENCE STATISTICS

During a training period of N vector symbols, we concatenate the received signal vectors, the training signal vectors and the interference vectors as $\bar{\mathbf{y}} = [\mathbf{y}_0^T \cdots \mathbf{y}_{N-1}^T]^T$, $\bar{\mathbf{x}} = [\mathbf{x}_0^T \cdots \mathbf{x}_{N-1}^T]^T$, and $\bar{\mathbf{n}} = [\mathbf{n}_0^T \cdots \mathbf{n}_{N-1}^T]^T$, respectively. The

received signal in (1) is rewritten as the vector

$$\bar{\mathbf{y}} = (\mathbf{I}_N \otimes \mathbf{H})\bar{\mathbf{x}} + \bar{\mathbf{n}}, \tag{3}$$

where $\bar{\mathbf{n}}$ is circularly symmetric complex Gaussian with zero-mean and covariance matrix $\mathbf{\Lambda}_N \otimes \mathbf{R}$. Assuming prior knowledge of temporal interference correlation matrix $\mathbf{\Lambda}_N$, we need to estimate channel matrix $\mathbf{H}$ and spatial interference correlation matrix $\mathbf{R}$. If $\mathbf{R}$ and $\mathbf{\Lambda}_N$ are nonsingular, the conditional probability density function (pdf) is

$$\begin{aligned}\Pr(\bar{\mathbf{y}}|\mathbf{H},\mathbf{R}) = {} & \frac{1}{\pi^{N\cdot N_r}\det(\mathbf{\Lambda}_N \otimes \mathbf{R})} \\ & \times \exp\Big\{ - [\bar{\mathbf{y}} - (\mathbf{I}_N \otimes \mathbf{H})\bar{\mathbf{x}}]^{\dagger} \\ & \qquad \times (\mathbf{\Lambda}_N \otimes \mathbf{R})^{-1}[\bar{\mathbf{y}} - (\mathbf{I}_N \otimes \mathbf{H})\bar{\mathbf{x}}]\Big\}.\end{aligned} \tag{4}$$

3.1. ML solution

The ML estimate of the pair of matrices $(\mathbf{H},\mathbf{R})$ is the value of $(\mathbf{H},\mathbf{R})$ that maximizes the conditional pdf in (4), which is equivalent to maximizing $\ln \Pr(\bar{\mathbf{y}}|\mathbf{H},\mathbf{R})$.

Letting $\mathbf{A}$ and $\mathbf{B}$ denote $m\times m$ and $n\times n$ square matrices, and using identities [19]

$$\begin{aligned}\det(\mathbf{A}\otimes\mathbf{B}) &= \det(\mathbf{A})^n \det(\mathbf{B})^m, \\ (\mathbf{A}\otimes\mathbf{B})^{-1} &= \mathbf{A}^{-1}\otimes\mathbf{B}^{-1},\end{aligned} \tag{5}$$

where $\mathbf{A},\mathbf{B}$ are nonsingular, it can be shown that maximizing (4) is equivalent to minimizing

$$\begin{aligned}f(\mathbf{H},\mathbf{R}) = {} & \ln\det(\mathbf{R}) \\ & + \frac{1}{N}[\bar{\mathbf{y}} - (\mathbf{I}_N \otimes \mathbf{H})\bar{\mathbf{x}}]^{\dagger} \\ & \times (\mathbf{\Lambda}_N^{-1} \otimes \mathbf{R}^{-1})[\bar{\mathbf{y}} - (\mathbf{I}_N \otimes \mathbf{H})\bar{\mathbf{x}}].\end{aligned} \tag{6}$$

Denoting the elements of $\mathbf{\Lambda}_N^{-1}$ as

$$\mathbf{\Lambda}_N^{-1} = \begin{bmatrix} \alpha_{0,0} & \cdots & \alpha_{0,N-1} \\ \vdots & & \vdots \\ \alpha_{N-1,0} & \cdots & \alpha_{N-1,N-1} \end{bmatrix}, \tag{7}$$

we rewrite (6) as

$$\begin{aligned}f(\mathbf{H},\mathbf{R}) & \\ = {} & \ln\det(\mathbf{R}) \\ & + \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\alpha_{i,j}(\mathbf{y}_i - \mathbf{H}\mathbf{x}_i)^{\dagger}\mathbf{R}^{-1}(\mathbf{y}_j - \mathbf{H}\mathbf{x}_j) \\ = {} & \ln\det(\mathbf{R}) \\ & + \operatorname{trace}\left\{\mathbf{R}^{-1}\frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\alpha_{i,j}(\mathbf{y}_i - \mathbf{H}\mathbf{x}_i)(\mathbf{y}_j - \mathbf{H}\mathbf{x}_j)^{\dagger}\right\}.\end{aligned} \tag{8}$$

To find the value of $(\mathbf{H},\mathbf{R})$ that minimizes $f(\mathbf{H},\mathbf{R})$ in (8), we set $\partial f(\mathbf{H},\mathbf{R})/\partial\mathbf{H} = \mathbf{0}$. Define the weighted sample correlation matrices[1] as

$$\begin{aligned}\tilde{\mathbf{R}}_{yy} &= \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\alpha_{i,j}\mathbf{y}_i\mathbf{y}_j^{\dagger}, \\ \tilde{\mathbf{R}}_{xy} &= \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\alpha_{i,j}\mathbf{x}_i\mathbf{y}_j^{\dagger}, \\ \tilde{\mathbf{R}}_{xx} &= \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\alpha_{i,j}\mathbf{x}_i\mathbf{x}_j^{\dagger}.\end{aligned} \tag{9}$$

Using the identities of matrix derivative [19], it can be shown [20] that (8) is minimized by

$$\hat{\mathbf{H}} = \tilde{\mathbf{R}}_{xy}^{\dagger}\tilde{\mathbf{R}}_{xx}^{-1}. \tag{10}$$

Setting $\partial f(\hat{\mathbf{H}},\mathbf{R})/\partial\mathbf{R} = \mathbf{0}$, it can also be shown that the estimate of spatial interference correlation matrix is given by

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\alpha_{i,j}(\mathbf{y}_i - \hat{\mathbf{H}}\mathbf{x}_i)(\mathbf{y}_j - \hat{\mathbf{H}}\mathbf{x}_j)^{\dagger} \tag{11}$$

$$= \tilde{\mathbf{R}}_{yy} - \hat{\mathbf{H}}\tilde{\mathbf{R}}_{xy}. \tag{12}$$

We remark that if $\tilde{\mathbf{R}}_{xy}$ and $\tilde{\mathbf{R}}_{xx}$ in (10) were known cross- and auto-correlation matrices, the estimate for $\mathbf{H}$ would represent the Wiener solution.

3.2. Special case: temporally white interference

If an interference is temporally white, with loss of generality, we may substitute $\mathbf{\Lambda}_N = \mathbf{I}_N$ into (9), (10), (11), and (12), and obtain estimates

$$\hat{\mathbf{H}}_w = \mathbf{R}_{xy}^{\dagger}\mathbf{R}_{xx}^{-1}, \tag{13}$$

$$\hat{\mathbf{R}}_w = \mathbf{R}_{yy} - \hat{\mathbf{H}}_w\mathbf{R}_{xy}, \tag{14}$$

where the subscript w indicates temporally white interference, and the sample correlation matrices are

$$\mathbf{R}_{yy} = \frac{1}{N}\sum_{i=0}^{N-1}\mathbf{y}_i\mathbf{y}_i^{\dagger}, \tag{15}$$

$$\mathbf{R}_{xy} = \frac{1}{N}\sum_{i=0}^{N-1}\mathbf{x}_i\mathbf{y}_i^{\dagger}, \tag{16}$$

$$\mathbf{R}_{xx} = \frac{1}{N}\sum_{i=0}^{N-1}\mathbf{x}_i\mathbf{x}_i^{\dagger}. \tag{17}$$

Note that $\hat{\mathbf{H}}_w$ in (13) is the same as the channel estimate used in [15].

[1] To distinguish weighted sample correlation matrices from conventional sample correlation matrices in Section 3.2, we denote the former by a tilde and the latter without a tilde.

3.3. Whitening filter interpretation

To obtain insight on the estimates in (10) and (12), we let the received signal vectors during the training period undergo a linear transformation where the transformed received signal vectors are

$$[\mathbf{y}'_0 \cdots \mathbf{y}'_{N-1}] = [\mathbf{y}_0 \cdots \mathbf{y}_{N-1}]\boldsymbol{\Lambda}_N^{-1/2}. \tag{18}$$

At the output of the transformation, we have

$$\mathbf{y}'_i = \mathbf{H}\mathbf{x}'_i + \mathbf{n}'_i, \quad i = 0, \ldots, N-1, \tag{19}$$

where the transformed training signal vectors and interference vectors are

$$\begin{aligned}[\mathbf{x}'_0 \cdots \mathbf{x}'_{N-1}] &= [\mathbf{x}_0 \cdots \mathbf{x}_{N-1}]\boldsymbol{\Lambda}_N^{-1/2},\\ [\mathbf{n}'_0 \cdots \mathbf{n}'_{N-1}] &= [\mathbf{n}_0 \cdots \mathbf{n}_{N-1}]\boldsymbol{\Lambda}_N^{-1/2},\end{aligned} \tag{20}$$

respectively. Concatenating the transformed interference vectors as $\tilde{\mathbf{n}}' = [\mathbf{n}_0'^T \cdots \mathbf{n}_{N-1}'^T]^T$, it can be shown that

$$\tilde{\mathbf{n}}' = (\boldsymbol{\Lambda}_N^{-1/2} \otimes \mathbf{I}_{N_r})\tilde{\mathbf{n}}, \tag{21}$$

where $\tilde{\mathbf{n}} = [\mathbf{n}_0^T \cdots \mathbf{n}_{N-1}^T]^T$. Since the covariance matrix of $\tilde{\mathbf{n}}$ is $\boldsymbol{\Lambda}_N \otimes \mathbf{R}$, the covariance matrix of $\tilde{\mathbf{n}}'$ is

$$\begin{aligned}\operatorname{cov}(\tilde{\mathbf{n}}') &= (\boldsymbol{\Lambda}_N^{-1/2} \otimes \mathbf{I}_{N_r}) \operatorname{cov}(\tilde{\mathbf{n}}) (\boldsymbol{\Lambda}_N^{-1/2} \otimes \mathbf{I}_{N_r})^\dagger\\ &= (\boldsymbol{\Lambda}_N^{-1/2} \otimes \mathbf{I}_{N_r}) (\boldsymbol{\Lambda}_N \otimes \mathbf{R}) (\boldsymbol{\Lambda}_N^{-1/2} \otimes \mathbf{I}_{N_r})\\ &= \mathbf{I}_N \otimes \mathbf{R},\end{aligned} \tag{22}$$

where we used $(\mathbf{A} \otimes \mathbf{B})^\dagger = \mathbf{A}^\dagger \otimes \mathbf{B}^\dagger$ and $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$ [19]. We also used the fact that the temporal correlation matrix $\boldsymbol{\Lambda}_N$ is symmetric, as well as $\boldsymbol{\Lambda}_N^{-1/2}$. From (22), it is obvious that the transformed interference vectors $\{\mathbf{n}'_0 \cdots \mathbf{n}'_{N-1}\}$ are temporally white with spatial correlation matrix $\mathbf{R}$.

As a result, we can estimate $\mathbf{H}$ and $\mathbf{R}$ from the sample correlation matrices of transformed signal vectors as in Section 3.2. The sample correlation matrix

$$\begin{aligned}\mathbf{R}_{y'y'} &= \frac{1}{N}\sum_{i=0}^{N-1} \mathbf{y}'_i \mathbf{y}_i'^\dagger\\ &= \frac{1}{N}[\mathbf{y}'_0 \cdots \mathbf{y}'_{N-1}][\mathbf{y}'_0 \cdots \mathbf{y}'_{N-1}]^\dagger\\ &= \frac{1}{N}[\mathbf{y}_0 \cdots \mathbf{y}_{N-1}]\boldsymbol{\Lambda}_N^{-1/2}\boldsymbol{\Lambda}_N^{-\dagger/2}[\mathbf{y}_0 \cdots \mathbf{y}_{N-1}]^\dagger\\ &= \frac{1}{N}[\mathbf{y}_0 \cdots \mathbf{y}_{N-1}]\boldsymbol{\Lambda}_N^{-1}[\mathbf{y}_0 \cdots \mathbf{y}_{N-1}]^\dagger = \tilde{\mathbf{R}}_{yy},\end{aligned} \tag{23}$$

which shows that the weighted sample correlation matrix of $\{\mathbf{y}_0 \cdots \mathbf{y}_{N-1}\}$ is equivalent to the sample correlation matrix of $\{\mathbf{y}'_0 \cdots \mathbf{y}'_{N-1}\}$. Similarly, the weighted sample correlation matrices $\tilde{\mathbf{R}}_{xy}$ and $\tilde{\mathbf{R}}_{xx}$ are equivalent to the sample correlation matrices $\mathbf{R}_{x'y'}$ and $\mathbf{R}_{x'x'}$, respectively. Therefore, the estimates in (10) and (12) can also be realized by first temporally whitening the interference, and then forming the estimates from the sample correlation matrices of the transformed signal vectors.

4. DATA DETECTION

We focus on ordered MMSE detection due to the better performance of MMSE compared to ZF detection [21]. For received signal vector $\mathbf{y}_i = \mathbf{H}\mathbf{x}_i + \mathbf{n}_i$, modifying the BLAST algorithm in [16], the steps of ordered MMSE detection of $\mathbf{x}_i$ from $\mathbf{y}_i$ with estimated channel and interference spatial correlation matrices are as follows:

Step 1. Initialization: set $k = 1$, $\mathbf{H}_k = \hat{\mathbf{H}}$, $\tilde{\mathbf{x}}_k = \mathbf{x}_i$, $\tilde{\mathbf{y}}_k = \mathbf{y}_i$.

Step 2. Calculate the estimation error covariance matrix $\mathbf{P}_k = (\mathbf{I}_{N_t+1-k} + \mathbf{H}_k^\dagger \hat{\mathbf{R}}^{-1}\mathbf{H}_k)^{-1}$. Find $m = \arg\min_j \mathbf{P}_k(j,j)$, where $\mathbf{P}_k(j,j)$ denotes the jth diagonal element of $\mathbf{P}_k$. Hence, the mth signal component of $\tilde{\mathbf{x}}_k$ has the smallest estimation error variance.

Step 3. Calculate the weighting matrix $\mathbf{A}_k = (\mathbf{I}_{N_t+1-k} + \mathbf{H}_k^\dagger \hat{\mathbf{R}}^{-1}\mathbf{H}_k)^{-1}\mathbf{H}_k^\dagger \hat{\mathbf{R}}^{-1}$. The mth element of $\tilde{\mathbf{x}}_k$ is estimated by $\hat{x}_k^m = Q(\mathbf{A}_k(m,:)\tilde{\mathbf{y}}_k)$, where $\mathbf{A}_k(m,:)$ denotes the mth row of matrix $\mathbf{A}_k$, and $Q(\cdot)$ denotes the slicing operation appropriate to the signal constellation.

Step 4. Assuming that the detected signal is correct, remove the detected signal from the received signal $\tilde{\mathbf{y}}_{k+1} = \tilde{\mathbf{y}}_k - \hat{x}_k^m \mathbf{H}_k(:,m)$, where $\mathbf{H}_k(:,m)$ denotes the mth column of $\mathbf{H}_k$.

Step 5. $\mathbf{H}_{k+1}$ is obtained by eliminating the mth column of matrix $\mathbf{H}_k$ and $\tilde{\mathbf{x}}_{k+1}$ is obtained by eliminating the mth component of vector $\tilde{\mathbf{x}}_k$.

Step 6. If $k < N_t$, increment k and go to Step 2.

We refer to this scheme as *one-vector-symbol detection*, as we detect $\mathbf{x}_i$ using $\mathbf{y}_i$ only.

When an interference is temporally colored, there may exist a performance to be gained by taking the temporal interference correlation into account. That is, we may use $\mathbf{y}_{N+1}, \ldots, \mathbf{y}_M$ to detect $\mathbf{x}_{N+1}, \ldots, \mathbf{x}_M$ jointly where N is the training length and M is the frame length. Due to the complexity of using all the received signal vectors and for simplicity of presentation, we consider a *two-vector-symbol detection* in which $(\mathbf{y}_i, \mathbf{y}_{i+1})$ is used to detect $(\mathbf{x}_i, \mathbf{x}_{i+1})$ jointly. The one-vector-symbol algorithm can be easily extended to the two-vector-symbol version by writing

$$\underbrace{\begin{bmatrix}\mathbf{y}_i\\ \mathbf{y}_{i+1}\end{bmatrix}}_{\check{\mathbf{y}}_i} = \underbrace{\begin{bmatrix}\mathbf{H} & \mathbf{0}\\ \mathbf{0} & \mathbf{H}\end{bmatrix}}_{\check{\mathbf{H}}} \underbrace{\begin{bmatrix}\mathbf{x}_i\\ \mathbf{x}_{i+1}\end{bmatrix}}_{\check{\mathbf{x}}_i} + \underbrace{\begin{bmatrix}\mathbf{n}_i\\ \mathbf{n}_{i+1}\end{bmatrix}}_{\check{\mathbf{n}}_i}. \tag{24}$$

With the estimated channel, an estimate of $\check{\mathbf{H}}$, denoted as $\hat{\check{\mathbf{H}}}$, can be obtained. Using the estimated spatial interference correlation and the known temporal interference correlation, we are able to estimate the covariance matrix of $\check{\mathbf{n}}_i$, denoted as $\hat{\check{\mathbf{R}}}$. Replacing $\mathbf{x}_i$, $\mathbf{y}_i$, $\hat{\mathbf{H}}$, and $\hat{\mathbf{R}}$ in the one-vector-symbol algorithm by $\check{\mathbf{x}}_i$, $\check{\mathbf{y}}_i$, $\hat{\check{\mathbf{H}}}$, and $\hat{\check{\mathbf{R}}}$, respectively, we obtain the two-vector-symbol detection algorithm.

5. APPLICATIONS

In this section, we apply the channel estimation in Section 3 and data detection in Section 4 to the case of a single-user link with one dominant cochannel interferer operating at different data rates.

5.1. System model

Consider a desired user with one dominant cochannel interferer. The assumption of one cochannel interferer can apply to cellular TDMA or FDMA systems when sectoring is used. For example, in 7-cell reuse systems, with 60 degree sectors, the number of cochannel interfering cells would be reduced to one [22]. We assume that the desired and interfering users have N_t and L transmitting antennas, respectively, and that there are N_r receiving antennas. Assuming that the thermal noise is small relative to the interference, we ignore the thermal noise in the problem formulation. An investigation of this assumption in channels with noise appears in Section 5.4.3. We also assume that over the duration of a transmitted frame, a randomly delayed replica of the interfering signal is transmitted continuously, and that the interference statistics do not change. This assumption may not hold for asynchronous packet transmission systems. In a slow flat fading environment, the vector signal at the receiving antennas is

$$\begin{aligned}\mathbf{y}(t) = {} & \sqrt{\frac{P_s T}{N_t}}\mathbf{H}\sum_{k=0}^{M-1}\mathbf{x}_k\tilde{g}(t-kT) \\ & + \sqrt{\frac{P_I T_I}{L}}\mathbf{H}_I\sum_{k=-\infty}^{\infty}\mathbf{b}_k\tilde{g}_I(t-kT_I-\tau),\end{aligned} \tag{25}$$

where M is the frame length, and $\mathbf{H}$ ($N_r \times N_t$) and $\mathbf{H}_I$ ($N_r \times L$) are the channel matrices of the desired and interfering users, respectively. The channel matrices are also assumed fixed over a frame and have independent realizations from frame to frame. The data transmission rates of the desired and interfering users are $1/T$ and $1/T_I$, respectively. The spectra of transmit impulse responses $\tilde{g}(t)$ and $\tilde{g}_I(t)$ are square root raised cosines with parameters T and T_I, respectively. The same roll-off factor, β, is assumed for both $\tilde{g}(t)$ and $\tilde{g}_I(t)$. The data vectors of the desired and interfering users are $\mathbf{x}_k$ ($N_t \times 1$) and $\mathbf{b}_k$ ($L \times 1$), respectively. We assume that the data symbols in $\mathbf{x}_k$'s and $\mathbf{b}_k$'s are mutually independent, zero mean, and with unit variance. We denote P_s and P_I as the transmit powers of the desired and interfering users, respectively. The delay of the interfering user relative to the desired user is τ, assumed to lie in $0 \le \tau < T_I$.

Passing $\mathbf{y}(t)$ in (25) through a filter matched to the transmit impulse response of the desired user, $\tilde{g}(t)$, the vector signal at the output of the matched filter is

$$\begin{aligned}\mathbf{y}_{\mathrm{MF}}(t) = {} & \sqrt{\frac{P_s T}{N_t}}\mathbf{H}\sum_{k=0}^{M-1}\mathbf{x}_k g(t-kT) \\ & + \sqrt{\frac{P_I T_I}{L}}\mathbf{H}_I\sum_{k=-\infty}^{\infty}\mathbf{b}_k g_I(t-kT_I-\tau),\end{aligned} \tag{26}$$

where $g(t) = \tilde{g}(t) * \tilde{g}(t)$, $g_I(t) = \tilde{g}_I(t) * \tilde{g}(t)$, and $*$ denotes convolution. As a result, $g(t)$ has a raised cosine spectrum and satisfies the Nyquist condition for zero intersymbol interference.

Assuming perfect synchronization for the desired user, as we sample the output of the matched filter (26) at time $t = jT$, we obtain

$$\mathbf{y}_j = \sqrt{\frac{P_s T}{N_t}}\mathbf{H}\mathbf{x}_j + \underbrace{\sqrt{\frac{P_I T_I}{L}}\mathbf{H}_I\sum_{k=-\infty}^{\infty}\mathbf{b}_k g_I(jT-kT_I-\tau)}_{\mathbf{n}_j}. \tag{27}$$

The interference vector $\mathbf{n}_j$ is zero mean since the data vector of interferer $\mathbf{b}_k$ is zero mean. Note that there is no intersymbol interference for the desired user. However, due to the interferer's delay and/or mismatch between the transmit and receive impulse responses, intersymbol interference exists for the interferer.

5.2. Interference statistics

The cross correlation between the interference vectors in (27) at time jT and qT is

$$\begin{aligned}E\{\mathbf{n}_j\mathbf{n}_q^\dagger\} = {} & \frac{P_I T_I}{L}\mathbf{H}_I \\ & \cdot E\left\{\left(\sum_{k_1=-\infty}^{\infty}\mathbf{b}_{k_1}g_I(jT-k_1T_I-\tau)\right)\right. \\ & \left.\times\left(\sum_{k_2=-\infty}^{\infty}\mathbf{b}_{k_2}^\dagger g_I(qT-k_2T_I-\tau)\right)\right\}\mathbf{H}_I^\dagger \\ = {} & \frac{P_I T_I}{L}\mathbf{H}_I\mathbf{H}_I^\dagger \\ & \cdot\sum_{k=-\infty}^{\infty}\{g_I(jT-kT_I-\tau)g_I(qT-kT_I-\tau)\},\end{aligned} \tag{28}$$

where the last equality is due to the facts that $E\{\mathbf{b}_{k_1}\mathbf{b}_{k_2}^\dagger\} = \mathbf{0}$ for $k_1 \ne k_2$ and $E\{\mathbf{b}_k\mathbf{b}_k^\dagger\} = \mathbf{I}_L$.

During a training period of N vector symbols, the covariance matrix of the concatenated interference vector $\bar{\mathbf{n}} = [\mathbf{n}_0^T \cdots \mathbf{n}_{N-1}^T]^T$ has the form of (2), where

$$\begin{aligned}\mathbf{\Lambda}_N(j,q) = {} & \sum_{k=-\infty}^{\infty}\{g_I(jT-kT_I-\tau)g_I(qT-kT_I-\tau)\}, \\ & 0 \le j,q \le N-1,\end{aligned} \tag{29}$$

$$\mathbf{R} = \frac{P_I T_I}{L}\mathbf{H}_I\mathbf{H}_I^\dagger. \tag{30}$$

The $N_r \times N_r$ spatial correlation matrix $\mathbf{R}$ is determined by the interferer's channel matrix. The $N \times N$ temporal correlation matrix $\mathbf{\Lambda}_N$ depends on parameters T and T_I, delay τ, and pulse $g_I(t)$; it can be calculated a priori if these parameters

are known. The temporal correlation is due to intersymbol interference in the sampled interfering signal. We remark that for the case of multiple interferers with the same delay, the covariance matrix of interference also has the form of (2).

We study temporal interference correlation in the cases where (1) the interferer has the same data rate as that of the desired signal ($T = T_I$) and (2) the data rate of the desired user is an integer multiple of that of the interferer ($T_I = mT$, $m > 1$).

5.2.1. Interferer at the same data rate as the desired signal

With $T = T_I$, $g_I(t)$ has a raised cosine spectrum and is given by [23]

$$g_I(t) = \operatorname{sinc}\left(\frac{\pi t}{T}\right)\frac{\cos(\pi\beta t/T)}{1-4\beta^2t^2/T^2}. \tag{31}$$

We note that $\mathbf{\Lambda}_N(j,q)$ depends on $j-q$. This indicates that the sequence consisting of interference vectors is stationary. Hence, the temporal correlation matrix is a symmetric Toeplitz. By appropriate truncation of the infinite series in (29), we can numerically calculate the temporal correlation matrix. For the case of $\beta = 1$, $T = 1$, and $\tau = 0.5$, the elements of the temporal correlation matrix are

$$\mathbf{\Lambda}_N(j,q) = \begin{cases} 0.5 & j = q, \\ 0.25 & |j-q| = 1 \quad \text{for } 0 \le j,q \le N-1, \\ 0 & \text{otherwise.} \end{cases} \tag{32}$$

5.2.2. Interferer at a lower data rate than the desired signal

It can be shown that $g_I(t)$ is given by

$$g_I(t) = \mathcal{F}^{-1}\left\{\sqrt{G_{rc,T_I}(f)}\sqrt{G_{rc,T}(f)}\right\}, \tag{33}$$

where $\mathcal{F}^{-1}$ denotes the inverse Fourier transform and $G_{rc,T}(f)$ is the raised cosine Fourier spectrum with parameter T and roll-off factor β. Unlike the case of the same data rate interferer where $\mathbf{\Lambda}_N(j,q)$ depends on $j-q$, in the case of lower data rate interferer, $\mathbf{\Lambda}_N(j,q)$ depends on the values of j and q. This indicates that the sequence consisting of interference vectors is cyclostationary [23, 24]. With $T_I = mT$, it can be shown that $\mathbf{\Lambda}_N(j,q)$ is periodic with period m, that is, $\mathbf{\Lambda}_N(j,q) = \mathbf{\Lambda}_N(j+m,q+m)$. As a result, the temporal correlation matrix $\mathbf{\Lambda}_N$ is symmetric, but not Toeplitz. Furthermore, for $N \ge m$, the number of nontrivial eigenvalues of $\mathbf{\Lambda}_N$ is $\lceil N/m \rceil$, where $\lceil \cdot \rceil$ rounds the argument to the nearest integer towards infinity [25]. For the case of $T_I = 2T$, $T = 1$, $\beta = 1$, $\tau = 0.25$, and training length $N = 6$, by numerical calculation of (29) with appropriate series truncation, the temporal correlation matrix is

$$\mathbf{\Lambda}_6 = \begin{bmatrix} 0.648 & 0.400 & -0.048 & -0.006 & -0.010 & -0.001 \\ 0.400 & 0.277 & 0.105 & 0.084 & 0.002 & 0.011 \\ -0.048 & 0.105 & 0.648 & 0.400 & -0.048 & -0.006 \\ -0.006 & 0.084 & 0.400 & 0.277 & 0.105 & 0.084 \\ -0.010 & 0.002 & -0.048 & 0.105 & 0.648 & 0.400 \\ -0.001 & 0.011 & -0.006 & 0.084 & 0.400 & 0.277 \end{bmatrix}. \tag{34}$$

Note that $\mathbf{\Lambda}_6$ in (34) is singular as the number of nontrivial eigenvalues is 3.

5.3. Data detection without estimating channel and interference

During a training period of N symbol vectors, instead of estimating the channel matrix and interference statistics, one can alternatively employ a least squares (LS) estimate of matrix $\mathbf{M}$ which minimizes the average estimation error

$$f_2(\mathbf{M}) = \operatorname{trace}\left\{\frac{1}{N}\sum_{i=0}^{N-1}(\mathbf{x}_i - \mathbf{M}\mathbf{y}_i)(\mathbf{x}_i - \mathbf{M}\mathbf{y}_i)^{\dagger}\right\}. \tag{35}$$

By setting $\partial f_2(\mathbf{M})/\partial\mathbf{M} = \mathbf{0}$, we obtain

$$\mathbf{M} = \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}, \tag{36}$$

where the sample correlation matrices $\mathbf{R}_{xy}$ and $\mathbf{R}_{yy}$ are defined in (16) and (15), respectively. The transmitted signal vector $\mathbf{x}_i$ is detected as $Q(\mathbf{M}\mathbf{y}_i)$, where $Q(\cdot)$ is the slicing operation appropriate to the signal constellation. We remark that (36) is the well-known DMI algorithm [17], generalized for multiple input signals. A significant loss in performance is expected for this LS detector, since without estimates of channel and spatial interference correlation matrices, iterative MMSE detection cannot be performed.

5.4. Simulation results

Monte Carlo simulations are used to assess the benefits of taking temporal and spatial interference correlation into account, for channel estimation and data detection in the case of one interferer. Although temporal interference correlation may be difficult to estimate in practice, we examine this as a benchmark and determine the performance loss due to ignoring this correlation. We evaluate average symbol error rates (SERs) in independent Rayleigh fading channels of rich scattering, that is, the elements in channel matrices $\mathbf{H}$ and $\mathbf{H}_I$ are independent, identically distributed (i.i.d.) zero-mean complex Gaussian with unit variance. We assume that the desired user has 5 transmitting and 5 receiving antennas, and the interfering user has 6 transmitting antennas.[2] Both the desired and interfering users employ uncoded quadrature phase shift keying (QPSK) modulation. The training signal vectors are columns of a fast Fourier transform (FFT) matrix

[2] For a nonsingular spatial interference correlation matrix, we set $N_r \le L$.

[16] to guarantee orthogonal training sequences from different transmitting antennas. We define $\mathrm{SIR(dB)} = 10\log P_s/P_I$. Without loss of generality, we set $P_I = 1$ in the simulation. The SERs of two cases are simulated: (1) interferer at the same data rate as the desired signal and (2) the data rate of the desired user is twice that of the interferer.

In Figures 1, 2, 3, and 4, with solid and dashed lines representing one- and two-vector-symbol data detection, respectively, we plot average SERs for the following cases:

(a) perfectly known channel parameters and interference statistics, with one-vector-symbol (curve 1) and two-vector-symbol (curve 2) detection;
(b) channel and spatial interference correlation matrices are estimated assuming known temporal interference correlation, with one-vector-symbol (curve 3) and two-vector-symbol (curve 4) detection;
(c) channel and spatial interference correlation matrices are estimated assuming temporally white interference, with one-vector-symbol detection (curve 5);
(d) only the channel matrix $\mathbf{H}$ is estimated assuming temporally white interference; an identity spatial interference correlation matrix is used in one-vector-symbol data detection (curve 6);
(e) LS estimate of the transmitted signal vector without ordered detection (Section 5.3) (curve 7).

We remark that cases (a) and (b) are benchmarks presented for reference, while case (d) corresponds to the well-known BLAST system in [4, 16].

5.4.1. Interferer at the same data rate as the desired signal

We examine the case of $T = 1$, $\beta = 1$, and $\tau = 1/2$, and the nonsingular temporal interference correlation matrix shown in (32). Figures 1 and 2 show the average SERs for training lengths $2N_t$ and $4N_t$, respectively. Comparing the LS detection (curve 7) with other methods, much lower SERs can be achieved by using ordered MMSE detection as expected.

Comparing curves 5 and 6, we observe that for a training length of $4N_t$ symbols, gains can be obtained by estimating spatial interference correlation. However, shorter training lengths such as $2N_t$ produce inaccurate estimates of spatial interference correlation which in turn do not yield any benefit over assuming spatially white interference. As expected, we observe that the improvement by taking into account estimated spatial correlation increases with longer training lengths.

Examining curves 3 and 5 in Figure 2, we observe that the improvement in taking temporal interference correlation into account in *channel estimation* is not significant. Moreover, this rate of improvement rapidly diminishes as the training length increases. This can be explained by noting that in estimating channel and spatial interference correlation matrices for temporally colored interference, the received signal vectors first undergo a transformation which temporally whitens the interference vectors as discussed in Section 3.3. Since the temporal correlation in (32) drops quickly to zero after one time lag, the benefit in temporal whitening of interference vectors is not significant, especially for long training lengths.

By comparing curves 3 and 4 in Figure 2, there is a slight improvement in using two-vector-symbol over one-vector-symbol detection. This implies that not much gain can be achieved by taking temporal interference correlation into account in *data detection*, owing to the low temporal correlation. Due to better estimates of channel and interference spatial correlation matrices obtained with a longer training length, the performance gap between curves 3 and 4 should increase as the training length increases.

By comparing curves 4 and 6 in Figure 2, we observe a 1.5 dB gain in SIR obtained by estimating spatial interference correlation and taking explicit advantage of known temporal interference correlation in channel estimation and data detection using a training length of $4N_t$. About 1 dB of that gain is due to the estimation of spatial interference correlation, and the remaining 0.5 dB gain is due to exploiting temporal interference correlation in channel estimation and data detection.

5.4.2. Interferer at a lower data rate than the desired signal

We examine the case of $T_I = 2T$, $T = 1$, $\beta = 1$, $\tau = 0.25$ and the temporal interference correlation matrix for training length $N = 6$ shown in (34). Recall that the temporal correlation matrix for the lower-data-rate-interferer case is singular. To avoid the singularity, the diagonal elements of $\mathbf{\Lambda}_N$ are increased by a small amount; hence, the temporal correlation matrix used for channel estimation may be modified to $\mathbf{\Lambda}_N + \delta\mathbf{I}_N$ within the proposed framework. In our simulation, we chose $\delta = 0.01$.

The same set of average SER curves as in the same-data-rate-interferer case are simulated. Figures 3 to 4 show the SERs for different training lengths. As in the case of the same-data-rate interferer, curve 7 illustrates the poor performance without ordered detection. Curves 5 and 6 suggest that for short training lengths it is better to estimate only the channel matrix and assume spatially white interference in data detection; however, for moderately long training lengths, gains can be obtained by estimating spatial interference correlation.

By examining curves 3 and 5 in Figure 4, we observe that the improvement in taking temporal interference correlation into account in *channel estimation*, although larger than that in the same-data-rate-interferer case due to the high temporal correlation in the lower-data-rate-interferer case, is still not that significant.

In contrast to the same-data-rate-interferer case, curves 3 and 4 in Figure 4 show that the improvement of two-vector-symbol over one-vector-symbol detection is significant due to the higher temporal interference correlation. This implies that a significant gain can be achieved by taking the known temporal interference correlation into account in *data detection* for the lower-data-rate-interferer case.

By comparing curves 4 and 6 in Figure 4, for the training length $4N_t$, there is a total of 4 dB gain in SIR by estimating spatial interference correlation and taking advantage of the

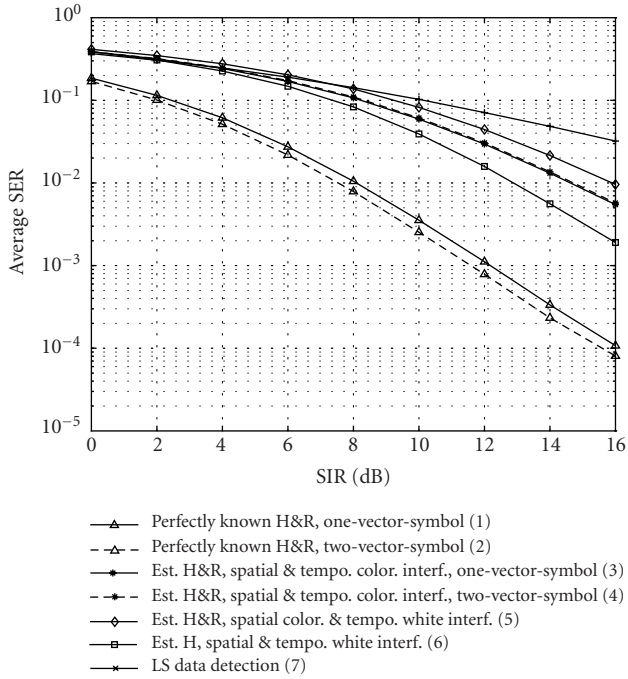


FIGURE 1: Average SER versus SIR with $N_t = N_r = 5$, $L = 6$, and training length $2N_t$ under independent Rayleigh fading. Both the desired and the interfering users are at the same data rate.

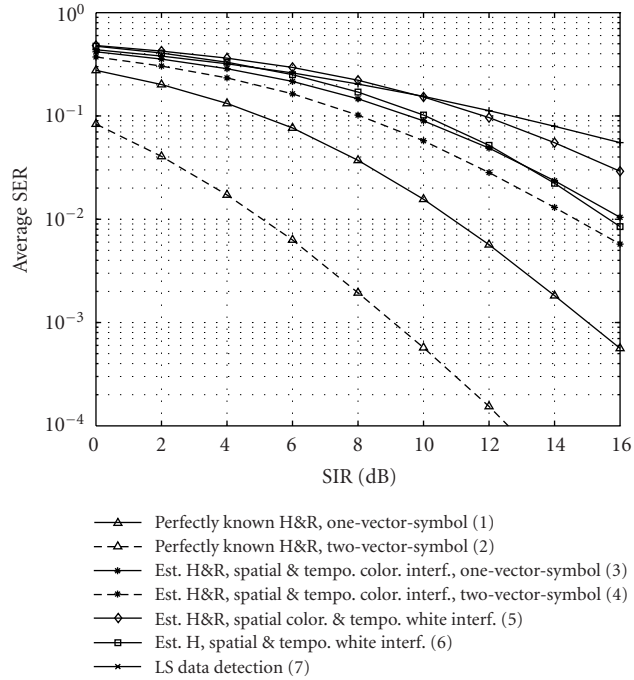


FIGURE 3: Average SER versus SIR with $N_t = N_r = 5$, $L = 6$, and training length $2N_t$ under independent Rayleigh fading. The data rate of the desired user is twice that of the interfering user.

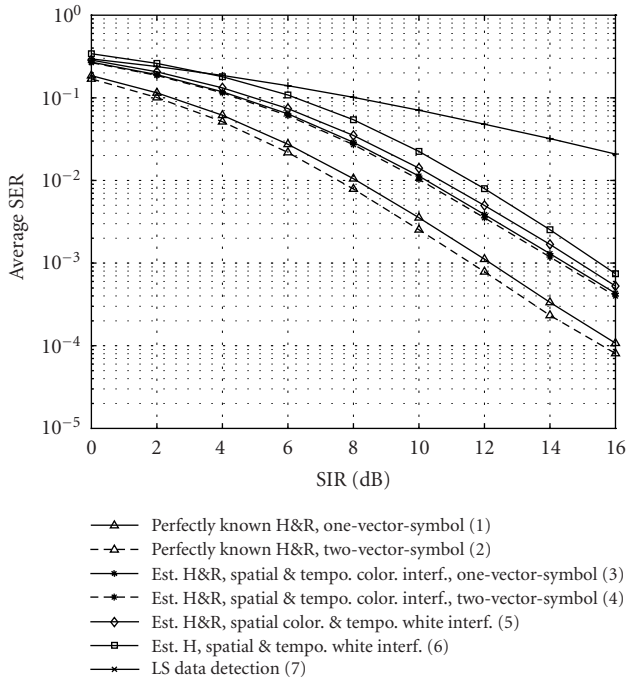


FIGURE 2: Average SER versus SIR with $N_t = N_r = 5$, $L = 6$, and training length $4N_t$ under independent Rayleigh fading. Both the desired and the interfering users are at the same data rate.

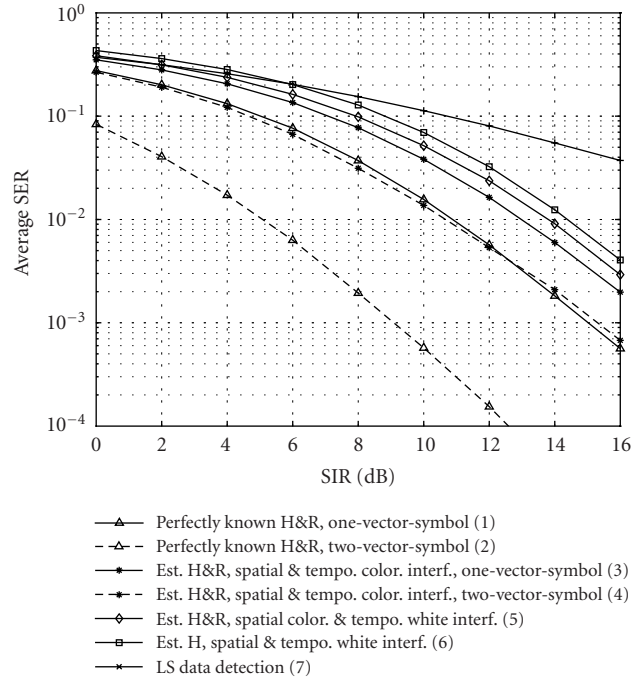


FIGURE 4: Average SER versus SIR with $N_t = N_r = 5$, $L = 6$, and training length $4N_t$ under independent Rayleigh fading. The data rate of the desired user is twice that of the interfering user.

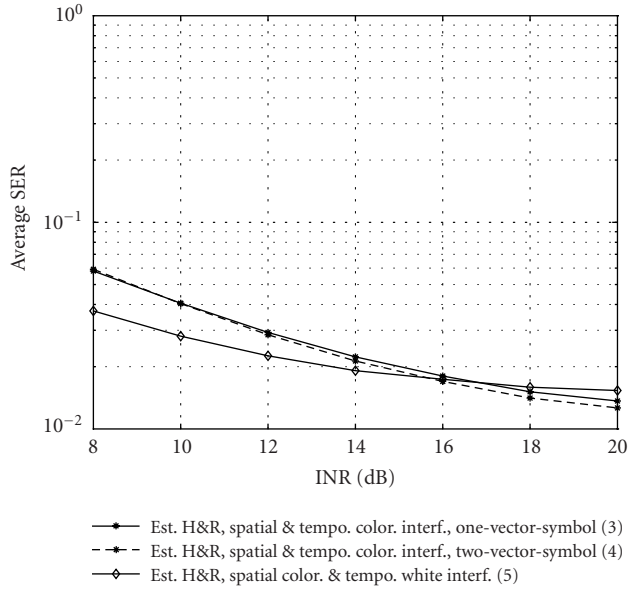


FIGURE 5: Average SER versus INR with $N_t = N_r = 5$, $L = 6$, SIR=10 dB, and training length $4N_t$ under independent Rayleigh fading. Both the desired and the interfering users are at the same data rate.

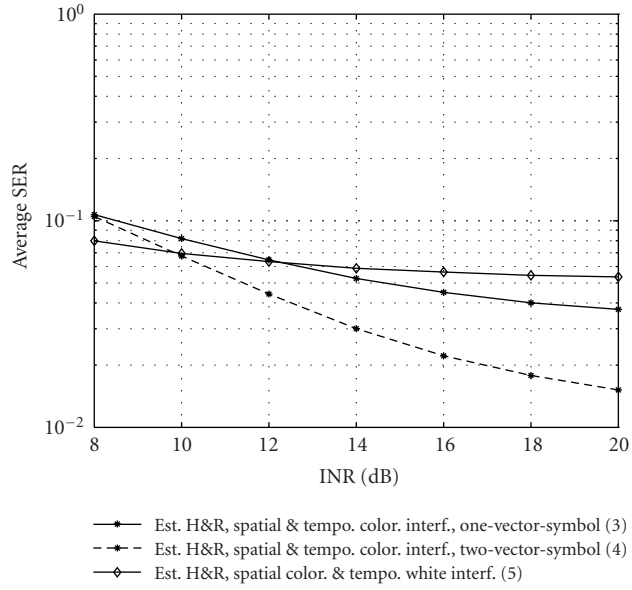


FIGURE 6: Average SER versus INR with $N_t = N_r = 5$, $L = 6$, SIR=10 dB, and training length $4N_t$ under independent Rayleigh fading. The data rate of the desired user is twice that of the interfering user.

known temporal interference correlation in channel estimation and data detection. About 3.5 dB of the gain is due to exploiting temporal interference correlation in channel estimation and data detection.

5.4.3. Effect of model mismatch

With one interferer and a finite SNR, the interference-plus-noise statistics can only be approximately modelled using a Kronecker product. Here, we investigate when this approximation breaks down. We model thermal noise as a zero-mean circularly symmetric complex Gaussian vector with covariance matrix $\sigma^2\mathbf{I}_{N_r}$, that is, independent from antenna to antenna, with noise power σ^2 on each antenna. We define (interference-to-noise power ratio) INR $= 10\log P_I/\sigma^2$, where $P_I = 1$ is used in the simulations. For the case of an interferer at the same data rate and using a training length $4N_t$, we observe in curves 3 and 5 in Figure 5 that, at INRs below 17 dB, taking interference temporal correlation into account appears not to be of benefit. Figure 6 shows the corresponding comparison for the case of the lower-data-rate interferer. In this case, temporal correlation is larger and the decoupled model of interference-plus-noise statistics breaks down at INRs lower than 12 dB.

5.4.4. Effect of exploiting spatial interference-plus-noise correlation

From the above results, temporal interference correlation, even if known, may not result in a performance benefit at lower INRs due to model mismatch. Therefore, we assess the benefit of taking only the *spatial* correlation of interference-plus-noise into account. As a reference, we compare the performance to the case of assuming the interference-plus-noise to be spatially white. With total interference power fixed, Figure 7 compares the average SER for one (solid lines) and two (dashed lines) interferers. In the case of two interferers, the interferers have equal power and random relative delays. Both desired and interfering users employ a $(5, 5)$ MIMO link, a total-interference-to-noise ratio of 12 dB, and a training length of $4N_t$. Both the desired and the interfering users operate at the same data rate. Figure 7 shows that for one interferer, there is 1.2 dB gain over a wide range of signal-to-interference-plus-noise ratio (SINRs), by estimating the spatial correlation of interference-plus-noise. For the case of two equal-powered interferers, the corresponding gain in SINR is negligible.

6. CONCLUSIONS

By modelling interference statistics as approximately temporally and spatially separable, we have investigated ML joint estimation of channel parameters and spatial interference correlation matrices. We have assessed the impact of temporal and spatial interference correlation on channel estimation and data detection. For training lengths of at least four times the number of transmitting antennas, gains of around 1 dB are observed by estimating spatial interference correlation. We determine that an additional 0.5 to 3.0 dB in performance gain would result if the known temporal correlation was exploited. For shorter training lengths, however, it is better to estimate only the channel matrix and assume spatially white interference in data detection. One source of temporal correlation occurs where a cochannel interferer operates

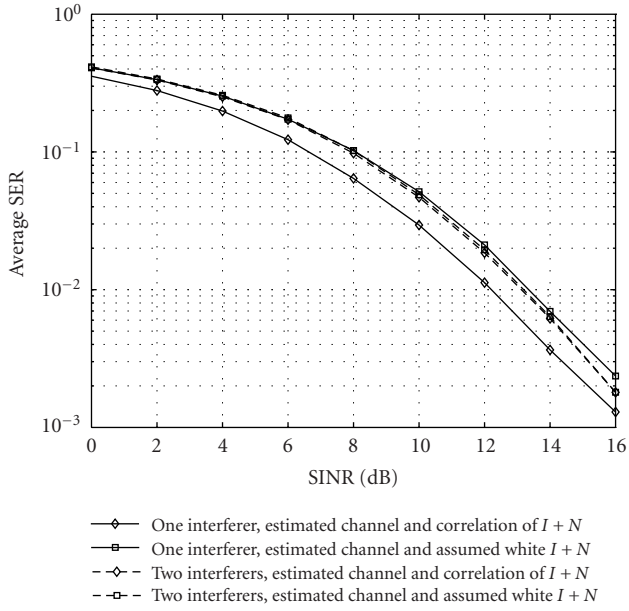


Figure 7: The improvement of estimating spatial correlation of interference-plus-noise in practical systems.

at data rate lower than that of the desired user. Exploiting temporal interference correlation in channel estimation was found not to be of benefit. However, if temporal correlation is significant, as in case of lower-data-rate interference, significant performance gains by exploiting temporal interference correlation in data detection are theoretically possible. The minimum INR levels, where separable temporal and interference correlation statistics model was shown to break down and provide no benefit, ranged from 12 or 17 dB, depending on the level of temporal correlation. Of more practical significance, it was shown that at a total INR of 12 dB, 1.2 dB of performance gain can be obtained over a wide range of SINRs by estimating spatial correlation only and neglecting temporal correlation.

ACKNOWLEDGMENT

The material in this paper was presented in part at IEEE VTC, Fall 2002.

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[3] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[4] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using the V-BLAST space-time communication architecture," *Electronics Letters*, vol. 35, no. 1, pp. 14–16, 1999.

[5] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.

[6] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.

[7] F. R. Farrokhi, G. J. Foschini, A. Lozano, and R. A. Valenzuela, "Link-optimal space-time processing with multiple transmit and receive antennas," *IEEE Communications Letters*, vol. 5, no. 3, pp. 85–87, 2001.

[8] S. Catreux, P. F. Driessen, and L. J. Greenstein, "Simulation results for an interference-limited multiple-input multiple-output cellular system," *IEEE Communications Letters*, vol. 4, no. 11, pp. 334–336, 2000.

[9] S. Catreux, P. F. Driessen, and L. J. Greenstein, "Attainable throughput of an interference-limited multiple-input multiple-output (MIMO) cellular system," *IEEE Trans. Communications*, vol. 49, no. 8, pp. 1307–1311, 2001.

[10] Y. Song and S. D. Blostein, "MIMO channel capacity in co-channel interference," in *Proc. 21st Biennial Symposium on Communications*, pp. 220–224, Kingston, Ontario, Canada, June 2002.

[11] A. Lozano and A. M. Tulino, "Capacity of multiple-transmit multiple-receive antenna architectures," *IEEE Transactions on Information Theory*, vol. 48, no. 12, pp. 3117–3128, 2002.

[12] R. S. Blum, "MIMO capacity with interference," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 793–801, 2003.

[13] M. Kang and M. S. Alouini, "Performance analysis of MIMO systems with co-channel interference over Rayleigh fading channels," in *Proc. IEEE International Conference on Communications (ICC '02)*, pp. 391–395, New York, NY, USA, April–May 2002.

[14] B. Suard, A. F. Naguib, G. Xu, and A. Paulraj, "Performance of CDMA mobile communication systems using antenna arrays," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '93)*, pp. 153–156, Minneapolis, Minn, USA, April 1993.

[15] T. Marzetta, "BLAST training: estimating channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annual Allerton Conference on Communication, Control, and Computing*, pp. 958–966, Monticello, Ill, USA, September 1999.

[16] B. Hassibi, "A fast square-root implementation for BLAST," in *Proc. 34th Asilomar Conference on Signal, Systems, and Computers*, pp. 1255–1259, Pacific Grove, Calif, USA, October 2000.

[17] J. H. Winters, "Signal acquisition and tracking with adaptive arrays in the digital mobile radio system IS-54 with flat fading," *IEEE Trans. Vehicular Technology*, vol. 42, no. 4, pp. 377–384, 1993.

[18] A. F. Naguib and A. Paulraj, "Performance of wireless CDMA with M-ary orthogonal modulation and cell site antenna arrays," *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 9, pp. 1770–1783, 1996.

[19] H. Lütkepohl, *Handbook of Matrices*, John Wiley & Sons, Chichester, 1996.

[20] Y. Song, *Multiple-input multiple-output wireless communication systems with cochannel interference*, Ph.D. thesis, Department of Electrical and Computer Engineering, Queen's University, Kingston, Ontario, Canada, 2003.

[21] S. Bäro, G. Bauch, A. Pavlic, and A. Semmler, "Improving BLAST performance using space-time block codes and turbo decoding," in *Proc. Global Telecommunications Conference*, pp. 1067–1071, San Francisco, Calif, USA, November–December 2000.
[22] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice-Hall, Upper Saddle River, NJ, USA, 1996.
[23] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 1995.
[24] W. A. Gardner, *Introduction to Random Processes*, McGraw-Hill, New York, NY, USA, 1990.
[25] G. Long, F. Ling, and J. G. Proakis, "Fractionally-spaced equalizers based on singular value decomposition," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 1514–1517, New York, NY, USA, April 1988.

Yi Song received her B.S. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 1995 and her M.S. and Ph.D. degrees in electrical engineering from Queen's University, Kingston, Ontario, Canada, in 1998 and 2003, respectively. Her research interests include wireless communications and signal processing for multiple antenna systems.

Steven D. Blostein received his B.S. degree in electrical engineering from Cornell University in 1983, and his M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 1985 and 1988, respectively. He has been on the faculty board of Queen's University in Kingston, Ontario, Canada since 1988, where he currently holds the position of Professor and Head of the Department of Electrical and Computer Engineering. His current interests lie in signal processing for wireless communications as well as detection and estimation theory. He is a Senior Member of the IEEE and a Registered Professional Engineer in Ontario.

EURASIP Journal on Applied Signal Processing 2004:5, 696–706

A MIMO System with Backward Compatibility for OFDM-Based WLANs

Jianhua Liu
Department of Electrical and Computer Engineering, University of Florida, P.O. Box 116130, Gainesville, FL 32611-6130, USA
Email: jhliu@dsp.ufl.edu

Jian Li
Department of Electrical and Computer Engineering, University of Florida, P.O. Box 116130, Gainesville, FL 32611-6130, USA
Email: li@dsp.ufl.edu

Received 16 December 2002; Revised 28 June 2003

Orthogonal frequency division multiplexing (OFDM) has been selected as the basis for the new IEEE 802.11a standard for high-speed wireless local area networks (WLANs). We consider doubling the transmission data rate of the IEEE 802.11a system by using two transmit and two receive antennas. We propose a preamble design for this multi-input multi-output (MIMO) system that is backward compatible with its single-input single-output (SISO) counterpart as specified by the IEEE 802.11a standard. Based on this preamble design, we devise a sequential method for the estimation of the carrier frequency offset (CFO), symbol timing, and MIMO channel response. We also provide a simple soft detector based on the unstructured least square approach to obtain the soft information for the Viterbi decoder. This soft detector is very simple since it decouples the multidimensional QAM symbol detection into multiple one-dimensional QAM symbol—and further PAM symbol—detections. Both the sequential parameter estimation method and the soft detector can provide excellent overall system performance and are ideally suited for real-time implementations. The effectiveness of our methods is demonstrated via numerical examples.

Keywords and phrases: MIMO system, OFDM, WLAN, symbol timing, carrier synchronization, channel estimation.

1. INTRODUCTION

Orthogonal frequency division multiplexing (OFDM) has been selected as the basis for several new high-speed wireless local area network (WLAN) standards [1], including IEEE 802.11a [2], IEEE 802.11g, and HIPERLAN/2. IEEE 802.11g and HIPERLAN/2 are very similar to IEEE 802.11a in terms of signal generation and detection/decoding. We use IEEE 802.11a to exemplify our presentation in this paper.

The OFDM-based WLAN system, as specified by the IEEE 802.11a standard, uses packet-based transmission. Each packet, as shown in Figure 1, consists of an OFDM packet preamble, a signal field, and an OFDM data field. The preamble can be used to estimate the channel parameters such as the carrier frequency offset (CFO), symbol timing, as well as channel response. These parameters are needed for the data symbol detection in the OFDM data field. The preamble design adopted by the standard is specifically tailored to the single-input single-output (SISO) system case where both the transmitter and receiver deal with a single signal. This standard supports a data rate up to 54 Mbps.

Transmission data rates higher than 54 Mbps are of particular importance for future WLANs. Deploying multiple antennas at both the transmitter and receiver is a promising way to achieve a high transmission data rate for multipath-rich wireless channels without increasing the total transmission power or bandwidth [3]. The corresponding system, as shown in Figure 2, is referred to as a multi-input multi-output (MIMO) wireless communication system, where M and N in the figure denote the numbers of transmit and receive antennas, respectively.

Among the various popular MIMO wireless communication schemes, the BLAST (Bell Labs Layered Space Time) approaches [4, 5] are particularly attractive. BLAST attempts to achieve the potentially large channel capacity offered by the MIMO system [6, 7]. In BLAST systems, the data stream is demultiplexed into independent substreams that are referred to as layers. These layers are transmitted simultaneously, that is, one layer per transmit antenna. At the receiver, the multiple layers can be detected, for example, through successive detection via an interference cancellation and nulling algorithm (ICNA) [5]. The detection

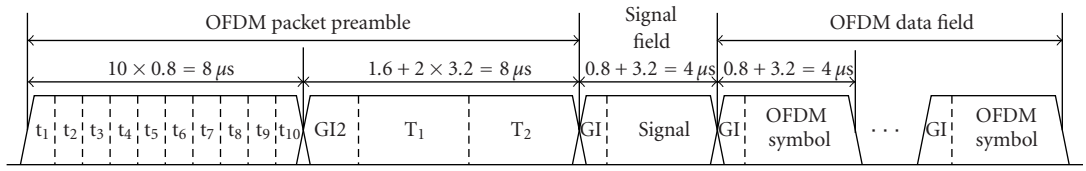


Figure 1: Packet structure of the IEEE 802.11a standard.

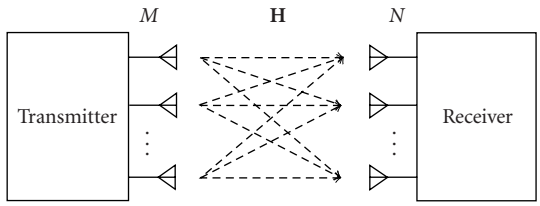


Figure 2: Diagram of a MIMO system.

can also be done via the sphere decoding (SPD) algorithm [8].

Our focus herein is on doubling the data rate of the SISO system as specified by the IEEE 802.11a standard by using two transmit and two receive antennas (referred to as the MIMO system in the sequel) based on the BLAST scheme. We propose a preamble design for this MIMO system that is backward compatible with its SISO counterpart as specified by the IEEE 802.11a standard. That is, a SISO receiver can perform CFO, symbol timing, and channel response estimation based on the proposed preamble design and detect up to the signal field. The SISO receiver is then informed, by using, for example, the reserved bit in the signal field, that a transmission is a SISO or not. Our preamble design can be used with two transmit and any number of receive antennas. However, we mainly focus on the two receive antenna case herein. Based on our MIMO preamble design, we propose a sequential method, ideally suited for real-time implementations, to estimate the CFO, symbol timing, and MIMO channel response.

The convolutional code specified in the IEEE 802.11a standard will also be used in our MIMO system for channel coding. As a result, soft information from the MIMO detector is needed by the Viterbi decoder to improve the decoding performance. Both the efficient ICNA and SPD algorithms offer only hard output. Soft output can be inferred with the ICNA-based algorithm for iterative detection and decoding [9]. However, this algorithm is computationally extremely heavy—exponentional in terms of transmit antenna number M as well as the constellation size. Although reduced complexity versions were alluded to in [9], the costs in performance degradation by using these versions were not clear. The space time bit-interleaved coded modulation (STBICM) approach [10] can deliver soft output, in both the iterative and noniterative modes, but it is also computationally extremely heavy. A list sphere decoder (LSD) algorithm [11] was recently proposed to reduce the computational complexity of STBICM with a small performance degradation. However, LSD is still very complicated and hard to implement in real time for OFDM-based MIMO WLAN applications due to the high data rate. We present herein a simple MIMO soft detector, based on the unstructured least square (LS) fitting approach. This LS-based soft detector is ideally suited for real-time implementations since it decouples the multidimensional quadrature amplitude modulation (QAM) symbol detection into multiple one-dimensional QAM symbol detections. We show that the real and imaginary parts of the noise of the decoupled detection output are independent of each other. Hence, the QAM symbol detection can be further simplified into two pulse amplitude modulation (PAM) symbol detections. As a result, this LS-based soft detector is orders of magnitude more computationally efficient than LSD; yet, the efficiency is achieved at a cost of a small performance degradation, due to the aforementioned decoupling. The LS-based detector can also be seen as related to the zero-forcing or to the linear decorrelating detector [12].

The remainder of this paper is organized as follows. Section 2 describes the MIMO system. The new preamble design is given in this section. Section 3 presents our sequential method for CFO, symbol timing, and MIMO channel response estimation. The MIMO soft detector is provided in Section 4. Numerical examples are given in Section 5 to demonstrate the effectiveness of the proposed methods. Finally, we end our paper with comments and conclusions in Section 6.

2. SYSTEM DESIGN

Our MIMO system closely resembles its SISO counterpart as specified by the IEEE 802.11a standard. We first give a brief overview of the IEEE 802.11a SISO system before we proceed to describe our MIMO system.

2.1. IEEE 802.11a standard

Figure 1 shows the packet structure as specified by the IEEE 802.11a standard. The nominal bandwidth of the OFDM signal is 20 MHz and the in-phase/quadrature (I/Q) sampling interval t_S is 50 nanoseconds. In this case, the number of samples $N_S = 64$ for an OFDM data symbol is equal to the number of subcarriers. The OFDM packet preamble consists of ten identical short OFDM training symbols t_i, $i = 1, 2, \ldots, 10$, each of which contains $N_C = 16$ samples, and two identical long OFDM training symbols T_i, $i = 1, 2$, each of which contains $N_S = 64$ samples. Between the short and

long OFDM training symbols, there is a long guard interval (GI2) consisting of $2N_C = 32$ data samples. GI2 is the cyclic prefix (CP) for the long OFDM training symbol T_1, that is, it is the exact replica of the last $2N_C$ samples of T_1.

The information carrying data are encoded in the OFDM data field. The binary source data sequence is first scrambled and then convolutionally encoded by an industrial standard constraint length $K = 7$, rate 1/2 encoder, which has generation polynomials $g_0 = (133)_8$ and $g_1 = (171)_8$. The encoded output is then punctured according to the data rate requirement and is segmented into blocks of length N_{CBPS} (number of coded bits per OFDM symbol), each of which corresponds to an OFDM data symbol. The binary data in each block is first interleaved among the subcarriers (referred to as the frequency domain (FD) interleaving in the sequel) and then mapped (in groups of $\log_2 A$ bits) into A-QAM symbols, which are used to modulate the different data carrying subcarriers. Each OFDM data symbol in the OFDM data field employs $N_S = 64$ subcarriers, 48 of which are used for data symbols and 4 for pilot symbols. There are also 12 null subcarriers with one in the center and the other 11 on the two ends of the frequency band. The OFDM data symbols, each of which consists of $N_S = 64$ samples, are obtained via taking the inverse fast Fourier transform (IFFT) of the data symbols, pilot symbols, and nulls on these N_S subcarriers. To eliminate the intersymbol interference (ISI), each OFDM data symbol is preceded by a CP or GI, which contains the last N_C samples of the OFDM data symbol.

The signal field contains the information including the transmission data rate and data length of the packet. The information is contained in 16 binary bits. There is also a reserved bit (which can be used to distinguish the MIMO from SISO transmissions) and a parity check bit. These 18 bits, padded with 6 zeros, are then encoded (by the same encoder as for the OFDM data field) to obtain a 48-bit binary sequence. The encoded sequence is then interleaved among subcarriers and used to modulate the 48 data carrying subcarriers using BPSK. The signal field consists of 64 samples and is obtained via taking the IFFT of these 48 BPSK symbols, 4 pilot symbols, and 12 nulls. Also, there is a CP of length N_C to separate the preamble from the signal field.

2.2. SISO data model

To establish the data model, consider first the generation of an OFDM data symbol in the OFDM data field. Let $\mathbf{x}_{\text{SISO}} = [x_1^1 \;\; x_2^1 \;\; \cdots \;\; x_{N_S}^1]^T$ be a vector of N_S data symbols, where $(\cdot)^T$ denotes the transpose and $x_{n_S}^1$, $n_S = 1, 2, \ldots, N_S$, is the symbol modulating the n_Sth subcarrier and is equal to 0 for null subcarriers, 1 or -1 for pilot subcarriers, and in $\mathcal{C}$ for data carrying subcarriers. Here $\mathcal{C}$ is a finite constellation, such as BPSK, QPSK, 16-QAM, or 64-QAM. Let $\mathbf{W}_{N_S} \in \mathbb{C}^{N_S \times N_S}$ be the fast Fourier transform (FFT) matrix. Then the OFDM data symbol $\mathbf{s}$ corresponding to $\mathbf{x}_{\text{SISO}}$ is obtained by taking the IFFT of $\mathbf{x}_{\text{SISO}}$. That is, $\mathbf{s} = \mathbf{W}_{N_S}^H \mathbf{x}_{\text{SISO}}/N_S$, where $(\cdot)^H$ denotes the conjugate transpose. To eliminate the ISI, each OFDM data symbol is preceded by a CP or GI $\mathbf{s}_C$ formed using $\mathbf{s}$.

Let

$$h^{(t)}(t) = \sum_p \alpha_p \delta(t - \tau_p t_S) \tag{1}$$

denote the time-domain analogue channel impulse response of the frequency-selective time-invariant fading channel, where α_p and $\tau_p t_S$, $0 \le \tau_p \le N_C$, $p \in \mathbb{Z}$, are the complex gain and time delay of the pth path, respectively. Let

$$\mathbf{h}^{(t)} = \left[h_0^{(t)} \;\; h_1^{(t)} \;\; \cdots \;\; h_{N_S-1}^{(t)}\right]^T \tag{2}$$

be the equivalent finite impulse response (FIR) filter response of $h^{(t)}(t)$, that is, if $\mathbf{h} = \mathbf{W}_{N_S}\mathbf{h}^{(t)} = [h_1 \;\; h_2 \;\; \cdots \;\; h_{N_S}]^T$ is the sampled frequency domain channel response, then for $n_S = 1, 2, \ldots, N_S$,

$$h_{n_S} = \sum_p \alpha_p e^{-j\tau_p t_S \omega}\big|_{\omega = 2\pi(n_S-1)/(N_S t_S)}. \tag{3}$$

The lth element of $\mathbf{h}^{(t)}$, $l = 0, 1, \ldots, N_S - 1$, can be written as

$$h_l^{(t)} = \sum_p \alpha_p e^{-j\pi(l+(N_S-1)\tau_p)/N_S} \frac{\sin(\pi\tau_p)}{\sin(\pi(\tau_p - l)/N_S)}, \tag{4}$$

which includes the leakage effect due to the frequency domain sampling [13].

By discarding the first N_C samples at the receiver (assuming a correct symbol timing), the noise-free and CFO free received signal vector $\mathbf{z}_{\text{SISO}}^{\text{ne}} \in \mathbb{C}^{N_S \times 1}$, due to sampling the received signal, is the circular convolution of $\mathbf{h}^{(t)}$ and $\mathbf{s}$. Hence the FFT output of the received data vector $\mathbf{z}_{\text{SISO}} = \mathbf{z}_{\text{SISO}}^{\text{ne}} + \mathbf{e}_{\text{SISO}}$, where $\mathbf{e}_{\text{SISO}} \sim \mathcal{N}(\mathbf{0}, (\sigma^2/N_S)\mathbf{I}_{N_S})$ is the additive zero-mean white circularly symmetric complex Gaussian noise with variance σ^2, can be written as [14]

$$\mathbf{y}_{\text{SISO}} = \mathbf{W}_{N_S}\mathbf{z}_{\text{SISO}} = \text{diag}\{\mathbf{h}\}\mathbf{x}_{\text{SISO}} + \mathbf{W}_{N_S}\mathbf{e}_{\text{SISO}} \in \mathbb{C}^{N_S \times 1}. \tag{5}$$

The data model in (5) can also represent the OFDM symbols in the signal field and the preamble.

Equation (5) can also be written as

$$\mathbf{y}_{\text{SISO}} = \text{diag}\{\mathbf{x}_{\text{SISO}}\}\mathbf{h} + \mathbf{W}_{N_S}\mathbf{e}_{\text{SISO}}. \tag{6}$$

Note that (5) is useful for symbol detection whereas (6) is used for channel estimation.

For the sake of simulation simplicity, the equivalent channel $\mathbf{h}^{(t)}$ is often approximated by an exponentially decaying FIR filter with length L_F [14], denoted as

$$\mathbf{h}_{L_F}^{(t)} = \left[h_0^{(t)} \;\; h_1^{(t)} \;\; \cdots \;\; h_{L_F-1}^{(t)}\right]^T. \tag{7}$$

In this case, the received signal can be easily simulated as the convolution of the channel $\mathbf{h}_{L_F}^{(t)}$ and the transmitted signal. Let t_r be the root mean square (RMS) delay spreading time and $t_n = t_r/t_S$. Then $L_F = \lceil 10t_n \rceil + 1$, where $\lceil x \rceil$ denotes the smallest integer not less than x. For $l_F = 0, 1, \ldots, L_F - 1$, we have

$$h_{l_F}^{(t)} \sim \mathcal{N}(0, (1 - e^{-1/t_n})e^{-l_F/t_n}). \tag{8}$$

This channel model is referred to as the Chayat model [15].

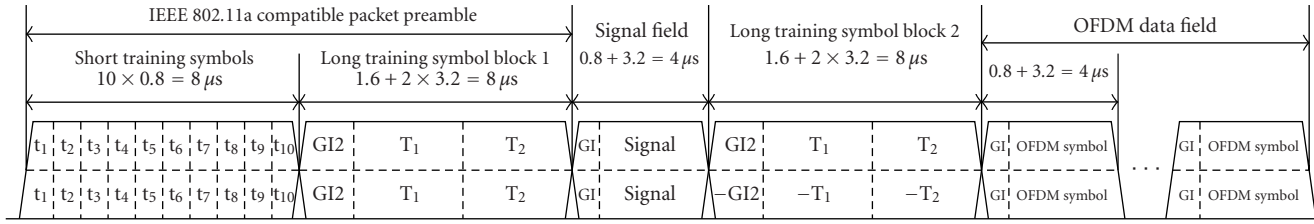


Figure 3: Proposed MIMO preamble (and signal field) structure.

Note that our symbol timing estimation method, which will be presented in Section 3, works equally well for the channel models given by both (4) and (7). Our MIMO channel response estimation method also works equally well for both models. We use (7) to generate channels to simplify our simulations.

2.3. MIMO preamble design

For the IEEE 802.11a SISO system, the short OFDM training symbols can be used to detect the arrival of the packet, allow the automatic gain control (AGC) to stabilize, compute a coarse CFO estimate, and obtain a coarse symbol timing, whereas the long OFDM training symbols can be used to calculate a fine CFO estimate, refine the coarse symbol timing, and estimate the SISO channel response.

The MIMO system considered herein has two transmit and two receive antennas (such as a crossed dipole pair for both the transmitter and the receiver). Two packets are transmitted simultaneously from the two transmit antennas. We design two preambles, one for each transmit antenna. We assume that the receiver antenna outputs suffer from the same CFO and has the same symbol timing. To be backward compatible with the SISO system, we use the same short OFDM training symbols as in the SISO preamble for both of the MIMO transmit antennas, as shown in Figure 3.

As for the long OFDM training symbols, they should be designed to support the MIMO channel response estimation. MIMO channel response estimation has attracted much research interest lately. Orthogonal training sequences tend to give the best performance (see, e.g., [16] and the references therein). We also adopt this idea of orthogonal training sequences in our preamble design. In the interest of backward compatibility, we use the same T_1 and T_2 (as well as GI2) as for the SISO system for both of the MIMO transmit antennas before the signal field, as shown in Figure 3. After the signal field, we use T_1 and T_2 (and GI2) for one transmit antenna, and $-T_1$ and $-T_2$ (and $-$GI2) for the other. This way, when the simultaneously transmitted packets are received by a single SISO receiver, the SISO receiver can successfully detect up to the signal field, which is designed to be the same for both transmit antennas. The reserved bit in the signal field can tell the SISO receiver to stop its operation whenever a MIMO transmission follows or otherwise to continue its operation. The long OFDM training symbols before and after the signal field are used in the MIMO receivers for channel estimation. Although the employment of an additional pair of long OFDM training symbols can increase the overhead, the corresponding loss of efficiency is not significant for larger packet. The reserved bit in the signal field can also inform the MIMO receiver that the transmission is a SISO one. When this occurs, the MIMO receiver can modify its channel estimation and the data bit detection steps slightly, as detailed at the end of Sections 3 and 4, respectively.

Other MIMO preamble design options with backward compatibility are possible. For example, by exploiting the transmit/receive diversities, we may get improved symbol timing or CFO correction. However, these improvements do not necessarily result in improved packet error rate (PER). Hence, we prefer the straightforward MIMO preamble design shown in Figure 3.

2.4. MIMO data model

To stay as close to the IEEE 802.11a standard as possible, we use in our MIMO system the same scrambler, convolutional encoder, puncturer, FD interleaver, symbol mapper, pilot sequence, and CP as specified in the standard. To improve diversity, we add a simple spatial interleaver to scatter every two consecutive bits across the two transmit antennas. This spatial interleaving is performed before the FD interleaving.

Consider the n_Sth subcarrier (for notational convenience, we drop the notational dependence on n_S below). Consider the case of N receive antennas. (Note that considering the general case of N receive antennas does not add extra difficulties for the discussions below.) Let $\mathbf{H}$ denote the MIMO channel matrix for the n_Sth subcarrier:

$$\mathbf{H} = \begin{bmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & h_{2,2} \\ \vdots & \vdots \\ h_{N,1} & h_{N,2} \end{bmatrix} \in \mathbb{C}^{N\times 2}, \tag{9}$$

where $h_{n,m}$ denotes the channel gain from the mth transmit antenna to the nth receive antenna for the n_Sth subcarrier. Let $\mathbf{y}$ denote a received data vector for the n_Sth subcarrier, which can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e} \in \mathbb{C}^{N\times 1}, \tag{10}$$

where $\mathbf{x} = [x_1 \;\; x_2]^T$ is the 2×1 QAM symbol vector sent on the n_Sth subcarrier and $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_N)$ is the additive white circularly symmetric complex Gaussian noise with variance σ^2. In Section 4, we will provide a soft detector based on this model.

3. CFO, SYMBOL TIMING, AND CHANNEL ESTIMATION

In this section, we present our sequential CFO, symbol timing, and MIMO channel response estimation approach based on our MIMO preamble design. The CFO can be estimated from the samples of two consecutive data blocks due to the periodic inputs (the short OFDM training symbols $t_1, \ldots, t_{10}$). Because of the fact that the CFO can be outside the unambiguous range measurable by the long OFDM training symbols, we have to estimate the CFO in two steps: (a) a coarse CFO estimation using the short OFDM training symbols and then (b) a fine CFO estimation, to determine the residue of the coarse CFO correction, using the long OFDM training symbols. After estimating and accounting for the CFO, we can obtain the symbol timing. We estimate the symbol timing also in two steps: the coarse symbol timing and fine symbol timing. The former is obtained by using the later portion of the short OFDM training symbols in the packet preamble. The fine symbol timing is obtained by using the long OFDM training symbols before the signal field. Finally, we obtain the MIMO channel response estimate. The parameter estimates are obtained in the order presented below.

3.1. Coarse CFO estimation

Let $z_n(l) = z_n^{\text{ne}}(l) + e_n(l)$, $n = 1, \ldots, N$, denote the lth time sample of the signal received from the nth receive antenna, starting from the moment that the receiver AGC has become stationary (the receiver AGC is assumed to become stationary at least before receiving the last two short OFDM training symbols and remain stationary while receiving the remainder of the packet). In the presence of CFO, we have [17]

$$z_n^{\text{ne}}(l + N_C) = z_n^{\text{ne}}(l) e^{j2N_C\pi\epsilon}, \quad n = 1, \ldots, N, \tag{11}$$

where ϵ is the normalized CFO (with respect to the sampling frequency), which we still refer to as CFO for convenience. For each receive antenna output, consider the correlation between two consecutive noise-free received data blocks, each of which is of length N_C. Then the sum of the correlations for all receive antennas can be written as

$$\begin{aligned}\sum_{n=1}^{N} \sum_{l=k}^{k+N_C-1} z_n^{\text{ne}}(l) \left(z_n^{\text{ne}}(l + N_C)\right)^* &= e^{-j2N_C\pi\epsilon} \sum_{n=1}^{N} \sum_{l=0}^{N_C-1} \left| z_n^{\text{ne}}(l) \right|^2 \\ &\triangleq P e^{-j2N_C\pi\epsilon},\end{aligned} \tag{12}$$

where $(\cdot)^*$ denotes the complex conjugate and k is any nonnegative integer such that $z_n^{\text{ne}}(k + 2N_C - 1)$ is a sample of the nth receive antenna output due to the input (transmit antenna output) being a sample of the short OFDM training symbols of the MIMO packet preamble. Let

$$\begin{aligned}P_S &= \sum_{n=1}^{N} \sum_{l=0}^{N_C-1} z_n(l) z_n^*(l + N_C) \\ &= P e^{-j2N_C\pi\epsilon} + e_P,\end{aligned} \tag{13}$$

where e_P is due to the presence of the noise. We calculate the coarse CFO as [18]

$$\hat{\epsilon}_C = -\frac{1}{2N_C\pi} \angle P_S, \tag{14}$$

where $\angle x$ denotes taking the argument of x.

We next correct the CFO using $\hat{\epsilon}_C$ to get the data samples $z_n^{(C)}(l)$, $n = 1, 2, \ldots, N$, as follows:

$$z_n^{(C)}(l) = z_n(l) e^{-j2l\pi\hat{\epsilon}_C}. \tag{15}$$

Correspondingly, we have

$$P_S^{(C)} = P_S e^{j2N_C\pi\hat{\epsilon}_C}. \tag{16}$$

In the sequel, we only consider the CFO corrected data given above. For notational convenience, we drop the superscript of $z_n^{(C)}(l)$, $n = 1, 2, \ldots, N$.

3.2. Coarse symbol timing estimation

Now we can use a correlation method, modified based on the approach presented in [17] to estimate the coarse symbol timing. The symbol timing is referred to as the starting time sample due to the input being the long OFDM training symbol T_1 (before the signal field). Once the starting time sample due to the long OFDM training symbol T_1 is determined, we can determine the starting time sample due to every OFDM data symbol thereafter. According to the specification of the IEEE 802.11a standard and the sampling rate of 20 MHz, the true symbol timing T_0 is 193, as shown in Figure 4.

From (13) and (16), we note that the correlation (after the CFO correction) is approximately the real-valued scalar P plus a complex-valued noise. Hence we propose to use the following real-valued correlation sequence for coarse symbol timing determination. We calculate the correlation sequence in an iterative form similar to the complex-valued approach in [17] as follows:

$$\begin{aligned}&P_R(k+1) \\ &= P_R(k) \\ &\quad + \text{Re}\left\{ \sum_{n=1}^{N} \left[z_n(k + N_C) z_n^*(k + 2N_C) - z_n(k) z_n^*(k + N_C) \right] \right\} \\ &= P_R(k) + \sum_{n=1}^{N} \{ \bar{z}_n(k + N_C) \left[\bar{z}_n(k + 2N_C) - \bar{z}_n(k) \right] \\ &\qquad + \tilde{z}_n(k + N_C) \left[\tilde{z}_n(k + 2N_C) - \tilde{z}_n(k) \right] \},\end{aligned} \tag{17}$$

where both $\text{Re}(\cdot)$ and $\bar{(\cdot)}$ denote the real part of a complex entity and $\tilde{(\cdot)}$ stands for the imaginary part. We start the iteration by using $P_R(0) = \text{Re}(P_S)$. Note that the real-valued correlation approach given in (17) is superior to the absolute-valued one given in [17] since the former uses fewer computations, lowers the noise level (variance reduced in half) in the correlation sequence, and decreases closer to zero when the data samples in the sliding data blocks are due to the

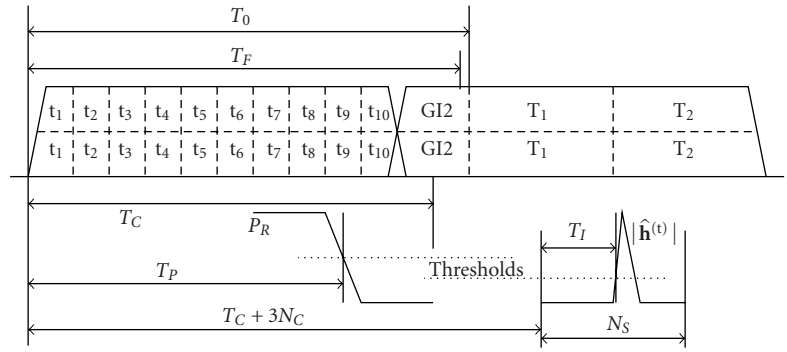


Figure 4: Illustration of symbol timing determination.

input being GI2 or the long OFDM training symbols following the short OFDM training symbols in the preamble.

When some of the data samples of the sliding data blocks are taken from the received data due to the input being GI2 or the long training symbols following the short OFDM training symbol, $P_R(k)$ will drop since (11) no longer holds. This property is used to obtain the coarse symbol timing. Let T_P, as shown in Figure 4, denote the first time sample when $P_R(k)$ drops to less than half of its peak value. The coarse symbol timing

$$T_C = T_P + \frac{3}{2}N_C + N_C \tag{18}$$

is the coarse estimate of the beginning time sample due to the input being the long OFDM training symbol T_1 before the signal field. The second term at the right-hand side of the above equation is due to the fact that $P_R(k)$ will drop to approximately one half of its maximum value when the data samples of the second half of the second of the two sliding blocks are due to the first GI2 in the preamble as input; the third term is due to one half of the length of GI2. When the channel spreading delay $t_D = \max\{\tau_p - \tau_l\}$ is assumed to satisfy $t_D \leq N_C$, only the first half of GI2 can suffer from ISI. Hence our goal of coarse timing determination is to place the coarse timing estimate between the true timing $T_0 = 193$ and $T_0 - N_C = 177$ to make accurate fine CFO estimation possible. This explains why we use N_C instead of $2N_C$ for the third term in (18).

3.3. Fine CFO estimation

For each receive antenna output, we calculate the correlation between the two long OFDM training symbols before the signal field. We then sum the correlations for all receive antennas as follows:

$$P_L = \sum_{n=1}^{N}\sum_{l=0}^{N_S-1} z_n(l+T_C)z_n^*(l+T_C+N_S). \tag{19}$$

Then the fine CFO estimate can be computed as

$$\hat{\epsilon}_F = -\frac{1}{2N_S\pi}\angle P_L. \tag{20}$$

We can use $\hat{\epsilon}_F$ in the same way as $\hat{\epsilon}_C$ to correct the CFO. We assume that for the data we use below, $\hat{\epsilon}_F$ has been already corrected.

Note that the aforementioned simple fine CFO estimation approach may not be optimal. For example, the CFO estimation accuracy could be improved by using the long OFDM training symbols after the signal field as well; however, our simple fine CFO estimation approach is sufficiently accurate in that the overall system performance can no longer be improved with a more accurate CFO estimate, especially when pilot symbols are exploited. For the data bit detection, no matter how accurate the CFO estimate is, it can never be perfect due to the presence of noise. Pilot symbols are used to track the CFO residual phase for each OFDM data symbol before data bit detection. A maximum likelihood (ML) CFO residual tracking scheme is given in Appendix C.

3.4. Fine symbol timing estimation

We now move on to obtain the fine symbol timing by using the long OFDM training symbols before the signal field. The fine symbol timing is estimated by using the N data blocks of length N_S, starting from the time sample $T_C + 3N_C$. With this choice, due to the fact that T_1 is identical to T_2, the data blocks are most likely due to the input being the second half of T_1 and the first half of T_2, even when the coarse symbol timing has a large error.

Let $\mathbf{y}_n$ denote the N_S-point FFT of the data block from the nth receive antenna and let $\mathbf{h}_{n,m}^{(t)}$ be the equivalent FIR channel in the time domain between the mth transmit antenna and the nth receive antenna, $n = 1, 2, \ldots, N$, $m = 1, 2$. Then, by neglecting the existence of the residual CFO, $\mathbf{y}_n$ can be written as (cf. (6))

$$\mathbf{y}_n = \mathbf{X}_B\mathbf{W}_{N_S}\sum_{m=1}^{2}\mathbf{h}_{n,m}^{(t)} + \mathbf{W}_{N_S}\mathbf{e}_n, \tag{21}$$

where $\mathbf{X}_B$ is a diagonal matrix with the 52 known BPSK symbols and 12 zeros, which form the T_1 in Figure 3, on the diagonal. Since the Moore-Penrose pseudoinverse of $\mathbf{X}_B$ is $\mathbf{X}_B$ itself and $\mathbf{W}_{N_S}/N_S^{1/2}$ is unitary, we get an estimate of $\mathbf{h}_n^{(t)} = \sum_{m=1}^{2}\mathbf{h}_{n,m}^{(t)}$ as

$$\hat{\mathbf{h}}_n^{(t)} = \frac{1}{N_S}\mathbf{W}_{N_S}^H \mathbf{X}_B \mathbf{y}_n. \tag{22}$$

Let T_I, as shown in Figure 4, denote the index of the first element of $|\hat{\mathbf{h}}^{(t)}| = \sum_{n=1}^{N} |\hat{\mathbf{h}}_n^{(t)}|$ that is above 1/3 of the maximum value of the elements of $\sum_{n=1}^{N} |\hat{\mathbf{h}}_n^{(t)}|$. (Our empirical experience suggests that selecting the threshold to be 1/3 gives the best result.) Then the fine symbol timing T_F is obtained as

$$T_F = T_C - N_C + T_I - 3. \tag{23}$$

The second term above is used to compensate for the aforementioned $3N_C$ shift due to the fact that $N_S - 3N_C = N_C$ and the last term above is chosen to be 3 to ensure that $T_F > T_0$ occurs with very low probability.

3.5. MIMO channel response estimation

After we obtained T_F, we can now estimate the MIMO channel response. Let $\mathbf{y}_{n,1}$ denote the N_S-point FFT of the average of the two consecutive blocks, each of which is of length N_S, associated with the two long training symbols before the signal field, from the nth receive antenna. Let $\mathbf{y}_{n,2}$ denote the counterpart of $\mathbf{y}_{n,1}$ after the signal field. Then, for the n_Sth subcarrier, we have

$$y_{n,1} \approx x_B(h_{n,1} + h_{n,2}), \tag{24}$$

$$y_{n,2} \approx x_B(h_{n,1} - h_{n,2}), \tag{25}$$

where x_B denotes the n_Sth diagonal element of $\mathbf{X}_B$, $y_{n,i}$ denotes the n_Sth element of $\mathbf{y}_{n,i}$, $i = 1, 2$, and we have dropped the dependence on n_S for notational simplicity. Solving (24) and (25) yields

$$\hat{h}_{n,1} = \frac{x_B(y_{n,1} + y_{n,2})}{2}, \tag{26}$$

$$\hat{h}_{n,2} = \frac{x_B(y_{n,1} - y_{n,2})}{2}. \tag{27}$$

When the reserved bit in the signal field indicates a SISO transmission, we only need to estimate $h_{n,1}$, $n = 1, 2, \ldots, N$, in a way similar to (26).

4. A SIMPLE MIMO SOFT DETECTOR

With the CFO, symbol timing, and MIMO channel response determined and accounted for, we can proceed to detect the data bits contained in each BLAST layer and subcarrier of the OFDM data symbols in the OFDM data field. In the sequel, we present a very simple soft detector for the MIMO system. Note that this soft detector can be used in a general setting of the BLAST system and hence we present it in a general framework based on the data model of (10), where $\mathbf{H}$ is assumed to be $N \times M$ and $\mathbf{x}$ to be $M \times 1$. (We use $\hat{\mathbf{H}}$ to replace $\mathbf{H}$ in our simulations.)

Consider first the ML hard detector of the BLAST system. For the data model of (10), the ML hard detector is given by

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{C}^{M\times 1}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||^2, \tag{28}$$

where $\|\cdot\|^2$ denotes the Euclidean norm. The cost function in (28) can be written as

$$\begin{aligned} ||\mathbf{y} - \mathbf{H}\mathbf{x}||^2 &= \mathbf{y}^H\mathbf{y} + \mathbf{x}^H\mathbf{H}^H\mathbf{H}\mathbf{x} - \mathbf{y}^H\mathbf{H}\mathbf{x} - \mathbf{x}^H\mathbf{H}^H\mathbf{y} \\ &= \left(\mathbf{x}^H - \mathbf{y}^H(\mathbf{H}^\dagger)^H\right)\mathbf{H}^H\mathbf{H}(\mathbf{x} - \mathbf{H}^\dagger\mathbf{y}) \\ &\quad + \mathbf{y}^H\mathbf{y} - \mathbf{y}^H(\mathbf{H}^\dagger)^H\mathbf{H}^H\mathbf{H}\mathbf{H}^\dagger\mathbf{y}, \end{aligned} \tag{29}$$

where $\mathbf{H}^\dagger = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H$. We note, from the above equation, that by ignoring the constellation constraint on $\mathbf{x}$, we can obtain an unstructured LS estimate $\hat{\mathbf{x}}_{\text{us}}$ of $\mathbf{x}$, which is given by

$$\hat{\mathbf{x}}_{\text{us}} = \mathbf{H}^\dagger\mathbf{y} = \mathbf{x} + \mathbf{H}^\dagger\mathbf{e} \triangleq \mathbf{x} + \mathbf{c}. \tag{30}$$

Note that $\hat{\mathbf{x}}_{\text{us}}$ is the soft decision statistic that we are interested in. We refer to this simple scheme of obtaining a soft decision statistic as the MIMO soft detection scheme. Note that a necessary condition for $\mathbf{H}^H\mathbf{H}$ to be nonsingular is $N \geq M$. Also note that $\mathbf{c}$ is still Gaussian with zero mean and covariance matrix

$$\mathrm{E}[\mathbf{c}\mathbf{c}^H] = \sigma^2\mathbf{H}^\dagger(\mathbf{H}^\dagger)^H = \sigma^2(\mathbf{H}^H\mathbf{H})^{-1}. \tag{31}$$

Due to the use of the interleaver and deinterleaver, the data bits contained in $\mathbf{x}$ are independent of each other. By ignoring the dependence among the elements of $\mathbf{c}$, we can consider only the marginal probability density function (pdf) for the elements $\hat{x}_{\text{us}}(m)$, $m = 1, 2, \ldots, M$, of $\hat{\mathbf{x}}_{\text{us}}$. (Note that an approximation is made here, which can lead to performance degradation. However, the computation is greatly simplified by the approximation.) Let

$$\mathbf{H}^\dagger \triangleq \begin{bmatrix} \check{\mathbf{h}}_1^T \\ \check{\mathbf{h}}_2^T \\ \vdots \\ \check{\mathbf{h}}_M^T \end{bmatrix} \in \mathbb{C}^{M\times N}. \tag{32}$$

Then the mth element of $\mathbf{c}$, $m = 1, 2, \ldots, M$, can be written as

$$c_m = \check{\mathbf{h}}_m^T\mathbf{e}. \tag{33}$$

Obviously, c_m is still Gaussian with zero mean and variance

$$\sigma_m^2 = \mathrm{E}\left[|c_m|^2\right] = ||\check{\mathbf{h}}_m||^2\sigma^2. \tag{34}$$

The estimate of the above noise variance σ^2 can be easily obtained via the difference of the two consecutive blocks of the nth receive antenna, from which we got $\mathbf{y}_{n,1}$ (cf. (24)). Note that σ_m^2 along with $\hat{x}_{\text{us}}^{(m)}$ provide the soft information for the mth, $m = 1, 2, \ldots, M$, symbol in $\hat{\mathbf{x}}_{\text{us}}$, needed by the Viterbi decoder. Note also that the noises corresponding to different layers have different variances which means that the symbols corresponding to different layers have different quality. This unbalanced layer quality is the reason why we have used a spatial interleaver before the FD interleaver.

Note that for SISO systems we usually consider an ordinary QAM symbol as two PAM symbols (e.g., a 64-QAM symbol can be considered as two 8-ary PAM symbols) due to the orthogonality between the real and imaginary parts of a QAM symbol as well as the independence between the real and imaginary parts of the additive circularly symmetric Gaussian error. A bit metric computation scheme for PAM symbols is presented in Appendix A. In Appendix B, we show that the real and imaginary parts of c_m are independent of each other. Hence we can significantly simplify the bit metric computations by exploiting these independencies.

The minimum mean square error (MMSE) detector is often deemed to be better than the LS-based one [12]. Although this can be true for the constant modulus constellations, such as PSK, it is not necessarily true for QAM symbols, as suggested by our simulations due to the different power levels of the QAM symbols. Hence, we do not provide an MMSE counterpart of the LS-based soft detector.

When the reserved bit in the signal field indicates a SISO transmission, the $\mathbf{H}$ in (28) is in fact a vector. Hence the $\hat{\mathbf{x}}_{\text{us}}$ in (30) and the $\mathrm{E}[\mathbf{cc}^H]$ in (31) are scalars, and they are the soft information used as in the SISO system for data bit detection.

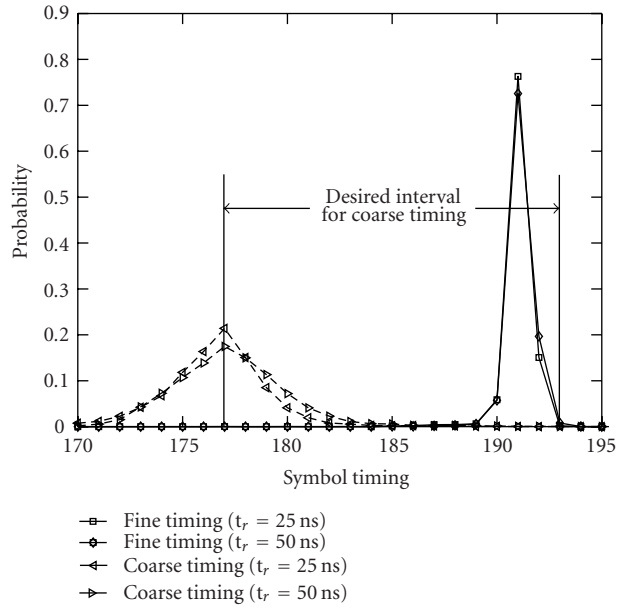


Figure 5: Coarse and fine symbol timing estimates.

5. NUMERICAL EXAMPLES

In this section, we provide numerical examples to demonstrate the effectiveness and performance of our sequential estimation method for CFO, symbol timing, and MIMO channel response based on our MIMO preamble design as well as the simple MIMO soft detector.

In the IEEE 802.11a standard, the maximum transmission data rate is 54 Mbps; in this case, the 64-QAM constellation is used and the channel coding rate is $R = 3/4$, which comes from puncturing the $R_C = 1/2$ encoded sequence with the puncturing rate $R_P = 2/3$. We consider doubling the maximum 54 Mbps transmission data rate by using two transmit and two receive antennas, that is, $M = N = 2$. In our simulations, each of the $MN = 4$ time domain MIMO channels is generated according to the Chayat model; the 4 channels are independent of each other.

Due to the fact that 52 out of 64 subcarriers are used in the OFDM-based WLAN system, the signal-to-noise ratio (SNR) for the SISO system used in this paper is defined as $52/(64\sigma^2)$ for the constellations whose average energies are normalized to 1. Whereas for the MIMO system, the SNR is defined as $52/(128\sigma^2)$ (i.e., we use the same total transmission power for the MIMO system as for its SISO counterpart).

We first provide a simulation example for symbol timing estimation. Two curves in Figure 5 show the 10^4 Monte Carlo simulation results of the coarse symbol timing estimates for the Chayat channels with $t_r = 25$ and $t_r = 50$ nanoseconds, respectively, when SNR = 10 dB. Note that the coarse symbol timing estimates fall within the desired interval with a high probability. Note also that the adverse effect of the coarse symbol timing estimate being smaller than $T_0 - N_C = 177$ is usually not significant since, due to the exponentially decaying property of the channels, the ISI in the receiver output due to the input being the latter portion of the first half of GI2 is minimal. (Since the adverse effect of the coarse symbol timing estimate being larger than $T_0 = 193$ is usually troublesome for the fine CFO estimation, we prefer the coarse timing estimate T_C to be well ahead of T_0.) The other two curves in Figure 5 show the 10^4 Monte Carlo simulation results of the fine symbol timing estimates for the Chayat channels with $t_r = 25$ and $t_r = 50$ nanoseconds, respectively, when SNR = 10 dB. Note that our simple fine symbol timing approach gives highly accurate timing estimates.

We then provide a simulation example to show the effectiveness of the MIMO channel estimator and the PER performance of the MIMO soft detector. (One packet consists of 1000 bytes. Based on the IEEE 802.11a standard, even if only one error occurs in a packet, the entire packet is discarded.) In Figure 6, we show the 10^4 Monte Carlo simulation results of the PER performance of our soft detector as a function of the SNR for the MIMO system, with t_r being 50 nanoseconds for the Chayat channels, when the transmission data rate is 108 Mbps. We consider two cases: the case of perfect channel knowledge and the case of estimated channel parameters. For the former case, we assume the exact knowledge of CFO, symbol timing, and MIMO channel, whereas for the latter case, we use the estimates of all of the aforementioned parameters obtained with our sequential approach from the MIMO packet preamble as well as the CFO residual phase tracking. As a reference, we also give the PER curves of the soft detector for the SISO system (with the data rate being 54 Mbps) as a reference. We note, from the PER curves, that both the MIMO preamble design and the sequential channel parameter estimation algorithm are effective in that the gap between the PER curves corresponding to the perfect channel knowledge case and the estimated channel parameter case for the MIMO system is no more than that of the SISO system.

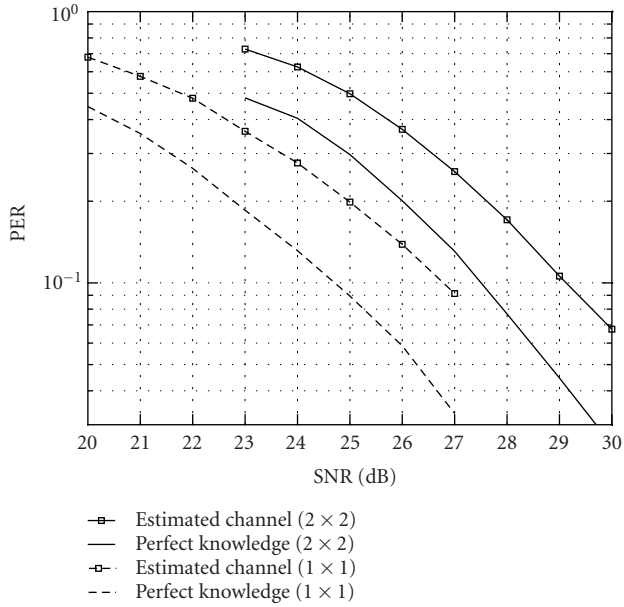


Figure 6: PER versus SNR at the 108 Mbps data rate for the Chayat channels with $t_r = 50$ nanoseconds.

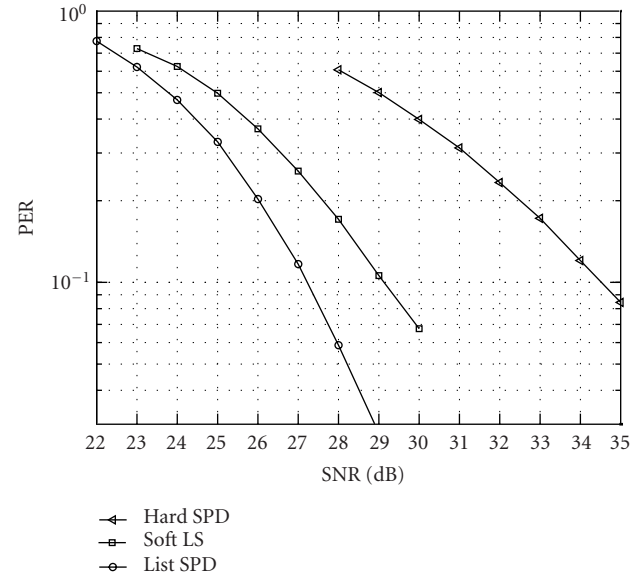


Figure 7: PER versus SNR at the 108 Mbps data rate for the Chayat channels with $t_r = 50$ nanoseconds.

Note also that the MIMO soft detector is effective in that the MIMO system needs only 2 to 3 dB extra total transmission power to keep the same PER (we are mostly interested in PERs being 0.1, according to the IEEE 802.11a standard) as its SISO counterpart, but with the data rate doubled.

Finally, we show the performance comparisons of the SPD hard detector [8], the MIMO soft detector, as well as the LSD-based soft detector [11], an approximation of STBICM, the ideal MIMO soft detector. Figure 7 gives PER curves obtained from 10^4 Monte Carlo simulations for these detectors as a function of SNR for the MIMO system, with estimated channel parameters. (The simulation parameters are the same as those in the previous example.) Note that the MIMO soft detector is much better than SPD and is outperformed by LSD in terms of PER. However, the MIMO soft detector is much more efficient than LSD. We did not attempt to optimize our Matlab simulation codes. Even so, our preliminary results indicate that the LS-based MIMO soft detector requires only about one fifth of the computations needed by SPD. Unlike SPD, whose sphere radius shrinks when finding better solutions, LSD keeps the sphere radius constant, which means that it is computationally much more demanding than SPD, especially when the sphere radius is large. We do not have an exact flop comparison, yet we believe LSD flops $\gg$ SPD flops $\approx$ 5 times MIMO soft detector flops, which means that the LS-based soft detector can be orders of magnitude more computationally efficient than the LSD-based one.

6. CONCLUDING REMARKS

We have proposed a preamble design for the MIMO system with two transmit and two receive antennas. This MIMO preamble design is backward compatible with its SISO counterpart as specified by the IEEE 802.11a standard. Based on this MIMO preamble design, we have devised a sequential method for the estimation of CFO, symbol timing, and MIMO channel responses. We have also provided a simple soft detector for the MIMO system based on the unstructured LS approach to obtain the soft information for the Viterbi decoder. Both the sequential parameter estimation method and the soft detector are very efficient and ideally suited for real-time implementations. The effectiveness of our methods has been demonstrated via numerical examples.

APPENDICES

A. BIT METRIC CALCULATION FOR THE QAM SYMBOL

To make this paper self-contained, we describe in this appendix briefly our bit metric calculation method for the QAM symbol. Note that the real and imaginary parts of a QAM symbol plus the additive circularly symmetric complex Gaussian noise are independent of each other. Hence the soft decision statistic corresponding to a transmitted QAM symbol can be easily divided into the real and imaginary parts, which correspond to the soft decision statistic of two real valued PAM symbols. The variance of the noise additive to the PAM symbols is halved as compared to the QAM symbols. In view of this, we only present the method for calculating the bit metric for a symbol in the PAM constellation $\mathcal{R}$.

Let $\mathcal{D}_{i,j} = \{s : s \in \mathcal{R}\}$ denote the set of all the possible PAM symbols with the ith bit $v_i = j$, $i = 1, 2, \ldots, \log_2 A/2$, $j = 0, 1$. The formation of $\mathcal{D}_{i,j}$ depends on the way the PAM symbols are labeled. For example, for the Gray indexed 8-ary

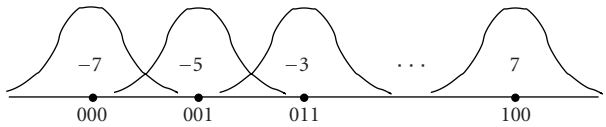


Figure 8: Illustration of 8-ary PAM symbols with Gray labeling and their pdf curves.

PAM constellation shown in Figure 8, we have

$$\mathcal{D}_{1,0} = \{-7, -5, -3, -1\}, \tag{A.1}$$

where the first bit is the left most one shown in Figure 8. Then, for a given soft information (x, σ) of the PAM symbol, the bit metric for v_i is given by

$$\hat{v}_i = \log \frac{p_\sigma(v_i = 1\,|\,x)}{p_\sigma(v_i = 0\,|\,x)}, \tag{A.2}$$

where

$$\begin{aligned} p_\sigma(v_i = j\,|\,x) &= p_\sigma(\mathcal{D}_{i,j}\,|\,x) \\ &= \sum_{s\in\mathcal{D}_{i,j}} p_\sigma(s|x) \\ &= \sum_{s\in\mathcal{D}_{i,j}} \frac{f_\sigma(x|s)p(s)}{p(x)}, \end{aligned} \tag{A.3}$$

with

$$f_\sigma(x|s) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-s)^2/2\sigma^2} \tag{A.4}$$

being the pdf given the symbol s and the variance σ^2, as shown in Figure 8. The occurrence of each symbol in $\mathcal{R}$ is often assumed to be equally likely, that is, $p(s) = (1/2)^{B/2}$, for all $s \in \mathcal{R}$. In this case, we have

$$p_\sigma(v_i = j|x) = \frac{1}{2^{B/2}p(x)} \sum_{s\in\mathcal{D}_{i,j}} f_\sigma(x|s), \tag{A.5}$$

which leads to

$$\hat{v}_i = \log\left\{\sum_{s\in\mathcal{D}_{i,1}} e^{-(x-s)^2/2\sigma^2}\right\} - \log\left\{\sum_{s\in\mathcal{D}_{i,0}} e^{-(x-s)^2/2\sigma^2}\right\}. \tag{A.6}$$

To speed up the bit metric calculation in practical applications, we can make a grid for x and σ to precalculate a lookup table for the $\hat{v}_i$'s. The bit metric calculation in our simulations is based on such a table.

B. A PROPERTY OF THE c_m IN (33)

With $\bar{x}$ and $\tilde{x}$ denoting the real and imaginary parts of x, respectively, we get from (33)

$$\begin{aligned} \mathbf{c} &= (\bar{\mathbf{H}}^\dagger + j\tilde{\mathbf{H}}^\dagger)(\bar{\mathbf{e}} + j\tilde{\mathbf{e}}) \\ &= (\bar{\mathbf{H}}^\dagger\bar{\mathbf{e}} - \tilde{\mathbf{H}}^\dagger\tilde{\mathbf{e}}) + j(\tilde{\mathbf{H}}^\dagger\bar{\mathbf{e}} + \bar{\mathbf{H}}^\dagger\tilde{\mathbf{e}}). \end{aligned} \tag{B.1}$$

Then

$$\bar{\mathbf{c}} = \bar{\mathbf{H}}^\dagger\bar{\mathbf{e}} - \tilde{\mathbf{H}}^\dagger\tilde{\mathbf{e}}, \qquad \tilde{\mathbf{c}} = \tilde{\mathbf{H}}^\dagger\bar{\mathbf{e}} + \bar{\mathbf{H}}^\dagger\tilde{\mathbf{e}}. \tag{B.2}$$

Hence, we have

$$\begin{aligned} \mathrm{E}\left[\tilde{\mathbf{c}}\tilde{\mathbf{c}}^T\right] &= \mathrm{E}\left[(\bar{\mathbf{H}}^\dagger\bar{\mathbf{e}} - \tilde{\mathbf{H}}^\dagger\tilde{\mathbf{e}})\left(\tilde{\mathbf{e}}^T(\bar{\mathbf{H}}^\dagger)^T + \tilde{\mathbf{e}}^T(\tilde{\mathbf{H}}^\dagger)^T\right)\right] \\ &= \mathrm{E}\left[\bar{\mathbf{H}}^\dagger\tilde{\mathbf{e}}\tilde{\mathbf{e}}^T(\bar{\mathbf{H}}^\dagger)^T - \tilde{\mathbf{H}}^\dagger\tilde{\mathbf{e}}\tilde{\mathbf{e}}^T(\tilde{\mathbf{H}}^\dagger)^T\right] \\ &\quad + \mathrm{E}\left[\bar{\mathbf{H}}^\dagger\tilde{\mathbf{e}}\tilde{\mathbf{e}}^T(\tilde{\mathbf{H}}^\dagger)^T - \tilde{\mathbf{H}}^\dagger\tilde{\mathbf{e}}\tilde{\mathbf{e}}^T(\bar{\mathbf{H}}^\dagger)^T\right] \\ &= \frac{1}{2}\sigma^2\left[\bar{\mathbf{H}}^\dagger(\tilde{\mathbf{H}}^\dagger)^T - \tilde{\mathbf{H}}^\dagger(\bar{\mathbf{H}}^\dagger)^T\right] \\ &= \frac{1}{2}\sigma^2\left[\bar{\mathbf{H}}^\dagger(\tilde{\mathbf{H}}^\dagger)^T - \left(\bar{\mathbf{H}}^\dagger(\tilde{\mathbf{H}}^\dagger)^T\right)^T\right], \end{aligned} \tag{B.3}$$

where we have used the fact that $\mathrm{E}[\bar{\mathbf{e}}\bar{\mathbf{e}}^T] = \mathrm{E}[\tilde{\mathbf{e}}\tilde{\mathbf{e}}^T] = \sigma^2\mathbf{I}/2$ and $\mathrm{E}[\bar{\mathbf{e}}\tilde{\mathbf{e}}^T] = \mathrm{E}[\tilde{\mathbf{e}}\bar{\mathbf{e}}^T] = \mathbf{0}$. Equation (B.3) implies that the diagonal elements of $\mathrm{E}[\tilde{\mathbf{c}}\tilde{\mathbf{c}}^T]$ are zero and hence $\mathrm{E}[\tilde{c}_m\tilde{c}_m] = 0$, $m = 1, 2, \ldots, M$.

C. PHASE CORRECTION USING PILOT SYMBOLS

The CFO correction will never be perfect in practice due to the presence of noise. Hence, there will be a phase error ϕ for each OFDM data symbol caused by the error in the fine CFO estimate $\hat{\epsilon}_F$. The error ϕ increases linearly with time.

As we mentioned earlier, each OFDM data symbol contains four known pilot symbols. We denote these pilot symbols by a 4×1 vector $\mathbf{p}$. The pilot symbols can be used to correct ϕ for each OFDM data symbol after CFO correction using $\hat{\epsilon}_F$. Let $\mathbf{y}_n^{(\mathrm{p})}$ be the vector containing the corresponding four elements of the FFT output of an OFDM data symbol in the OFDM data field received from the nth antenna, $n = 1, 2, \ldots, N$. Let $\hat{\mathbf{h}}_{n,m}^{(\mathrm{p})}$ be the 4×1 estimated channel vector from transmit antenna m to receive antenna n for the four corresponding subcarriers. Let $\mathbf{P} = \mathrm{diag}\{\mathbf{p}\}$. We have

$$\mathbf{y}_n^{(\mathrm{p})} = e^{j\phi}\mathbf{P}\sum_{m=1}^{2}\hat{\mathbf{h}}_{n,m}^{(\mathrm{p})} + \mathbf{e}_n^{(\mathrm{p})}, \quad n = 1, 2, \ldots, N, \tag{C.1}$$

where $\{\mathbf{e}_n^{(\mathrm{p})}\}_{n=1}^N$ are zero-mean white circularly symmetric complex Gaussian noise vectors that are independently and identically distributed. Then the ML criterion leads to

$$\begin{aligned} \hat{\phi}_{\mathrm{ML}} &= \arg\min_\phi \sum_{n=1}^{N}\left\|\mathbf{y}_n^{(\mathrm{p})} - e^{j\phi}\mathbf{P}\sum_{m=1}^{2}\hat{\mathbf{h}}_{n,m}^{(\mathrm{p})}\right\|^2 \\ &= \angle\left\{\sum_{n=1}^{N}\left[\sum_{m=1}^{2}\left(\hat{\mathbf{h}}_{n,m}^{(\mathrm{p})}\right)^H\right]\mathbf{P}^H\mathbf{y}_n^{(\mathrm{p})}\right\}. \end{aligned} \tag{C.2}$$

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation Grant CCR-0097114 and the Intersil Corporation Contract 2001056.

REFERENCES

[1] R. van Nee, G. Awater, M. Morikura, H. Takanashi, M. Webster, and K. W. Halford, "New high-rate wireless LAN standards," *IEEE Communications Magazine*, vol. 37, no. 12, pp. 82–88, 1999.

[2] IEEE 802.11a-1999, "IEEE Standard for Information technology—Telecommunications and information exchange between systems—Local and metropolitan area networks—specific requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications—Amendment 1: High-speed Physical Layer in the 5 GHz band," 1999.

[3] A. F. Naguib, N. Seshadri, and A. R. Calderbank, "Increasing data rate over wireless channels," *IEEE Signal Processing Magazine*, vol. 17, no. 3, pp. 76–92, 2000.

[4] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[5] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using the V-BLAST space-time communication architecture," *Electronics Letters*, vol. 35, no. 1, pp. 14–15, 1999.

[6] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[7] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[8] O. Damen, A. Chkeif, and J.-C. Belfiore, "Lattice code decoder for space-time codes," *IEEE Communications Letters*, vol. 4, no. 5, pp. 161–163, 2000.

[9] X. Li, H. Huang, G. J. Foschini, and R. A. Valenzuela, "Effects of iterative detection and decoding on the performance of BLAST," in *Proc. IEEE GLOBECOM'00*, vol. 2, pp. 1061–1066, San Francisco, Calif, USA, November 2000.

[10] A. M. Tonello, "Space-time bit-interleaved coded modulation with an iterative decoding strategy," in *Proc. IEEE 52th Vehicular Technology Conference*, vol. 1, pp. 473–478, Boston, Mass, USA, September 2000.

[11] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Transactions on Communications*, vol. 51, no. 3, pp. 389–399, 2003.

[12] C. Z. W. Hassell Sweatman, J. S. Thompson, B. Mulgrew, and P. M. Grant, "A comparison of the MMSE detector and its BLAST version for the MIMO channel," in *IEE Seminar on MIMO: Communications Systems from Concept to Implementations*, pp. 19/1–19/6, London, UK, December 2001.

[13] J.-J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Börjesson, "On channel estimation in OFDM systems," in *Proc. IEEE 45th Vehicular Technology Conference*, vol. 2, pp. 815–819, Chicago, Ill, USA, July 1995.

[14] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications," *IEEE Signal Processing Magazine*, vol. 17, pp. 29–48, May 2000.

[15] N. Chayat, "Tentative criteria for comparison of modulation methods," IEEE P802.11-97/96, September 1997.

[16] E. G. Larsson and J. Li, "Preamble design for multiple-antenna OFDM-based WLANs with null subcarriers," *IEEE Signal Processing Letters*, vol. 8, no. 11, pp. 285–288, 2001.

[17] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Transactions on Communications*, vol. 45, no. 12, pp. 1613–1621, 1997.

[18] J. Li, G. Liu, and G. B. Giannakis, "Carrier frequency offset estimation for OFDM-based WLANs," *IEEE Signal Processing Letters*, vol. 8, no. 3, pp. 80–82, 2001.

Jianhua Liu received the B.S. degree in electrical engineering from Dalian Maritime University, Dalian, China, in 1984, the M.S. degree in electrical engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 1987, and the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in 1998. From March 1987 to February 1999, he worked at the Communications, Telemetry, and Telecontrol Research Institute, Shijiazhuang, China, where he was an Assistant Engineer, Engineer, Senior Engineer, and Fellow Engineer. From March 1995 to August 1998, he was also a Research Assistant at Tsinghua University. From February 1999 to June 2000, he worked at Nanyang Technological University, Singapore, as a Research Fellow. Since June 2000, he has been a Research Assistant in the Department of Electrical and Computer Engineering, the University of Florida, Gainesville, working towards a Ph.D. degree majoring in electrical engineering and minoring in Statistics. His research interests include wireless communications, statistical signal processing, and sensor array processing.

Jian Li received the M.S. and Ph.D. degrees in electrical engineering from The Ohio State University, Columbus, in 1987 and 1991, respectively. From April 1991 to June 1991, she was an Adjunct Assistant Professor with the Department of Electrical Engineering, The Ohio State University, Columbus. From July 1991 to June 1993, she was an Assistant Professor with the Department of Electrical Engineering, University of Kentucky, Lexington. Since August 1993, she has been with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, where she is currently a Professor. Her current research interests include spectral estimation, array signal processing, and their applications. Dr. Li is a member of Sigma Xi and Phi Kappa Phi. She received the 1994 National Science Foundation Young Investigator Award and the 1996 Office of Naval Research Young Investigator Award. She was an Executive Committee Member of the 2002 International Conference on Acoustics, Speech, and Signal Processing, Orlando, Florida, May 2002. She has been an Associate Editor of the IEEE Transactions on Signal Processing since 1999 and an Associate Editor of the IEEE Signal Processing Magazine since 2003. She is presently a member of the Signal Processing Theory and Methods (SPTM) Technical Committee of the IEEE Signal Processing Society.

EURASIP Journal on Applied Signal Processing 2004:5, 707–726

High Capacity Downlink Transmission with MIMO Interference Subspace Rejection in Multicellular CDMA Networks

Henrik Hansen
INRS-Télécommunications, Université du Québec, Place Bonaventure, 800 de la Gauchetière Ouest, Suite 6900, Montréal, Québec, Canada H5A 1K6
Email: henrik.b.hansen@ericsson.com

Sofiène Affes
INRS-Télécommunications, Université du Québec, Place Bonaventure, 800 de la Gauchetière Ouest, Suite 6900, Montréal, Québec, Canada H5A 1K6
Email: affes@inrs-emt.uquebec.ca

Paul Mermelstein
INRS-Télécommunications, Université du Québec, Place Bonaventure, 800 de la Gauchetière Ouest, Suite 6900, Montréal, Québec, Canada H5A 1K6
Email: mermel@inrs-emt.uquebec.ca

Received 31 December 2002; Revised 18 September 2003

We proposed recently a new technique for multiuser detection in CDMA networks, denoted by interference subspace rejection (ISR), and evaluated its performance on the uplink. This paper extends its application to the downlink (DL). On the DL, the information about the interference is sparse, for example, spreading factor (SF) and modulation of interferers may not be known, which makes the task much more challenging. We present three new ISR variants which require no prior knowledge of interfering users. The new solutions are applicable to MIMO systems and can accommodate any modulation, coding, SF, and connection type. We propose a new code allocation scheme denoted by DACCA which significantly reduces the complexity of our solution at the receiving mobile. We present estimates of user capacities and data rates attainable under practically reasonable conditions regarding interferences identified and suppressed in a multicellular interference-limited system. We show that the system capacity increases linearly with the number of antennas despite the existence of interference. Our new DL multiuser receiver consistently provides an Erlang capacity gain of at least 3 dB over the single-user detector.

Keywords and phrases: CDMA, downlink multiuser detection, interference rejection, space-time processing, code allocation, MIMO.

1. INTRODUCTION

Third generation wireless systems will deploy wideband CDMA (W-CDMA) [1, 2] access technology to achieve data transmission at variable rates. Standards [1] call for transmission rates up to 384 Kbps for mobile users and 2 Mbps for portable terminals. On the downlink (DL), high-speed DL packet access (HSDPA) [3, 4] allows for transmission rates up to about 10 Mbps in the conventional single-input single-output (SISO) channel and about 20 Mbps in the multiple-input multiple-output (MIMO) channel. It is expected that most of the traffic will be DL due to asymmetrical services like FTP and web browsing. The DL will therefore become the limiting link, and only high DL performance can give the network operator maximal revenue from advanced radio-network technologies.

MIMO [5] and multiuser detection (MUD) [6, 7, 8] are both very promising techniques for high capacity on the DL in wireless systems. In a noise-limited MIMO system, Shannon capacities increase linearly in SNR with the number of antennas [5] instead of logarithmically as in the SISO system. Recent studies, however, have shown that in an interference-limited MIMO system, this linear relationship is not achieved due to the multiple-access interference (MAI) [8]. In [9, 10], it was shown that the gain in such systems is basically limited to the antenna beamforming gain at the receiver. In terms

of system capacity,[1] this means that the Erlang capacity increases linearly with the number of antennas. MUD can significantly increase the capacity further especially when interference is pronounced [11]. It is therefore of prime concern to establish a cost-effective solution that combines MIMO and MUD for optimal DL performance.

MUD is a challenging problem, not only for the uplink (UL), but even more so for the DL. On the UL, the receiving base station knows the connection characteristics of all in-cell users. The DL MUD problem is more difficult because the terminal has no knowledge of active interference, its spreading codes, SF, modulation, coding, and the connection type (packet switched or circuit switched). Furthermore, complexity considerations are more important because terminals are limited by size and price and are restricted in available power.

Most previous work was aimed at the UL (e.g., [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21]). For the DL, blind adaptive MMSE solutions based on generalizations of single-user detectors (SUDs) have previously been proposed for the STAR [22] receiver in [23], denoted STAR GSC, and for the RAKE [24] receiver in [25], denoted the generalized RAKE (G-RAKE). These solutions are characterized by low complexity and low risk because they impose the least change to an established technology. But they require the use of short codes and the capacity gain in a practical DL environment is limited to about 1.5–2.5 dB for the G-RAKE [26, 27] (and expectedly in the same range for STAR-GSC). In [28], a solution which offers potentially higher capacity gains is presented. Relying on the use of orthogonal variable spreading factor (OVSF) [29] codes, it probes for interference on the OVSF code tree at a high SF level in order to identify and reject codes with significant energy. This solution is complex because it rejects interference at a high SF level and is defined for rejection of in-cell interference only.

We propose a new class of MUD solutions for DL multicellular interference-limited CDMA-based MIMO systems. These new solutions are all DL variants of the previously presented interference subspace rejection (ISR) technique [30] and are therefore referred to as DLISR. The DLISR variants do not rely on prior knowledge of the interference and its properties (e.g., modulation, coding scheme, and connection type). Nor do they attempt to estimate the SF and modulation of the interference. DLISR takes advantage of a concept we denote by virtual interference rejection (VIR) combined with a new OVSF code allocation scheme denoted dynamic power-assisted channelization code allocation (DACCA). VIR reduces complexity in the receiver by attacking interference at a low SF. DACCA provides information to the terminal about the location of interference in the OVSF code-space. DLISR does not necessarily require VIR and DACCA. However, when combined with these new concepts, DLISR provides very high performance at very low complexity. As a benchmark, we consider the PIC [16, 17] with soft decision (PIC-SD), which can also exploit the VIR and DACCA techniques.

Performance of MUD detectors heavily relies on the distribution of interference. For instance, MUD typically offers very significant performance gains if the interference arrives from one strong source. However, if interference arrives from numerous weaker sources, MUD performance approaches SUD performance. In order to provide convincing results with regards to real-world applications, it follows that interference must be modelled realistically. We have therefore implemented a precise model as shown in Figure 1. First we establish a realistic realization of the interference using a radio-network simulator (RNS); then this information is used for the link-level simulations to assess the BER for DLISR, PIC-SD, and the SUD. Repeating the cycle many times and combining the results, we arrive at system-level capacity estimates. Our link-level simulator makes assumptions very similar to those in W-CDMA standards. We do not rely on any a priori knowledge of the channel; instead we employ the STAR receiver [22] to estimate the channel. Simulations show that our new MUD consistently offers a gain of at least 3 dB over SUD based on maximal ratio combining (MRC) for QPSK and as much as 6.5–8.1 dB for 16 QAM. Our solution demonstrates a linear growth in Erlang capacities with the number of receiving antennas.

The main contributions of this paper are as follows. Most importantly, we propose a new solution for DL MIMO MUD in CDMA-based systems. We present the concepts of VIR and DACCA to allow for effective operation of DLISR and to reduce the complexity at the receiver significantly. Finally, we propose an RNS to generate realistic realizations of the interference in the DL MIMO system.

The paper is organized as follows. We present our link-level signal model in Section 2. In Section 3, we derive DLISR and introduce DACCA and VIR. The RNS is presented in Section 4. Then our system-level simulation results are presented in Section 5. Finally, our conclusions are given in Section 6.

2. LINK-LEVEL SIGNAL MODEL

In this section, we discuss the link-level signal model and discuss briefly basic estimation issues. The radio-network model, which is important for the quality of our simulation results, is presented later in Section 4. Section 2.1 presents an overview of the MIMO model, Section 2.2 provides the mathematical model of the signals, and finally, Section 2.3 considers estimation of the basic parameters.

2.1. Overview of the MIMO model

We consider a DL MIMO CDMA system as illustrated in Figure 2. Let (u, v) denote the user with index $u = 1, \ldots, U_v$ connected to the cell with index $v = 1, \ldots, N_{\text{CELLS}}$. We define a cell as one site sector, that is, a three-sector site has three cells. U_v is the number of users connected to the cell with index v and N_{CELLS} is the number of cells considered.

[1] System capacity is a measure of the total system capacity. Shannon capacity is a measure of the single link spectral efficiency.

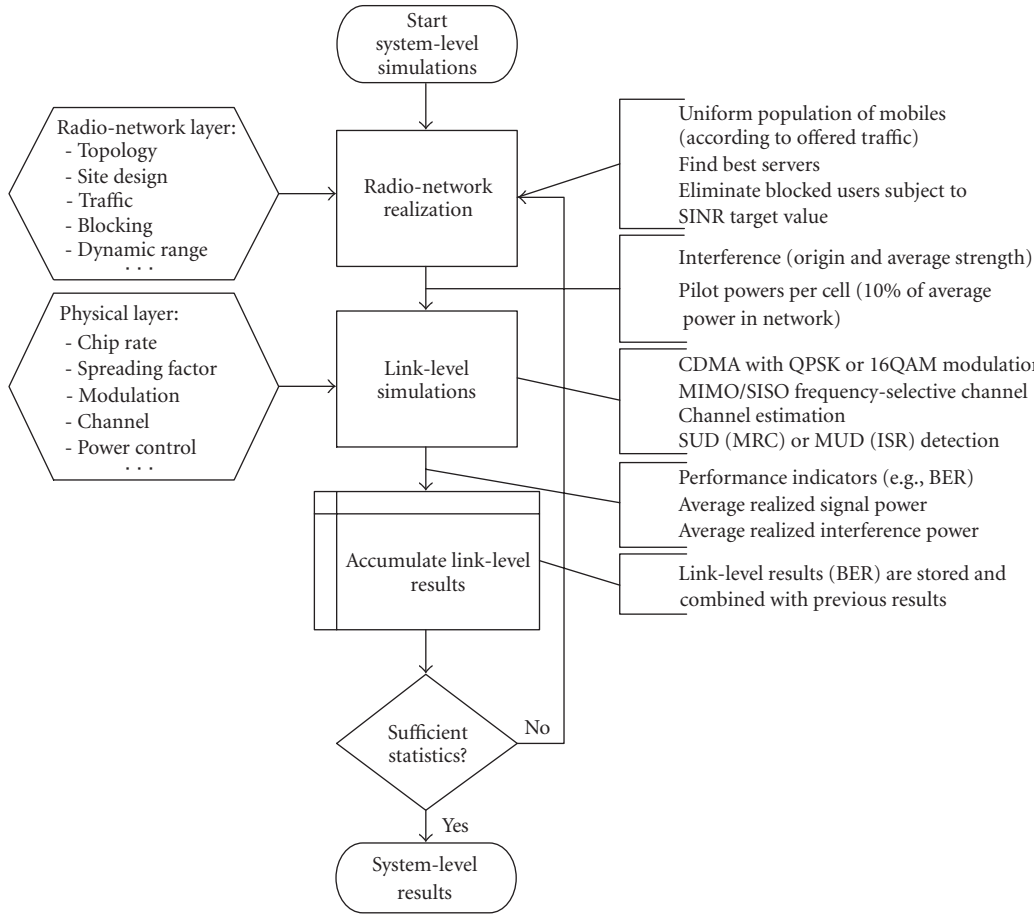


Figure 1: Organization of operations for radio-network and link-level simulations.

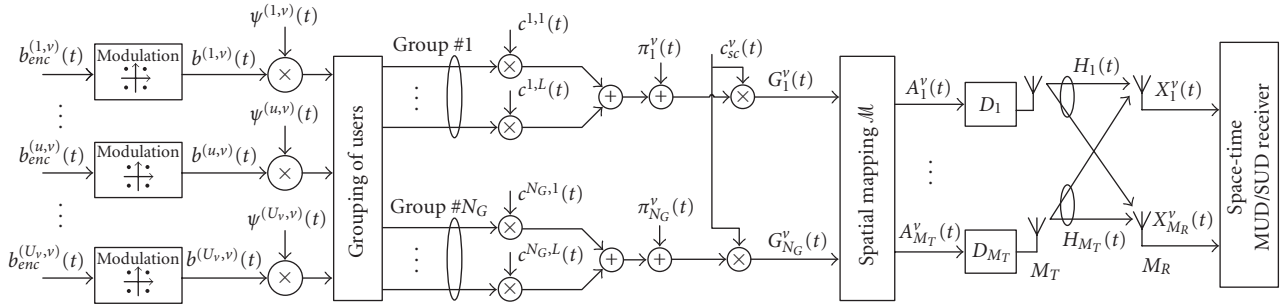


Figure 2: Block diagram of an $M_T \times M_R$ MIMO transceiver structure with emphasis on transmitter and channel.

Let $b_{\text{enc}}^{(u,v)}(t)$ represent a BPSK stream of encoded information bits. The encoded data bits are modulated according to the modulation scheme (we consider QPSK and 16-QAM in this paper) and scaled by the desired transmit amplitude $\psi^{(u,v)}(t)$. The stream of modulated channel symbols are switched to one of N_G groups such that the user (u, v) is assigned to the group $g_{(u,v)}$. The modulated symbols are then spread by a user-specific channelization code, increasing the rate by the SF, $L = T/T_c$, where T is the time duration of one modulated symbol and T_c is the chip duration. The channelization code is defined as $c_{\text{ch}}^{(u,v)}(t) = c^{(g_{(u,v)}, i_{(u,v)})}(t)$, where $i_{(u,v)}$ is the index to one of the codes of the group. Assignment

of groups and channelization codes are discussed below. We add a pilot unique to each group scaled by the desired pilot amplitude, that is, $\pi_g^v(t) = \psi_\pi^v(t)c_\pi^{g,v}(t)$, where $(\psi_\pi^v(t))^2$ is the desired pilot power and $c_\pi^{v,g}(t)$ is a PN code unique to the group. Finally the cell-specific scrambling code $c_{sc}^v(t)$ is applied to yield the group-specific signal $G_g(t)$, $g = 1, \ldots, N_G$. The N_G groups of signals, organized in the vector $G^v(t) = [G_1^v(t)^T, \ldots, G_{N_G}^v(t)^T]^T$, are next spatially mapped onto M_T antennas by the $M_T \times N_G$-dimensional matrix $\mathcal{M}$ to arrive at the M_T-dimensional signal, $A^v(t) = \mathcal{M}G^v(t)$. $A^v(t)$ is transmitted over the channel $H^v(t)$ and received by the mobile unit with M_R antennas. If $\mathcal{M}$ has full rank, the groups are mapped orthogonally in space onto the transmitting antennas. Orthogonal spatial mapping is possible as long as the condition $(M_T \geq N_G)$ is satisfied. In this paper, we assume that $M_T = N_G$ and therefore the Hadamard matrix is useful. The Hadamard matrix ensures both orthogonal transmission in space and equal distribution of power between the transmitting antennas.[2] If a different delay D_m is employed at each transmitting antenna, we obtain time diversity. This may be attractive in low-diversity situations, but in a typical multipath channel possibly with multiple receive antennas, the sufficient diversity is available and extra time diversity may degrade performance because channel identification is made more difficult [31] (see also footnote 16). In our simulations, we consider multipath mostly with antenna diversity reception and therefore we have used $D_m = 0$. Simulations (not shown herein) have demonstrated that using different antenna delays generally results in the same or slightly worse performance when multipath propagation is considered.

We now return to the concepts of grouping and channelization-code design. Channelization codes are grouped into N_G groups with L codes in each group. The purpose of grouping is to allow for user capacities beyond the SF. Each group will contain channelization codes unique to the group. Codes are correlated between groups but mutually uncorrelated within groups. The spatial mapping $\mathcal{M}$ serves to separate groups further by assigning orthogonal spatial signatures at transmission. Users are assigned a group and a channelization code pair $(g_{(u,v)}, i_{(u,v)})$ on a first-come first-serve basis in the following order: $(g, i) = (1, 1), (1, 2), \ldots, (1, L), (2, 1), \ldots, (N_G, L)$. Let $\mathcal{G}_g$ denote the set of channelization codes in group g. By wise definitions of the code groups, intragroup (preferably orthogonal) as well as intergroup correlations are controlled. It is noteworthy that since the same scrambling code is used across groups, cross-correlation properties, once set by proper choice of channelization code sets, are preserved after scrambling. As an example, we consider the following two groups of SF = 4 channelization codes:

$$\mathcal{G}_1 : [c^{1,1}, c^{1,2}, c^{1,3}, c^{1,4}] = \left[\begin{bmatrix}+1\\+1\\+1\\+1\end{bmatrix}, \begin{bmatrix}+1\\-1\\+1\\-1\end{bmatrix}, \begin{bmatrix}+1\\+1\\-1\\-1\end{bmatrix}, \begin{bmatrix}+1\\-1\\-1\\+1\end{bmatrix}\right];$$
$$\mathcal{G}_2 : [c^{2,1}, c^{2,2}, c^{2,3}, c^{2,4}] = \left[\begin{bmatrix}+1\\-1\\+1\\+1\end{bmatrix}, \begin{bmatrix}+1\\+1\\+1\\-1\end{bmatrix}, \begin{bmatrix}+1\\-1\\-1\\-1\end{bmatrix}, \begin{bmatrix}+1\\+1\\-1\\+1\end{bmatrix}\right]. \quad (1)$$

Intragroup correlations are zero for both groups and intergroup correlations are always −6 dB (relatively). Using these code groups as a baseline, we can easily derive an OVSF tree for both groups (see [29]). It is easy to show that intergroup correlations reduce with higher SFs. For SF lower than four some code pairs will have nonzero correlation. Lower SFs must therefore be employed in practice with extra coordination between groups. In this example, the two code groups have been rotated by 45° with respect to each other.

2.2. Multiuser multicell downlink signal model

We now present a mathematical formulation of the received signal. A useful diagram is shown in Figure 3. We consider the DL of a cellular CDMA system, where the mobile is equipped with an antenna array of M_R sensors. At time t, the observation vector received at the antenna array of M_R sensors at the mobile terminal can be defined as follows:

$$X(t) = \begin{bmatrix} X_1(t) \\ \vdots \\ X_{M_R}(t) \end{bmatrix} = \sum_{v=1}^{N_{\text{CELLS}}} X^v(t) + N(t), \quad (2)$$

where

$$X^v(t) = \sum_{u=1}^{U_v} X^{(u,v)}(t) + \sum_{g=1}^{N_G} X_\pi^{g,v}(t) \quad (3)$$

is the signal arriving from the vth cell, $X^{(u,v)}(t)$ is the contribution from the (u, v)th user, $X_\pi^{g,v}(t)$ is the pilot signal of the gth group of the vth cell, and $N(t)$ is the thermal noise assumed to be uncorrelated additive white Gaussian noise (AWGN).

The contribution of the (u, v)th user, $X^{(u,v)}(t)$, to the received signal $X(t)$ is given by

$$X^{(u,v)}(t) = \sum_{m=1}^{M_T} H_m^v(t) \otimes A_m^{(u,v)}(t), \quad (4)$$

where $H_m^v(t)$, $m = 1, \ldots, M_T$, is the M_R-dimensional channel vector from the mth transmitting antenna to the receiving antenna array with M_R sensors, and $A_m^{(u,v)}(t)$, $m = 1, \ldots, M_T$, is the contribution of the (u, v)th user to the signal transmitted at the mth antenna. Each dimension corresponds to one transmit antenna. The total transmitted signal arriving from

[2] We use Hadamard matrices with a power-2 number of transmit antennas. Otherwise, with an arbitrary number of transmit antennas, we resort to orthogonal Vandermonde-structured matrices. Current investigations suggest significant advantages due to exploitation of such spatial mapping matrices when combined with closed-loop PC and MIMO transmit diversity [31].

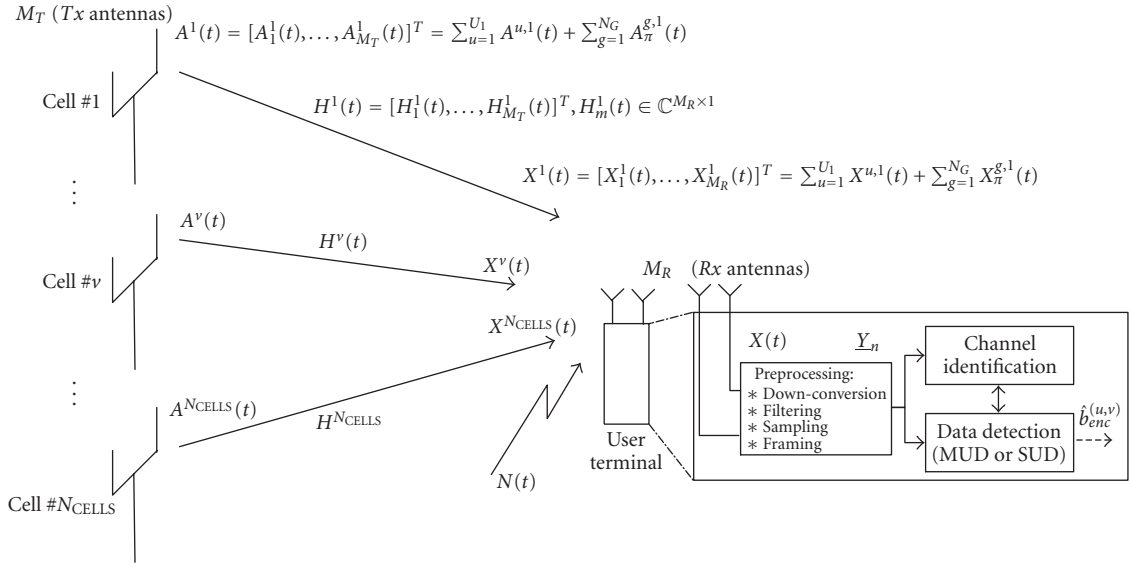


Figure 3: Network-level signal diagram.

the (u, v)th user is defined as follows:

$$A^{(u,v)}(t) = \begin{bmatrix} A_1^{(u,v)}(t) \\ \vdots \\ A_{M_T}^{(u,v)}(t) \end{bmatrix} = \mathcal{M} G^{(u,v)}(t) \tag{5}$$

with

$$G^{(u,v)}(t) = \begin{bmatrix} \vdots \\ G_g^{(u,v)}(t) \\ \vdots \end{bmatrix},$$
$$G_g^{(u,v)}(t) = \begin{cases} \psi^{u,v}(t) c^{u,v}(t) b^{(u,v)}(t) & \text{if } (u,v) \in \mathcal{G}_g, \\ 0 & \text{if } (u,v) \notin \mathcal{G}_g, \end{cases} \tag{6}$$

where $(\psi^{(u,v)}(t))^2$ is the power, $c^{(u,v)}(t) = c_{ch}^{(u,v)}(t) c_{\text{SC}}^{(u,v)}(t)$ is the spreading code (channelization code + scrambling code), and $b^{(u,v)}(t)$ denotes the modulated symbols. For lack of space, we do not detail the contribution of the pilots to the received signal, but it follows the pattern of (4), (5), and (6) by replacing $X^{(u,v)}(t)$ by $X_\pi^{(u,v)}(t)$, $A^{(u,v)}(t)$ by $A_\pi^{(u,v)}(t)$, and $G^{(u,v)}(t)$ by $G_\pi^{(u,v)}(t)$, respectively.

We adopt the common assumption that the channel response can be modeled as a tapped delay line with Rayleigh-faded tap gains [32]. The M_R-dimensional channel response vector from the transmitting cell to the mobile unit with M_R antenna elements is therefore given as follows:

$$H^v(t) = \begin{bmatrix} \vdots \\ H_m^v(t) \\ \vdots \end{bmatrix}, \quad m = 1, \ldots, M_T \tag{7}$$

with

$$H_m^v(t) = L_{\text{LOSS}} \sum_{p=1}^{P} h_{m,p}^v(t) \varepsilon_p^v(t) \delta\left(t - \tau_p^v - D_m\right), \quad m = 1, \ldots, M_T, \tag{8}$$

where $\delta(t)$ is the Dirac delta function, $\tau_p^v(t) \in [0, T)$ are the multipath time delays for $p = 1, \ldots, P$. Note that the physical path delays are the same for all receiving antennas but delay differences may optionally be imposed at transmission. $h_{m,p}^v(t) = [h_{1,m,p}^v(t), \ldots, h_{M_R,m,p}^v(t)]^T$ is the unit-norm propagation vector, $\varepsilon_p^v(t)^2$, $p = 1, \ldots, P$, are the power fractions along each path such that $\sum_{p=1}^{P} \varepsilon_p^v(t)^2 \doteq 1$, D_m is an additional transmit delay associated with each transmit antenna, and L_{LOSS} is the path loss. In practice, L_{LOSS} is largely compensated by power control and we therefore fix it to unity in what follows. Note that this implies that the expected gain of $H_m^v(t)$ is one (by definition).

At reception, the M_R-dimensional received signal is first filtered by the pulse-matched filter, then sampled and framed into observation vectors containing Q consecutive symbols of the desired user (the signal is first down converted in reality). We define the preprocessing step through the function $\mathcal{P}$, $\underline{V} = \mathcal{P}(U(t), n) : \mathcal{C}^{M_R \times 1} \to \mathcal{C}^{M_R(QL + L_\Delta) \times 1}$ as follows (see [30] for more details):

$$U^\phi(t) = \frac{1}{T_c} \int U(t + t') \phi(t') dt',$$
$$\underline{V} = \Big[U^\phi(nT + aT_c)^T, U^\phi(nT + (a+1)T_c)^T, \ldots, U^\phi(nT + (a + QL + L_\Delta - 1)T_c)^T \Big]^T, \tag{9}$$

where L_Δ is an extra margin to account for the delay spread,

$\phi(t)$ is the square-root raised-cosine shaping pulse, and a is an offset that guarantees that the targeted symbols $nQ + k$, $k = 0, \ldots, Q-1$, occur within the duration of the observation frame. Without loss of generality, we set $a = 0$ in what follows. With this definition, we can now define the preprocessed observation as

$$\underline{Y}_n = \sum_{v=1}^{N_{\text{CELLS}}} \sum_{u=1}^{U_v} \psi_n^{(u,v)} \underline{Y}_n^{(u,v)} + \sum_{v=1}^{N_{\text{CELLS}}} \sum_{g=1}^{N_G} \psi_{\pi,n}^{g,v} \underline{Y}_{\pi,n}^{g,v} + \underline{N}_n^{\text{pwn}}, \tag{10}$$

where $\underline{Y}_n = \mathcal{P}(X(t), n)$, $\underline{Y}_n^{(u,v)} = \mathcal{P}(X^{(u,v)}(t)/\psi^{(u,v)}(t), n)$, $\underline{Y}_{\pi,n}^{g,v} = \mathcal{P}(X_\pi^{g,v}(t)/\psi_\pi^{g,v}(t), n)$, $\underline{N}_n^{\text{pwn}} = \mathcal{P}(N(t), n)$, and $\psi_n^{(u,v)} = \psi^{(u,v)}(nQT)$. $\underline{Y}_n^{(u,v)}$ is to be understood as the contribution of the (u,v)th user to the nth observation. It is useful to decompose its contributions as follows:

$$\psi_n^{(u,v)} \underline{Y}_n^{(u,v)} = \psi_n^{(u,v)} \sum_{k'} b_{nQ+k'}^{(u,v)} \underline{Y}_{k',n}^{(u,v)}, \tag{11}$$

and $\underline{Y}_{k',n}^{(u,v)}$ is to be understood as the signature of the $nQ + k'$th symbol. We next define the user (d, v_d) as the desired user (v_d denotes the best server of user d) and let g_d denote the group to which the user is assigned. We now isolate the desired signal and pilot in (10) from intersymbol interference (ISI) and in-cell/out-cell MAI as follows:

$$\underline{Y}_n = \underbrace{b_{nQ+k}^{(d,v_d)} \psi_n^{(d,v_d)} \underline{Y}_{k,n}^{(d,v_d)}}_{\text{desired signal}} + \underbrace{\psi_{\pi,n}^{g_d,v_d} \underline{Y}_{\pi,n}^{g_d,v_d}}_{\text{desired pilot}} + \underbrace{\underline{I}_{\pi,n}^{(d,v_d)}}_{\text{pilot interference}} + \underbrace{\underline{I}_{\text{ISI},k,n}^{(d,v_d)}}_{\text{ISI}} + \underbrace{\sum_{u=1,\, u\neq d}^{U_{v_d}} \underline{I}_n^{(u,v_d)}}_{\text{in-cell MAI}} + \underbrace{\sum_{v=1,\, v\neq v_d}^{N_{\text{CELLS}}} \sum_{u=1}^{U_v} \underline{I}_n^{(u,v)}}_{\text{out-cell MAI}} + \underbrace{\underline{N}_n^{\text{pwd}}}_{\text{AWGN}}, \tag{12}$$

where with reference to (11), we have

$$\begin{aligned} \underline{I}_{\text{ISI},k,n}^{(d,v_d)} &= \sum_{k'\neq k} \psi_n^{(u,v)} b_{nQ+k'}^{(u,v)} \underline{Y}_{k',n}^{(d,v_d)}, \\ \underline{I}_n^{(u,v)} &= \sum_{k'} \psi_n^{(u,v)} b_{nQ+k'}^{(u,v)} \underline{Y}_{k',n}^{(u,v)}, \\ \underline{I}_{\pi,n}^{d,v_d} &= \sum_{v=1}^{N_{\text{CELLS}}} \sum_{g=1}^{N_G} \psi_{\pi,n}^{g,v} \underline{Y}_{\pi,n}^{g,v} - \psi_{\pi,n}^{g_d,v_d} \underline{Y}_{\pi,n}^{g_d,v_d}. \end{aligned} \tag{13}$$

2.3. Basic parameter estimation principles

In our simulations, we estimate every parameter as needed with no prior information assumed known to the receiver. To estimate the multipath delays and the multipath gains, we employ a variant of the STAR receiver [22] as discussed in Section 2.3.1. MRC data detection (used by the SUD considered herein), power estimation, and signal-to-interference-plus-noise ratio (SINR) estimation for PC are then discussed in Sections 2.3.2, 2.3.3, and 2.3.4, respectively.

2.3.1. STAR: the spatio-temporal array-receiver

We employ a variant of the STAR receiver [22] which mainly differs in the despreading operation. Instead of using the code of the desired user for despreading, we employ a more generalized code for despreading. We consider multicodes to represent one cooperative code for despreading, which is a combination of concatenating codes in time (i.e., consecutive symbols by data remodulation) and combining over channels. For the channel of the desired user, we combine the pilot code with the data remodulated spreading code over Q consecutive symbols. For other channels, we employ only the pilot for channel identification with STAR.

2.3.2. MRC beamforming and data detection

The signal component $s_{nQ+k}^{(u,v)} = \psi_n^{(u,v)} b_{nQ+k}^{(u,v)}$ contains sufficient statistics for the estimation of both data and power. The signal component can be estimated by MRC which is optimal in white noise. With reference to (12), the MRC combiner for the k'th symbol of user (u,v) is as follows:

$$\underline{W}_{\text{MRC},k',n}^{(u,v)} = \frac{\underline{Y}_{k',n}^{(u,v)}}{\left|\left|\underline{Y}_{k',n}^{(u,v)}\right|\right|^2}, \quad k' = 0, \ldots, Q-1, \tag{14}$$

and then the signal component is estimated as

$$\hat{s}_{nQ+k'}^{(u,v)} = \underline{W}_{\text{MRC},k',n}^{(u,v),H} \underline{Y}_n. \tag{15}$$

A beamformer for the pilots can be defined accordingly. Note that we use the term beamformer because $\underline{W}_{\text{MRC},k',n}$ works in both space and time. The transmitted symbol is estimated as the symbol in the signal constellation which is the closest to $\hat{b}_{nQ+k}^{(u,v)} = \hat{s}_{nQ+k}^{(u,v)}/\hat{\psi}_n^{(u,v)}$, where $\hat{\psi}_n^{(u,v)}$ is the estimated power (Section 2.3.3).

2.3.3. Power estimation

We consider two different power estimators. The first estimator first estimates the amplitude

$$\hat{\psi}_n^{(u,v)} = \alpha \hat{\psi}_{n-1}^{(u,v)} + (1-\alpha)\frac{1}{Q} \Re\left\{ \sum_{k'=0}^{Q-1} \left(\hat{b}_{nQ+k'}^{(u,v)}\right)^H \hat{s}_{nQ+k'}^{(u,v)} / \left|\hat{b}_{nQ+k'}^{(u,v)}\right| \right\}, \tag{16}$$

where α is a forgetting factor. The power estimate is then found by squaring the amplitude estimate. The second estimator estimates the power directly:

$$\left(\hat{\psi}_n^{(u,v)}\right)^2 = \alpha \left(\hat{\psi}_{n-1}^{(u,v)}\right)^2 + (1-\alpha)\frac{1}{Q} \sum_{k'=0}^{Q-1} \left|\hat{s}_{nQ+k'}^{(u,v)}\right|^2. \tag{17}$$

The latter is biased because it effectively estimates the combined signal and interference noise power. The estimator in (16) has less bias and is more accurate because the filtering appears before the squaring; but it requires that the decision feedback (DF) is decent. The estimator of (17) is useful to

Table 1: Definition of the constraint matrix of each mode. (Each generic column $\hat{\underline{C}}_{j,n}$ is normalized to one.)

ISR mode	$\hat{\mathbf{C}}_n = [\ldots, \hat{\underline{C}}_{j,n}, \ldots]$	N_c (number of constraints)
Hypotheses (H) (constraint/symbol/interferer)	$[\ldots, \hat{\underline{Y}}^i_{k,n}, \ldots]$	$(Q+2)NI$
Realizations (R) (constraint/interferer)	$\left[\ldots, \sum_{k=-1}^{Q} \hat{b}^i{}_{nQ+k} \hat{\underline{Y}}^i_{k,n}, \ldots\right]$	NI
Total realization (TR) (constraint/total MAI)	$\left[\sum_{i=1}^{NI} \hat{\psi}^i_n \sum_{k=-1}^{Q} \hat{b}^i{}_{nQ+k} \hat{\underline{Y}}^i_{k,n}\right]$	1

estimate the power of the interference (where decision feedback is difficult), whereas the estimator of (16) is used for the desired pilot and data signal.

2.3.4. SINR estimation

The PC command is determined by comparing the SINR estimate at the receiver with the target SINR. We use the following estimator for the SINR:

$$\hat{\gamma}_n^{(d,v_d)} = \left(\frac{\hat{\psi}_n^{(d,v_d)}}{\hat{\sigma}_n^{(d,v_d)}}\right)^2, \tag{18}$$

where $\hat{\psi}_n^{(d,v_d)}$ results from (16) and $\hat{\sigma}_n^{(d,v_d)}$ is an estimator for the postcombined noise, which is obtained by estimating the total received power (of all users) after combining and then subtracting the estimated power of the desired user.

3. DOWNLINK INTERFERENCE SUBSPACE REJECTION

Our main contribution is a new efficient and cost-effective MUD solution for DL MIMO, DLISR. DLISR is based on ISR previously presented for UL systems [30]. It incorporates new variants of ISR modes which are specially suited for the more problematic DL case. In particular, DLISR employs VIR, which involves rejection of virtual users instead of physical users. VIR has many benefits especially when it is combined with DACCA. Neither VIR nor DACCA are indispensable for DLISR; however, capacity gains and especially complexity reductions are achieved when combined. We next review ISR in Section 3.1. Then we define DACCA and VIR and introduce DLISR. Finally, we discuss the attractive complexity features of our new solutions.

3.1. Review of ISR

In this section, we provide an overview of ISR. For a more complete picture, see [30]. The basic ISR recipe is to form a constraint matrix $\hat{\mathbf{C}}$ with a column span which spans the estimated interference subspace. In a second step, the observation is mapped away from the interference subspace spanned by $\hat{\mathbf{C}}$ by constrained spatio-temporal projection; thereby, MAI and ISI are reduced significantly. The desired signal can then be estimated by conventional beamforming,[3] for example, MRC.

[3] We use the term beamforming because our solution works in space and time. However, the term filter-combiner could equally well be used.

The projection and combining steps can also be carried out in a single beamforming step. The ISR beamformer $\underline{W}^{(d,v_d)}_{k,n}$, $k = 0, \ldots, Q-1$, is defined by

$$\mathbf{Q}_n = \left(\hat{\mathbf{C}}_n^H \hat{\mathbf{C}}_n\right)^{-1}, \tag{19}$$

$$\mathbf{\Pi}_n = \mathbf{I}_{N_T} - \hat{\mathbf{C}}_n \mathbf{Q}_n \hat{\mathbf{C}}_n^H, \tag{20}$$

$$\underline{W}^{(d,v_d)}_{k,n} = \frac{\mathbf{\Pi}_n \hat{\underline{Y}}^{(d,v_d)}_{k,n}}{\hat{\underline{Y}}^{(d,v_d)H}_{k,n} \mathbf{\Pi}_n \hat{\underline{Y}}^{(d,v_d)}_{k,n}}, \tag{21}$$

where $\mathbf{I}_{N_T}$ denotes an $N_T \times N_T$ identity matrix, and $N_T = M_R(QL + L_\Delta)$ is the total space-time dimension. First, we form the projector $\mathbf{\Pi}_n$ orthogonal to the constraint matrix $\hat{\mathbf{C}}_n$. Second, we project the estimated response vector $\hat{\underline{Y}}^{(d,v_d)}_{k,n}$ and normalize it to yield the ISR beamformer $\underline{W}^{(d,v_d)}_{k,n}$.

3.1.1. ISR modes

The ISR modes differ in the construction of the constraint matrix. Table 1 defines the constraint matrix of each mode when considering only MAI rejection and a pedagogical illustration is provided in Figure 4 which links the modes to the composition of the constraint matrix. In the table, NI denotes the number of interfering signals to be rejected, and i is the index to a subset of MAI signals which we strive to reject. Note that for simplicity, Table 1 defines the composition of the constraint matrix when only MAI is rejected, but it is easily generalized to also incorporate ISI rejection by adding columns of the estimated ISI. Of the modes previously presented, three merit discussion here.

In the ISR-hypothesis mode (ISR-H), every symbol signature[4] of the selected interfering users is rejected individually. This mode does not require DF. If the channel is known, selected interfering users can be rejected perfectly but the white noise is enhanced. ISR-H was found to perform poorly on the UL because of the large noise enhancement associated with the many constraints [30]. Its application to the DL, however, is more appealing due to the adverse near-far situations there as we will witness later.

In the ISR-realizations mode (ISR-R), we do not form a null constraint for each symbol signature of each interfering user. Instead, we reconstruct the sequence of symbols over

[4] "Symbol signature" is understood as the unmodulated symbol.

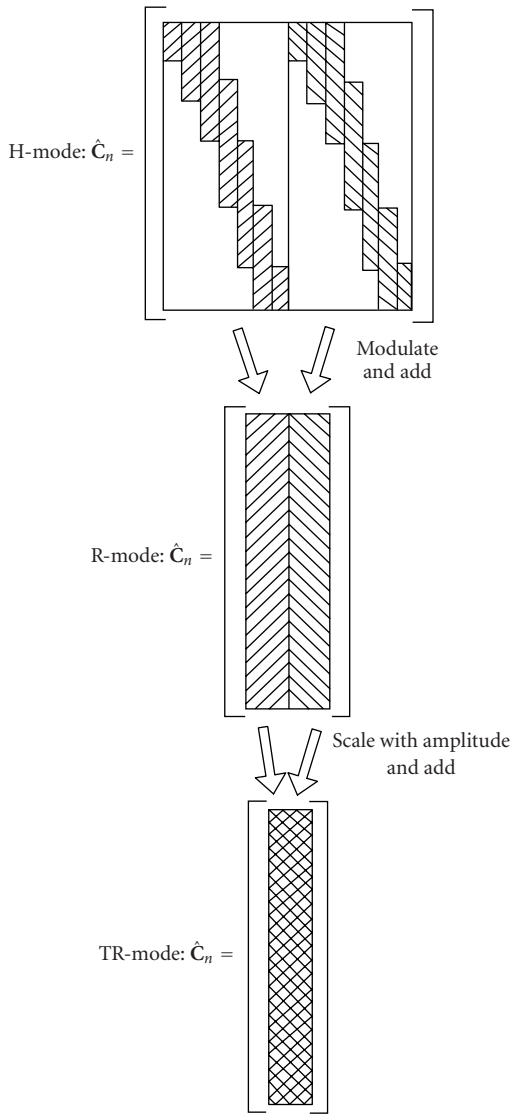


Figure 4: Relation between H, R, and TR modes can be illustrated from the composition of the constraint matrix.

the duration of the observation frame. The R mode therefore requires DF. These decisions are obtained from MRC-based decisions (Section 2.3.2). The number of constraints is reduced with ISR-R giving less white noise enhancement at the cost of reduced near-far resistance.

In the ISR-total realization (ISR-TR) mode, we reconstruct interference using DF as in the R mode, then we add the reconstructed interfering users scaled by their estimated amplitudes to form one constraint only. ISR-TR, in addition to DF, also requires power estimates (Section 2.3.3). The TR mode has negligible white noise enhancement but also the worst near-far resistance.

Before we introduce the proposed application of ISR to the DL (DLISR) in Section 3.4, we will present DACCA and VIR in Sections 3.2 and 3.3, respectively.

3.2. DACCA

We propose a strategy for channelization code allocation of user data channels at the base station, which we denote by DACCA. With DACCA, the base station dynamically reassigns channelization codes to the users at a low rate with the aim of concentrating energy in the left-hand side of the OVSF tree. We propose a simple metric for code assignment as the product between each user's output power and SF, denoted by the power-SF product (PSFP) in the following.[5] DACCA is illustrated in Figure 5a. The aim is to fill the OVSF tree from left to right subject to the PSFP of users. The desired outcome is a concentration of power at the left-hand side of the OVSF tree. Figure 6 shows the probabilistic origin of the interference for a random mobile in a network. The distributions were obtained with the aid of the RNS to be presented in Section 4 and corresponds to a soft-blocking rate (SBR) (see Section 4.2) of 20%, processing gain (PG) of 16, and an offered traffic of $T_{\text{OFF}} = 4$ Erl. In this paper, the PG is defined as the SF, L, multiplied by the number of receive antennas, that is, PG $= M_R L$. Otherwise, the assumptions specified in Section 5.2.1 apply. We observe that most of the interference is generated by just a few users. For example, 30% of the total interference arrives from the strongest in-cell interferer and the sum of only two interferers accounts for almost half the interference. With DACCA, therefore, most of the interference power can be concentrated in a relatively small portion of the OVSF code space. It is the pronounced near-far situations on the DL which make DACCA especially interesting.

Dynamic code assignment and reassignment strategies have previously been considered in [33, 34]. The goal in previous works was to reduce code blocking and limit the code reassignment rate. Instead, the purpose of DACCA is to provide the mobile with a priori knowledge on where to look for interference and at the same time concentrating the interference energy in a small portion of the OVSF tree. DACCA shares some similarities with the strategy denoted "leftmost" in [34], namely, users are assigned to the leftmost available code in the OVSF tree. DACCA imposes additional restrictions because it both strives to assign the leftmost codes and at the same time to achieve the best possible concentration of power at the left-hand side of the OVSF tree. Therefore, DACCA will exacerbate the probability of code blocking and more frequent code reassignments must be performed by UTRAN (UMTS terrestrial radio access network). The need for frequent reassignment is satisfied by reassigning codes at a low rate of 75 Hz in our simulations. Regarding code blocking, previous results [34] indicate that a load (i.e., number of OVSF codes in use divided by the SF) of 50% yields a code-blocking rate less than 1%. Comparing this blocking with the loads we can achieve (see Section 5) and the SBR on the air interface, it is reasonable to deem code blocking

[5] In practice, the assignment rule should be more complex because not all SFs are equally probable and because assigned codes mutually preclude each other; for example, assignment of a high SF code blocks any parents of that code to be assigned. This issue is irrelevant for this work because we consider only one SF for all users in our simulations.

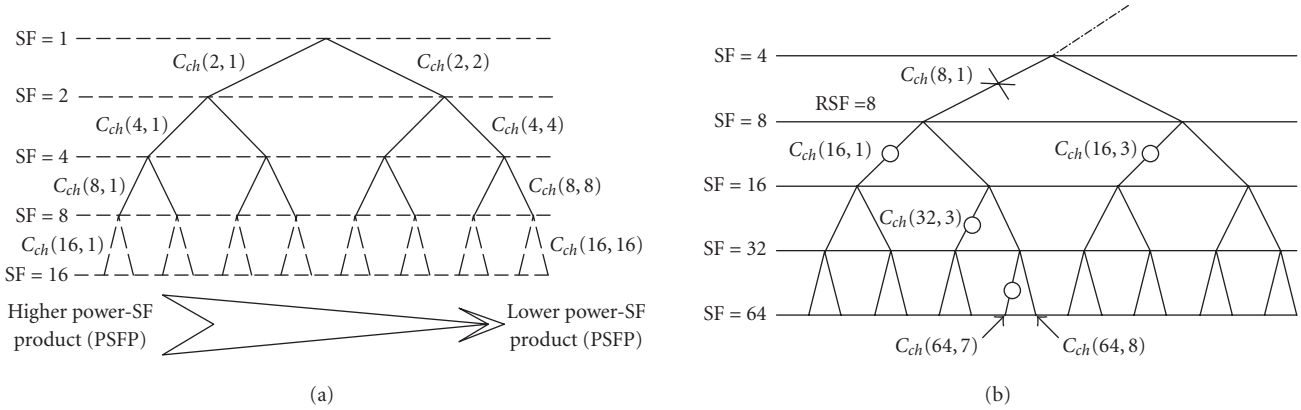


Figure 5: DACCA and VIR illustrated. (a) In DACCA, users are assigned channelization codes according to their PSFP. (b) Interference rejection is aimed at a low SF when VIR is employed.

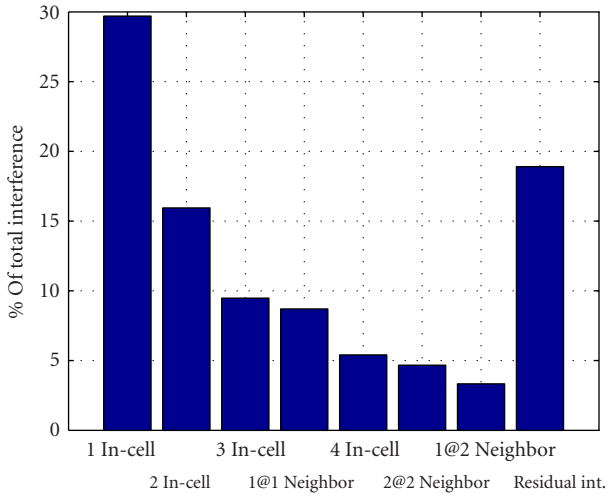


Figure 6: Relative power of interferers arriving from different sources. 1 In-cell is the strongest in-cell interferer, 1@1 neighbor is the strongest interference from first-tier neighbors.

to be a minor drawback of DACCA. Note that DACCA does not conflict with 3G standards because channelization codes can be allocated almost freely by UTRAN. Only the primary CPICH and the primary CCPCH have predefined channelization codes [29].

3.3. Virtual interference rejection

VIR involves rejection of interference targeting a channelization code with low SF (rejection SF (RSF)) although no physical users may be assigned this code. VIR is particularly interesting in the context of OVSF trees [29]. The idea is to target one or more virtual channelization codes with low RSF L_R and reject these codes as if they were physical users. The advantage is that any offspring (in the OVSF tree) from the rejected virtual code is also rejected; therefore, multiple interfering users are rejected, targeting only a few virtual channelization codes.

It is noteworthy that VIR targets the channelization codes. In practice, the channelization codes are repeated at the rate $L_R T_c$, scrambled by the scrambling code and filtered by the channel response. A mathematical formulation of VIR is provided in [35]; here we will provide an example of VIR. Consider the segment of an OVSF tree starting at an SF of 8 shown in Figure 5b. Codes that are circled are in active use. Consider the virtual channelization code $c_{ch}(8,1)$, marked with an "x." We reconstruct all required segments[6] of $c_{ch}(8,1)$, apply the appropriate scrambling code, and filter them by the estimated channel response. Then we reject all reconstructed segments. It then follows that all descendants are rejected irrespective of their SF and modulation; that is, the interferer with SF = 16 assigned to code $c_{ch}(16,1)$, the code with SF = 32 assigned to $c_{ch}(32,3)$, and the one with an SF of $L = 64$ assigned to $c_{ch}(64,7)$, respectively, are all rejected. The code $c_{ch}(64,8)$ is rejected although it is not active and the code $c_{ch}(16,3)$ is active but not rejected. Preferably, codes that are not active should not be rejected.

When VIR is combined with DACCA, cancelling the leftmost code at any RSF ideally causes the highest possible fraction of the interference to be rejected. The efficiency of VIR is, therefore, enhanced when DACCA is used. If DACCA is not employed, the RSF must be higher to minimize the number of rejected inactive codes. This will increase complexity significantly (see Section 3.5) and possibly degrade performance.

An idea similar to VIR was considered in [28]; however, the targeted SFs were very high SFs instead of very low SFs like in VIR. The idea there is that one interferer at a low SF is equivalent to numerous high SF virtual users. With VIR, the

[6] Required segments means those segments which will have contributions within the current observation frame. If the delay spread is low, there are approximately $QL/L_R + 1$ contributing segments per targeted virtual code (including two edge symbols).

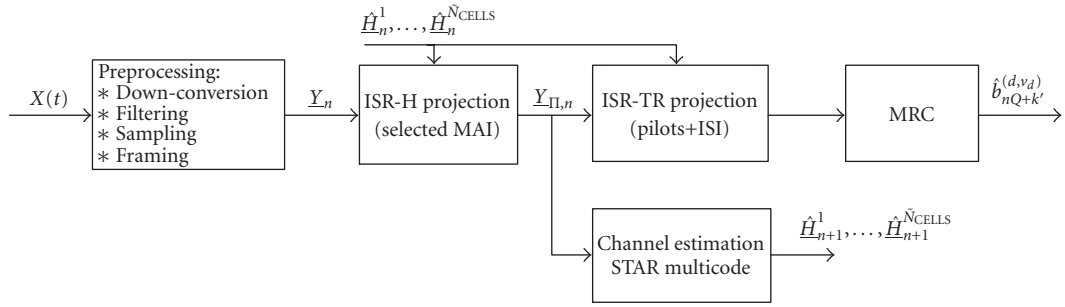


Figure 7: Proposed DLISR receiver structure. $\tilde{N}_{\text{CELLS}} \leq N_{\text{CELLS}}$ is the number of (virtual) interferers selected for rejection.

Table 2: Important characteristics of new ISR variants for DL MIMO.

Feature strategy	Requires DACCA?	Knows int. codes?	Knows int. SF?	Knows int. modulation?	Applicable to PS/CS int.?	Knows int. coding?
ISR-H-FC	Yes	No	No	No	Yes	No
ISR-H-BC	No[1,2]	No	No	No	Yes[3]	No
ISR-R-SD	No[1,2]	No	No	No	Yes[3]	No
PIC-SD	No[1,2]	No	No	No	Yes[3]	No
MRC	No	No	No	No	Yes	No

[1]Performance gain with DACCA.
[2]Complexity reduction with DACCA.
[3]Possible performance penalty for PS.

idea is opposite: one low SF code constitutes many interferers assigned to physical OVSF codes of higher SFs.

3.4. DLISR

Compared to the UL, DL MUD is characterized by a lack of information regarding the interference. A mobile generally has no knowledge of the interfering users' codes, modulation, connection type, and coding. This information is only available for the pilots and the desired signal. Therefore, the interference rejection is conveniently split into two steps: in the first step, we remove the MAI and in the second step, we remove the ISI and the pilots as shown in Figure 7. The TR mode has shown excellent performance in [30] with the lowest possible complexity. Therefore, the TR mode is well suited for application in the second step regardless of the solution applied in the first step. For lack of space, we disregard further details and focus on the more important first step in the following. Improved near-far resistant channel estimation [36] may be achieved by using the near-far resistant observation $\underline{Y}_{\Pi,n} = \mathbf{\Pi}_n \underline{Y}_n$ (see (20)) offered as an intermediate step according to Figure 7. It is therefore natural to use $\underline{Y}_{\Pi,n}$ for the purpose of channel identification because it is offered without additional complexity. In the following, we present three variants of DLISR. Two variants based on ISR-H and are denoted by DLISR-H with fixed constraints (DLISR-H-FC) and DLISR-H with best constraints (DLISR-H-BC), respectively. The final variant is based on the R mode with soft decision and is denoted by DLISR-R-SD. For the purpose of comparison, we also consider the PIC-SD. Important properties of the DLISR variants, PIC-SD, and MRC are summarized in Table 2.

3.4.1. DLISR-H-FC

DLISR-H-FC is the simplest of all variants. The idea is to blindly reject the same OVSF code subspace according to a fixed strategy. Obviously, this mode is relevant only when DACCA is employed.

Whenever a virtual-user code is rejected, white noise is enhanced. It can be shown that if the spreading is real, the noise enhancement is given as follows:[7]

$$\kappa \simeq \frac{N_T - 2}{N_T - 2 - N_c}, \tag{22}$$

where N_c is the number of interfering signals to be rejected. The observation frame with dimension $N_T = M_R(QL + L_\Delta)$ (see (10)) spans $(QL + L_\Delta)/L_R$ segments of the targeted code with SF L_R. Due to asynchronism and multipath propagation, additional symbols will contribute at the edges. Assuming that the delay spread is insignificant, it follows that the number of constraints in (22) is $N_c \simeq \lceil (QL + L_\Delta)/L_R \rceil + 1$.

Using (22) and the probabilistic distribution of interference (see Figure 6), we can identify a solution that optimizes the trade-off between noise enhancement and interference reduction. Table 3 lists the relative reduction of interference and noise enhancement for different strategies. The first row

[7]If we strive to reject a subspace with dimension N_c contained within the total dimension N_T, a fraction of the desired signal energy is rejected as well. It is reasonable to assume that this fraction is approximately $(N_T - N_c)/N_T$. Therefore, the noise compared to the desired signal is enhanced by $N_T/(N_T - N_c)$. A more accurate development of (22) will be shown in a later contribution.

Table 3: Choosing the best strategy.

Number of interferers to reject (in-cell/neighbor 1/neighbor 2)	1/0/0	2/0/0	3/0/0	3/1/0	4/1/0	4/2/0	4/2/1	4/3/1
Interference reduction (dB)	1.53	2.64	3.47	4.41	5.11	5.83	6.42	6.98
Noise enhancement (dB)	0.32	0.66	1.04	1.46	1.91	2.43	3.01	3.67
Net gain (dB)	1.20	1.97	2.43	2.95	3.19	3.40	**3.41**	3.31

Table 4: Complexity estimates of ISR variants in Mops.

Task	DLISR-H-FC	DLISR-H-BC	DLISR-R-SD	PIC-SD	MRC	Comment
STAR	300	300	300	300	300	100 Mops per channel [37]
Reconstruct, $\hat{\underline{Y}}_{nQ+k}^{(u,v)}$	259	363	363	311	0	—
$\mathbf{C}^H \underline{Y}_n$	61	61	31	0	0	$\mathbf{C}$ has higher dimension for H-variants but it is sparse
$\mathbf{Q} = \mathbf{C}^H\mathbf{C}$	246	246	61	0	0	—
$\mathbf{Q}^{-1}\mathbf{C}^H \underline{Y}_n$	282	282	1	0	0	—
Margin	40%	40%	40%	40%	40%	—
Total Mops at RSF = 8	1607	1753	1058	855	420	Appropriate when DACCA is employed
Total Mops at RSF = 16	4185	4476	1800	1218	420	—
Total Mops at RSF = 32	13728	14308	3828	1944	420	Appropriate when DACCA is absent

identifies the interferers rejected, for example, 2/1/0 means the two strongest in-cell virtual users plus the strongest out-cell user of the neighbor cell with the strongest pilot channel. In the second row, the noise enhancement is computed according to (22). The net gain peaks at 3.41 dB suggesting that the best strategy is to reject 4 in-cell virtual users, 2 virtual users from the strongest neighbor, and one virtual user from the second strongest neighbor. In reality, the strategy (which is fixed) should be selected according to the highest load during busy hour. This ensures optimal performance at peak load and always satisfactory performance at lower loads.

3.4.2. DLISR-H-BC

In the DLISR-H-BC variant, we estimate the power in the virtual subspace of the serving cell and all cells in the neighbor list. The power is estimated subject to the RSF which may represent many virtual users. The best constraints are computed along the same lines as in Table 3, but the interference reduction is based on the estimated power and not the statistical mean. This version hence adapts easily to fast fading and will attempt to reject interference most efficiently. This strategy therefore ensures that we always follow an optimal rejection strategy, provided that the powers are estimated properly and the update is done frequently.

DLISR-H-BC is more complex than DLISR-H-FC because it needs to probe the interference subspace and has to decide which constraints to reject for best performance. It can, however, work in the absence of DACCA although DACCA simplifies probing. In the absence of DACCA, interference is not generally concentrated at a low SF virtual code; it may therefore be necessary to probe the OVSF tree at higher RSF levels. This increases complexity and reduces the accuracy of probing because a few strong sources can be estimated more reliably than many weak sources.

3.4.3. DLISR-R-SD

In this variant, we reconstruct the virtual users using soft decision. Working at a low RSF, the N_v OVSF virtual codes which contain most power are selected. These codes are reconstructed as virtual users' signals, and soft decision estimates based on MRC estimation are used. Note that hard decision FB is not usually an option on the DL and the fact that one virtual code is the contribution of many physical interfering users makes hard decision even more complicated.

3.4.4. PIC-SD

As a benchmark, we consider the PIC [16, 17] with SD FB, and denote it by PIC-SD. We follow the same steps as for DLISR-R-SD; but the reconstructed interference is subtracted instead of nulled. Obviously, PIC-SD, like DLISR, takes advantage of both VIR and DACCA to improve performance and lower complexity.

3.5. Computational complexity of DLISR

We provide complexity estimates in Table 4 assuming VIR with an RSF of 8 for all DLISR variants, PIC-SD, and MRC. We have also listed results for an RSF of 16 and 32, respectively. We have detailed the most demanding tasks and a margin of 40% has been added to account for all other operations not listed. We assume that RSF/2 virtual codes are rejected and that three cells are actively monitored. Complexity is specified in Mops, where one operation is defined as a complex multiply-add. The numbers are appropriate for $M_R = 1$. Roughly speaking, complexity is invariant to the SF of the desired user, and grows linearly with the number of receiving antennas. The results for RSF = 8 relate to the situation where DACCA is employed (as in our later simulations). When DACCA is not employed, an RSF of 8 is too low. We simulated the leftmost and random code-allocation

schemes, for which details are omitted for lack of space, and found that an RSF of about 32 must be employed if the leftmost strategy is used instead of DACCA, and even higher RSF must be employed if random code allocation is employed.

The complexity of the matrix inversion is very modest. For the R-variant, it is negligible because the dimension is only 4 (with RSF = 8). H-variants have higher complexities associated with the inversion but, although not evident, there are huge savings because **Q** is band diagonal as a result of VIR (low RSF approach).[8] PIC-SD does not require matrix inversion and therefore has a complexity advantage over DLISR which, however, is vanishing for low RSFs.

When VIR and DACCA are employed, the complexity of our solution is moderate. Our MUD solutions require from about 1.1 to 1.7 Gops. Today's high-end signal processors offer speeds of more than 10 Gops. A requirement of 1.1–1.7 Gops is therefore reasonable for a mobile terminal application where cost and power consumption must be kept low. The feasibility becomes even more evident when compared with SUD (STAR-MRC); our solution requires only about 2.5–4 times the complexity of SUD. Note that our SUD candidate, STAR [22] with MRC, is comparable in complexity to the RAKE [37], which is used in current implementations. DLISR-R-SD is less complex than the H-variants but the difference is only about 50% which is considered unimportant.

If DACCA is not employed, VIR is still applicable (and should be used!) but it must target a higher RSF. This exacts a significant complexity increase of about 4–8 times[9] when comparing at RSF = 32 which as argued is a good choice when DACCA is not employed. The complexity of the *R*-variants is now four times less than the *H*-variants. It is therefore in much favor of DLISR-R-SD when higher RSFs are used.

4. RADIO-NETWORK SIMULATOR

The purpose of the RNS is to provide a realistic picture of the distribution of the users and how they interfere with each other. This information is then used for the link-level simulations.

The RNS starts by uniformly populating users in a homogeneous cell grid which we name the test network. Using propagation estimates, it iteratively blocks users either due to coverage or interference limitations. Once the network arrives at a stable condition, the RNS outputs the realized interference. A stable condition is characterized as one where all users can achieve the required SINR without being blocked (i.e., without exceeding the maximum power offered by the base station cell). First, we provide a mathematical formulation in Section 4.1. Then we outline the algorithm in Section 4.2.

[8] When the columns of **C** are arranged appropriately. Note that ISR is invariant to the arrangement of the columns of **C**.

[9] Values are in the high end. We feel confident that many computational tricks can be exploited to reduce the complexity of reconstruction, and so forth.

4.1. Network-level signal model

The mobile unit always strives to achieve a certain SINR which is sufficient to provide a certain QoS. If the serving cell is not able to supply the power required by a mobile, the mobile is *blocked*. Below we define the link budget which is useful to assess the SINR at the target mobile subject to transmitted power, propagation loss, interference, and so forth. First, we briefly discuss the propagation model which is essential to the later considerations.

4.1.1. Propagation model

We consider the following simplified form of the Okumura-Hata propagation model [38, 39]:

$$L_{\mathrm{PATH}}(\overline{u},\overline{v}_u) = L_0 + 10K_P \log_{10}\left(\frac{\max\{\mathrm{dist}(\overline{u},\overline{v}_u), d_0\}}{d_0}\right) + \Gamma_{\mathrm{LNF}}, \tag{23}$$

where K_P is the propagation exponent (typically 3.5–4 for urban environments), L_0 is an offset which relates to the morphology, $\overline{u} = 1,\ldots,N_U$ is the user index where N_U is the total number of users in the network attempting a call, $\overline{v} = 1,\ldots,\overline{N}_{\mathrm{CELLS}}$ is the cell index where $\overline{N}_{\mathrm{CELLS}}$ is the total number of cells, and $\mathrm{dist}(\overline{u},\overline{v})$ is the distance between the mobile and the cell. Finally, Γ_{LNF} models the log-normal fading (LNF) and is assumed to be a normally distributed random variable, that is, $\Gamma_{\mathrm{LNF}} \in N\{0,\sigma^2_{\mathrm{LNF}}\}$. Note that the variables $\overline{u}$, $\overline{v}$, and $\overline{N}_{\mathrm{CELLS}}$ by definition are different from u, v, and N_{CELLS} first introduced in Section 2.1.[10] Considering that signals arriving from the same spatial direction will experience similar LNF, we introduce the following location-dependent modeling of the LNF:

$$\Gamma_{\mathrm{LNF}} = X_{\mathrm{LNF}}\cos(\Theta) + Y_{\mathrm{LNF}}\sin(\Theta), \tag{24}$$

where Θ is the angle between the mobile and the cell, and where X_{LNF}, Y_{LNF} are independent zero-mean Gaussian distributed random variables with variance σ^2_{LNF}.

4.1.2. Generic multicell multiuser link budgets

We define the set $\mathcal{G}_B$ which contains the indices of all mobiles which are blocked. If the mobile $(\overline{u},\overline{v})$ is not blocked (i.e., $(\overline{u},\overline{v}) \notin \mathcal{G}_B$), we have

$$S(\overline{u},\overline{v}) = P_{\mathrm{OUT}}(\overline{u},\overline{v}) + G_T(\overline{u},\overline{v}) + G_R - \Delta_{\mathrm{MARG}}(\overline{u},\overline{v}) - L_{\mathrm{PATH}}(\overline{u},\overline{v}) \Longleftrightarrow \tag{25}$$

$$P_{\mathrm{OUT}}(\overline{u},\overline{v}) = S(\overline{u},\overline{v}) - G_T(\overline{u},\overline{v}) - G_R + \Delta_{\mathrm{MARG}}(\overline{u},\overline{v}) + L_{\mathrm{PATH}}(\overline{u},\overline{v}), \tag{26}$$

where S is the signal strength at the input of the receiving

[10] For instance, (u,v) is the uth user connected to the vth cell. However, $\overline{u}$ means the $\overline{u}$th user in the network (possibly blocked).

antenna,[11] P_{OUT} is the power fed to the transmitting antenna,[12] G_T and G_R are the gains of the transmitting and receiving antennas, respectively, Δ_{MARG} accounts for additional engineering margins (e.g., PC margin), and L_{PATH} is the path loss between the serving cell and the user equipment defined in (23). We assume that the mobile antenna gain is independent of its location and G_R is therefore location independent.

Let γ_{REQ} specify the required SINR in dB for a specified QoS. We assume that the required target value is the same for all mobiles.[13] Assume that $\overline{v}_u$ is the serving cell of the $\overline{u}$th mobile; then the required signal power at the input to the receiver, say $S_{\text{REQ}}(\overline{u}, \overline{v}_u)$, is given in dB as follows:

$$\begin{aligned} S_{\text{REQ}}(\overline{u}, \overline{v}_u) &= \gamma_{\text{REQ}} + 10\log_{10}\left(\mathcal{N}^{\#} + \mathcal{I}^{\#}(\overline{u})\right) \\ &= \gamma_{\text{REQ}} + \mathcal{N} + 10\log_{10}\left(1 + \frac{\mathcal{I}^{\#}(\overline{u})}{\mathcal{N}^{\#}}\right), \end{aligned} \tag{27}$$

where $\mathcal{N}$ is the user-independent thermal noise power, and $\mathcal{I}(\overline{u})$ is the total MAI received at the mobile $\overline{u}$ to be defined in (29). We use the # sign to differentiate a physical value from its dB equivalent. We next combine (26) and (27) to find the required transmitted power:

$$\begin{aligned} P_{\text{REQ}}(\overline{u}, \overline{v}_u) &= \gamma_{\text{REQ}} + \mathcal{N} + 10\log_{10}\left(1 + \frac{\mathcal{I}^{\#}(\overline{u})}{\mathcal{N}^{\#}}\right) \\ &\quad - G_T(\overline{u}, \overline{v}_u) - G_R + \Delta_{\text{MARG}} + L_{\text{PATH}}(\overline{u}, \overline{v}_u) \end{aligned} \tag{28}$$

and can now define the received interference as follows:

$$\begin{aligned} \mathcal{I}(\overline{u}) = 10\log_{10}\Bigg(&\sum_{\overline{u}' \neq \overline{u}} 10^{(P_{\text{REQ}}(\overline{u}', \overline{v}_u) - K_{\text{ORTH}} - L_{\text{LOSS}}(\overline{u}', \overline{v}_u))/10} \\ &+ \sum_{\overline{v}' \neq \overline{v}_u} \sum_{\overline{u}'} 10^{(P_{\text{REQ}}(\overline{u}', \overline{v}') - L_{\text{LOSS}}(\overline{u}', \overline{v}'_u))/10} \Bigg) \\ &+ 10\log_{10}(\text{PG}), \end{aligned} \tag{29}$$

where K_{ORTH} is the orthogonality factor, a measure of the orthogonality loss due to multipath propagation (a typical value is 2 dB), and $\text{PG} = LM_R$ is the PG.

We define the best server $\overline{v}_u$ of the mobile with index $\overline{u}$ as the serving cell $\overline{v}'$ which requires the lowest output power to satisfy the SINR target:

$$\overline{v}_u = \arg\min_{\overline{v}'} \left\{ P_{\text{REQ}}(\overline{u}, \overline{v}') \mid \mathcal{I}(\overline{u}) = 0 \right\}. \tag{30}$$

This definition suggests that we consider the best server as the cell with the strongest signal since the interference is assumed zero. In handover situations, this may not be true because of hand-over hysteresis, congestion, or load balancing. We also note that the best server, according to the definition used, may not be the best choice because of orthogonal transmission, which implies that a weaker server may occasionally have better effective SINR. If the best server cannot supply the required power needed, the user is blocked. Blocking occurs either due to *coverage blocking* (excessive path loss) or *interference blocking* (excessive interference).

The size of the test network can be limited using *wraparound* to mitigate the edge effect. To implement wraparound, we place nine virtual images of the test network in all directions (south, south-east, east, and so forth). In the computations, the image of a cell which gives the strongest signal is chosen. For instance, to compute the required power in (28), we compute it for both the target cell and also for all its nine replicas; and then select the replica which gives the highest signal strength. Nondocumented simulations support the efficiency of network wrapping, which allows for the use of test networks smaller than 25 sites (5 by 5 grid).

4.2. RNS algorithm

The object of the algorithm is to locate mobiles in the network so that all nonblocked mobiles experience satisfactory SINR. The algorithm estimates this by uniformly distributing $\overline{N}_{\text{CELLS}} T_{\text{OFF}}$ users, where T_{OFF} is the *offered traffic*. Then it blocks users until a stable solution is found. The RNS algorithm is illustrated by the flowchart in Figure 8. We identify a cell near the center of the grid as the target cell. Assume that the noise floor $\mathcal{N}$ and the maximum output power P_{MAX} have been defined. Initially, the sets of blocked mobiles are empty, that is, $\mathcal{S}_{\text{CB}} = \varnothing$ (coverage) and $\mathcal{S}_{\text{IB}} = \varnothing$ (interference).

(1) Distribute $\overline{N}_{\text{CELLS}}$ cells on a map in a hexagonal grid.
(2) Randomly populate the test network with $\overline{N}_{\text{CELLS}} T_{\text{OFF}}$ users.
(3) Compute for every mobile-cell pair the power required for service (see (28)).
(4) Identify the best server for every user (see (30)).
(5) Users with $P_{\text{REQ}} > P_{\text{MAX}}$, P_{MAX} being the maximum output power which can be assigned to any individual user, are deemed to be coverage blocked and are added to the set $\mathcal{S}_{\text{CB}}$. The fraction of users blocked estimates the coverage blocking probability, that is, $\hat{\text{Pr}}_{\text{CB}} = \text{size}\{\mathcal{S}_{\text{CB}}\}/(\overline{N}_{\text{CELLS}} T_{\text{OFF}})$.
(6) Compute the total received interference for all remaining users (see (29)).
(7) Compute required output power for all remaining users (see (28)).
(8) Block users which have $P_{\text{OUT}} > P_{\text{MAX}}$ and add to the set of interference blocked users $\mathcal{S}_{\text{IB}}$.
(9) If all users that are not blocked can achieve the required SINR, stop, otherwise go to 7.

We note that the noise floor and the maximum output powers are chosen arbitrarily. By appropriate choices, we can target any desired coverage blocking.

$\hat{\text{Pr}}_{\text{IB}} = \text{size}\{\mathcal{S}_{\text{IB}}\}/(\overline{N}_{\text{CELLS}} T_{\text{OFF}})$ estimates the probability of interference blocking. An estimate of the total SBR is then $\hat{\text{Pr}}_B = \hat{\text{Pr}}_{\text{CB}} + \hat{\text{Pr}}_{\text{IB}}$.

[11] We assume for simplicity that the reference point is the antenna connector and avoid this way to consider feeder losses and so forth which are immaterial for our purpose.

[12] The connector of the transmitting antenna is our reference point.

[13] In reality, the target value is determined by the outer-loop PC and differs, due to different channel conditions, slightly across mobiles which otherwise require the same service.

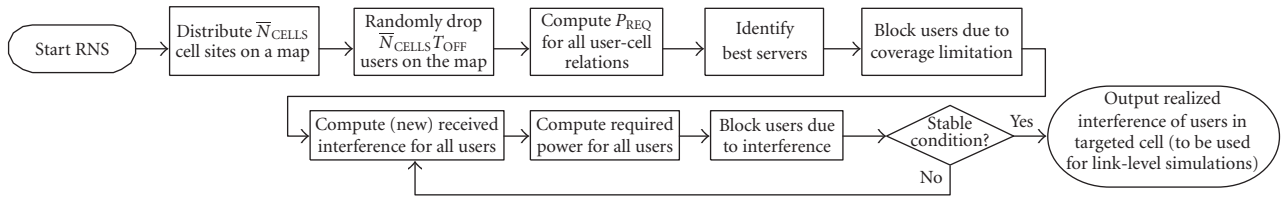


Figure 8: Flowchart of the radio-network simulation operations.

5. SYSTEM-LEVEL SIMULATIONS

5.1. From link-level to system-level results

The simulation model consists of the RNS (Section 4) and the link-level simulator as shown in Figure 1. The RNS provides realistic realizations of the radio-network and the link-level simulator uses this information for BER assessments. We note that a network realization from RNS will result in the same target SINR for all mobiles; however, it is the distribution of the interference which is particularly important on the DL. For a given average offered Erlang traffic, the actual carried traffic is determined by the radio-network SBR (coverage+interference blocking). The SBR is related to the average SINR of the mobiles. This is illustrated in Figure 9 which depicts SBR as a function of the required SINR. The coverage blocking was fixed at 10% by adjusting the maximal output power P_{MAX}. Each SBR estimate is based on 37500 observations with the conditions otherwise stated in Section 5.2.1. Each curve corresponds to the offered traffic level specified in the legend. For a given carried traffic and SBR, the SINR from these curves dictates the PC target SINR which must be used by the link-level simulator. For instance, if we target an SBR of 20% and 4 Erlangs of traffic, the SINR target is 4.5 dB.

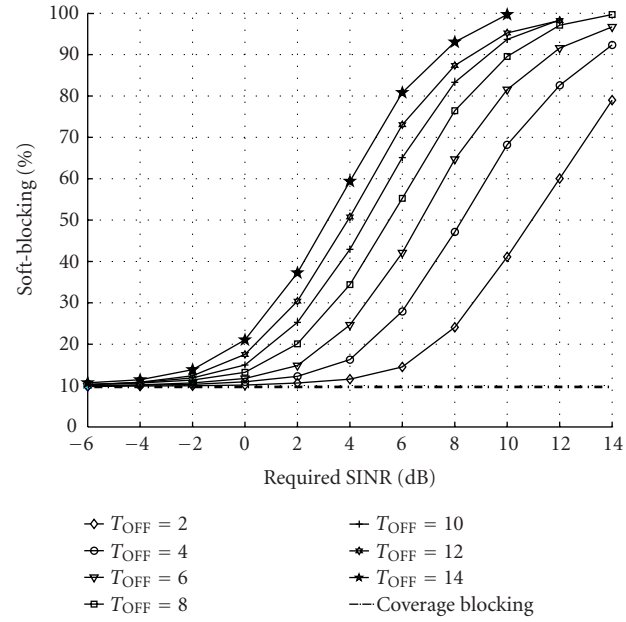


Figure 9: Estimated SBR as a function of the SINR target in a homogeneous system. The coverage blocking is fixed at 10% and the PG is 16.

5.2. Simulation setup

5.2.1. RNS simulations setup

We have considered a homogeneous hexagonal grid of 5 by 5 sites. Wrapping has been used to mitigate the edge effect. The sites have 3 sectors with pointing directions of 0°, 120°, and 240° azimuth, respectively. The antennas are 20 m high and the site-to-site distance is $250\sqrt{5}$ m. We use the vertical/horizontal antenna patterns of Kathrein Werke KG, type number 742212, 1950 MHz antennas with 6° electrical tilt.[14] The orthogonality factor is assumed to be 2.2 dB. We dedicate 10% of the average output power to the CPICH. Coverage blocking has been fixed at 10%. We consider high data rates herein. Therefore, the coverage blocking is chosen moderately high. Table 5 summarizes the settings otherwise used.

5.2.2. Link-level simulations setup

In the link-level simulations, we attempted to approximate the specifications for WCDMA [2]. We have considered low SF operation and high-order modulation schemes (e.g., HSPDA [3, 4]). We use the interference realizations and pilot powers as given by the RNS as inputs to the link-level simulator. We explicitly generate signals from the serving cell and the three strongest neighbors,[15] whereas the interference from the remaining cells is modeled as AWGN. In all our simulations, we consider an SF of $L = 8$ which corresponds to a coded information rate of 480 kbps with QPSK or 960 kbps with 16 QAM when a rate-1/2 coding is assumed. The channel is Rayleigh fading [32] with chip-rate normalized Doppler f_D/R_c, where $R_c = 3.86$ Mcps is the chip rate, and we consider frequency-selective fading with $P = 3$ equal-strength propagation paths with random delays and interpath delays limited to 10 chips. We consider both SISO and MIMO systems. For the desired user, we implement

[14] We tested antenna tilts in the range of 0° to 8° and selected 6° because it provided the highest coverage degree with the choice of site-to-site distance and antenna heights.

[15] Simulations with the RNS show that the desired cell and the three strongest neighbors contribute 95% of the total interference when the carried traffic is 1.6 Erlangs. This number increases with higher traffic loads.

Table 5: Parameters used for the RNS.

Parameter	Assumption	Comments
Cell layout	Hexagonal grid, three-sector sites	Wrapping used to mitigate edge effect
Site-to-site distance	$250\sqrt{5}$ m	—
Antenna pattern	Kathrein 742212 with 6° electrical tilt	Horizontal/vertical patterns
Antenna height	20 m	—
Antenna tilt	6°	Optimized for coverage
SBR	20%	10% due to coverage, 10% due to interference
CPICH power	10%	Relative to average cell output power
Propagation model	Constant + $40\log_{10}\left(\sqrt{x^2+y^2}\right)$	—
Processing gain	16	Equal to $M_R L$
LNF standard deviation	8 dB	—
LNF correlations	Yes	Determined by angle of arrival (see (24))

PC,[16] with a PC correction factor ΔP_{PC} to be updated at a rate of 1500 Hz. The PC message is determined by comparing the estimated SINR (see Section 2.3.4) to the target SINR (coordinated with the RNS). We further impose a transmission delay of $D_{\mathrm{PC}} = 1/(1600\,\mathrm{Hz}) = 0.625$ millisecond and a simulated error rate on the PC bit of $\mathrm{BER}_{\mathrm{PC}} = 10\%$. Modeling closed-loop PC for all users is costly and we have therefore used a simplified model for the interfering users as illustrated in Figure 10. The signal from the unit power source is first scaled by the PC feedback to yield the transmitted power $P_{TX}^{(u,v)}$. The transmitted power is attenuated by the channel, then a Gaussian random variable with variance 0.25 is added to model practical estimation errors in the receiver. This signal is squared and used by the PC decision device to adjust the transmitted power (δ_{bias} compensates for the bias imposed by the simulated noise), and fed back with a delay of D_{PC}. To find the power as experienced by the target mobile, we attenuate the transmitted power by the propagation loss from the serving cell of the interferer to the desired user $L_{\mathrm{LOSS}}^{(u,v)} - L_{\mathrm{LOSS}}^{(d,v_d)}$, to eventually yield $(\psi^{(u,v)}(t))^2$. The values of the propagation losses are obtained as a side product from the RNS.

We use STAR [22] to estimate the channels with the modifications formulated in Section 2.3.1 and Figure 7. DACCA is used with code reallocation at 75 Hz. It is further assumed that DLISR-H-BC updates its constraints at a rate of 300 Hz. Working at an RSF of 8, we found that $N_v = 2M_R$ is a good rule for good performance for DLISR-R-SD in the operating region of interest (about 5% BER). The PIC-SD interestingly shows strong sensitivity to this parameter and the best choice proves to be $N_v = M_R$. The parameters most commonly utilized in the simulations, unless otherwise specified, are summarized in Table 6. All BER estimates reported are derived from at least 150 RNS realizations and each realization was run for at least 19000 symbols.

5.3. SISO with QPSK modulation

We consider first a SISO system with QPSK modulation. The SBR is 20% and the SF is $L = 8$. We employ one channelization-code group composed of $L = 8$ orthogonal Walsh codes. Note that the high soft-blocking ratio considered reflect the high data rate services that we are considering. The carried traffic is hence hard limited to a maximum of 8 users. Code blocking occurs rarely with the traffic loads we consider and its influence is vanishing compared to the SBR of 20%. This claim is true for all simulations cited herein.

Figure 11 shows the uncoded BER as a function of the carried Erlang traffic in the network. Our proposed DLISR variants significantly outperform MRC. They provide Erlang capacity gains of 3.5 dB (DLISR-H-BC) > 3.2 dB (DLISR-R-SD) > 1.6 dB (DLISR-H-FC),[17] respectively, over MRC-based SUD at 5% BER. Although PIC-SD is similar to DLISR-R-SD, it can only offer a gain of 2.6 dB. This illustrates the advantage of linearly constrained beamforming compared to subtraction.[18] With DLISR-H-BC, we achieve the highest spectral efficiency of 0.78 bps/Hz, where spectral efficiency is defined as $\eta_S = \log_2(M_{\mathrm{Mod}})T_{\mathrm{Erl}}/L$, where M_{Mod} denotes the number of symbols in the signal constellation, and T_{Erl} is the carried Erlang traffic (at 5% BER). It is noteworthy that the H mode on the UL did not demonstrate as good performance. The pronounced near-far situations on the DL makes its application attractive.

[16] In the absence of PC, the received power ψ^2 has a χ^2 distribution with standard deviation $\sigma_{\psi^2} = 1/\sqrt{M_T \times M_R}$ that asymptotically approaches the AWGN channel at a very high diversity order $M_T \times M_R \to \infty$. With PC, however, ψ^2 has a log-normal distribution with much weaker standard deviation that quickly approaches the AWGN channel with few antenna elements only, as shown in [31]. Hence, PC significantly increases capacity and reduces the MIMO array size. Indeed, as noted in [31], if we apply the asymptotic expression for the BER in the absence of PC $\Pr[\hat{b} \neq b] = (E_b/N_0)^{-1/M_T \times M_R} = (E_b/N_0)^{-1/\sigma_{\psi^2}^2}$ to the case of active PC (as an approximation), we may expect to obtain (from standard deviation measurements) the same capacity with PC and 3×2 antennas as would be obtained without PC and 30×2 antennas!

[17] We use for simplicity ">" to say that the gain is "greater than."

[18] Note the similarity of these two: with PIC, interference is reconstructed and subtracted; with DLISR-R-SD, the reconstructed interference is nulled.

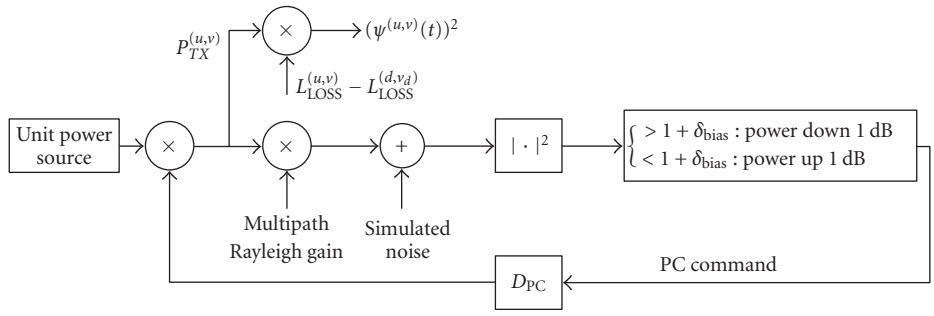


Figure 10: Simplified PC modeling used to model PC (for interfering users only).

Table 6: Parameters used in link-level simulations (unless otherwise specified).

Parameter	Value	Comment
R_c	3.84 Mcps	Chip rate
P	3 (0 dB, 0 dB, 0 dB)	Number of paths (relative average strength)
f_c	1.9 GHz	Carrier frequency
f_e	0 Hz	Frequency error (we assume that frequency offset errors have been compensated, see [40])
f_D	8.9 Hz	Doppler frequency (i.e., 5 Kmph)
L	8	SF
f_{PC}	1600 Hz	Frequency of PC updating
ΔP_{PC}	± 1 dB	PC adjustment
BER_{PC}	10%	Simulated PC BER
$\frac{\delta\tau}{\delta t} T_c$	2 ppm	Symbol clock drift (linear)
$\Delta\tau$	10 chips	Maximal delay spread
f_{DACCA}	75 Hz	DACCA reassignment rate

Note that if we instead compare capacities at BER levels below 3 dB and above 5%, respectively (i.e., 2.5% and 10%), the DLISR-H-BC capacity gains over MRC are 4.5 dB and 2.3 dB, respectively. It is therefore advantageous for DLISR (and MUD solutions in general) to compare along lower BER levels. Our internal studies have shown that 5% is an appropriate target if a rate-1/2 convolutional code with constraint length 9 is assumed. We therefore continue to aim at 5%.

5.4. 2×2 MIMO with QPSK modulation

We now consider a 2×2 MIMO system. The SF is still 8 but the PG is 16 because of the extra antenna. Since we have two transmitting antennas, we have defined two groups of channelization codes. One group consists of $L = 8$ orthogonal Walsh codes; the second group likewise consists of 8 orthogonal codes obtained from the first group by 45° rotation (see the example in Section 2.1). Results are shown in Figure 12.

DLISR-H-BC, DLISR-R-SD, PIC-SD, and MRC achieve the same relative capacity gain of about 3.9 dB compared to SISO. The advantage of linearly constrained beamforming (DLISR-R-SD) compared to subtraction (PIC-SD) is confirmed in this situation as well. The best spectral efficiency of 1.95 bps/Hz is again achieved by DLISR-H-BC. It is obvious that about 3 dB of these gains are due to the antenna gain. The rest is a combination of diversity and statistical multiplexing gain on the air interface. DLISR-H-FC improves in MIMO compared to SISO achieving a relative gain of 5.1 dB compared to SISO. DLISR-H-FC experiences a statistical gain because more users are active in the MIMO system and randomness hence plays a less dominant role. Since this variant uses completely fixed constraints, interference energy is more likely to be concentrated where expected.

5.5. 4×4 MIMO with QPSK modulation

We increase the number of receive and transmit antennas to four. Four code groups were determined by computer simulations where the objective was to minimize the intergroup cross-correlation. Results are shown in Figure 13 (for MRC, ISRDL-H-BC, and PIC-SD). The spectral efficiency of both DLISR and MRC doubles, compared to the 2×2 MIMO system. We are hence able to retain our MUD advantage of at least 3 dB over MRC-based SUD. PIC-SD as usual performs worse than DLISR and can only provide a gain of 2.1 dB over MRC. With DLISR-H-BC, we can now support 17 Erlangs of 480 kbps traffic per sector corresponding to a spectral efficiency of 4 bits/Hz/sector. Comparing SISO, 2×2 MIMO, and 4×4 MIMO, we notice that capacity increases linearly with the number of antennas. This linear relationship was also found by [9] for the MMSE MUD in an interference-limited cellular system. In cellular interference-limited systems, the

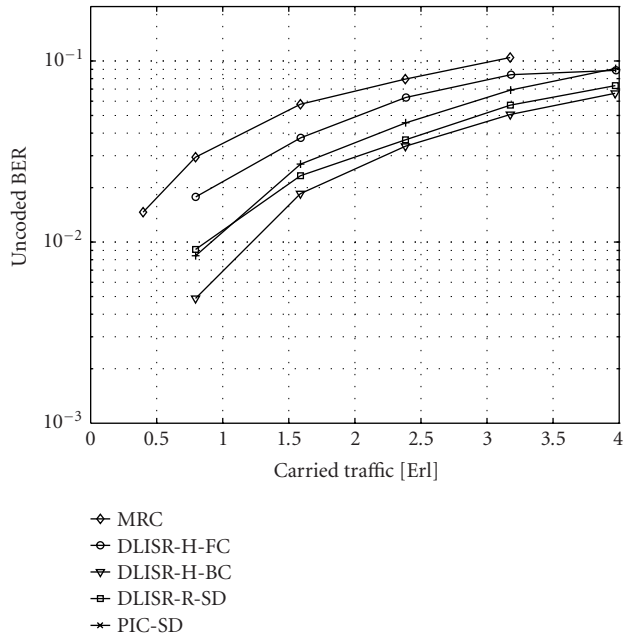


Figure 11: Uncoded BER performance as a function of the offered traffic. The modulation is QPSK and the channel is SISO. The SF is 8 corresponding to an information rate of 480 kbps (rate-1/2 coding assumed).

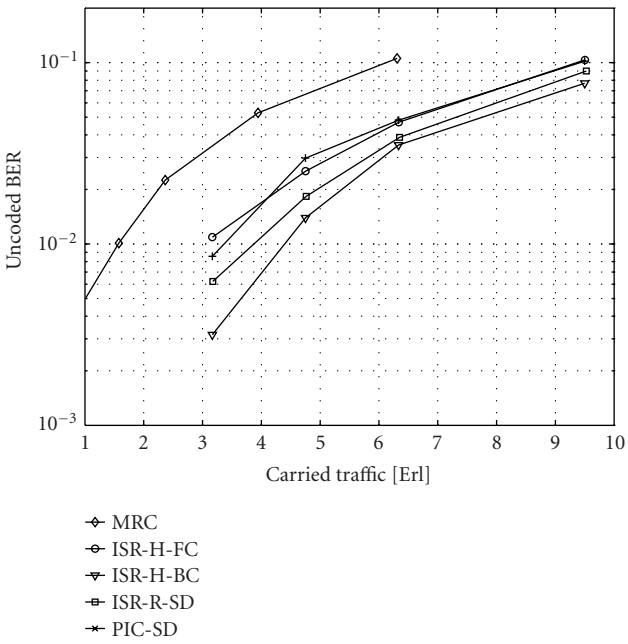


Figure 12: Uncoded BER performance as a function of the offered traffic. The modulation is QPSK and the channel is 2×2 MIMO. The SF is 8 corresponding to an information rate of 480 kbps (rate-1/2 coding assumed).

gain is limited to the antenna gain and is therefore dictated by the number of receive antennas. Note that multiple transmit

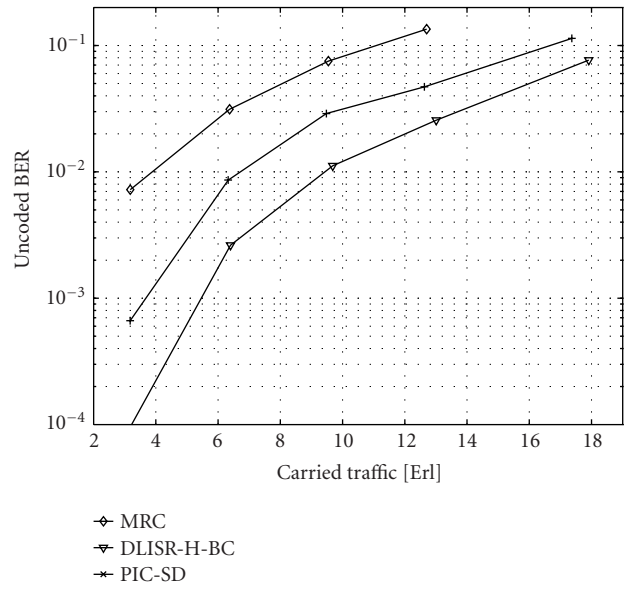


Figure 13: Uncoded BER performance as a function of the offered traffic. The modulation is QPSK, the channel is 4×4 MIMO, and the SBR is 20%. The SF is 8 corresponding to an information rate of 480 kbps (rate-1/2 coding assumed).

antennas still serve to alleviate the shortage of OVSF codes and can provide additional time diversity.

5.6. 2×2 MIMO with 16-QAM modulation

We use the same settings as in Section 5.4 but consider now 16-QAM modulation corresponding to a bit rate of 960 kbps after rate-1/2 coding. Figure 14a shows the uncoded BER as a function of the carried traffic. We have used the 16-QAM symbol constellation suggested in [4].

The capacity gain of DLISR compared to MRC becomes dominant offering 8.1 dB capacity increase achieved with DLISR-H-BC. DLISR-R-SD performs slightly worse with 7.7 dB gain over MRC, but as usual, outperforming PIC-SD which only provides a gain of 6.7 dB. The remarkable gains over MRC are a result of increased data rate which effectively exacerbates the near-far situations because interference is limited to fewer sources. Compared to the QPSK results, the carried Erlang traffic is reduced by about 5.4 dB for DLISR variants. The spectral efficiency, which decreases less due to the doubled symbol rate, is 1.1 bps/Hz for DLISR-H-BC corresponding to a reduction of 2.6 dB compared to MIMO QPSK.

Higher capacities can always be achieved at the expense of increased SBR because it implies higher SINR operating point, even though the carried traffic is constant. To see the effect, Figure 14b shows performance with SBR = 60%. The spectral efficiency is increased for all modes. For instance, the DLISR-H-BC spectral efficiency is increased by 1.8 dB, yielding an absolute spectral efficiency of about 1.5 bps/Hz. This illustrates the important trade-off between capacity and

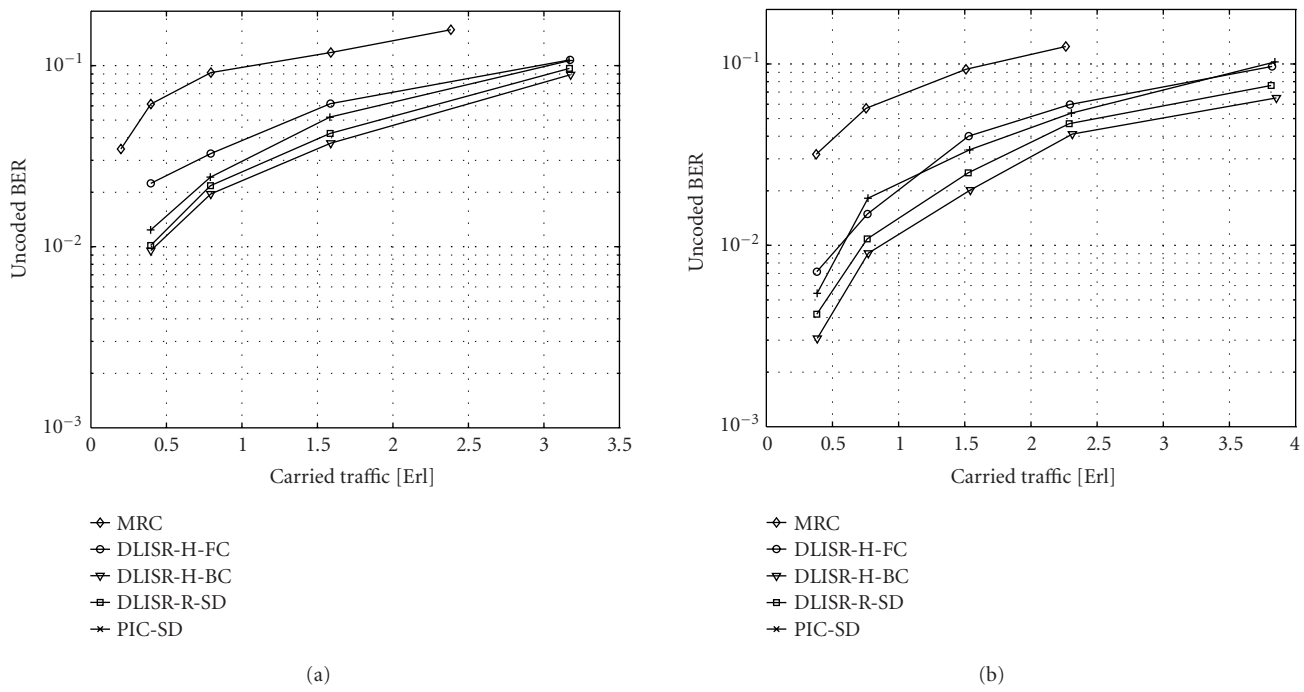


Figure 14: Uncoded BER performance as a function of the offered traffic. The modulation is 16 QAM, the channel is 2×2 MIMO, and the SF is 8 corresponding to an information rate of 960 kbps (rate-1/2 coding assumed). (a) The SBR is 20%. (b) The SBR is 60%.

network SBR. Higher SBR reduces the benefit of DLISR compared to MRC slightly, but it is still a significant 6.5 dB with DLISR-H-BC. MRC benefits more from increased SBR because in-cell interference becomes dominant and therefore an orthogonality gain, which is more pronounced for MRC, is achieved.

6. CONCLUSION

In this paper, we have presented a new MUD for DL MIMO systems. Our solution is based on previously presented ISR and is denoted DLISR. We have defined three variants of DLISR with different performances and complexities. The DLISR variants share one common feature, they employ VIR and they can benefit from dynamic allocation of channelization codes at the base station using the DACCA technique. VIR significantly reduces complexity because interference is rejected at a low (virtual) SF. With DACCA, the base station assigns channelization codes with the aim of concentrating interference in a small portion of the OVSF code tree. With DACCA, VIR therefore becomes even more efficient because we can attack interference at a lower SF hence reducing complexity further. We note that only one of our solutions requires DACCA. The remaining solutions benefit from DACCA in terms of complexity.

Performance of DLISR has been evaluated with the aid of a realistic simulation model consisting of an RNS and a link-level simulator. The RNS generates interference scenarios similar to those experienced in real life. These realizations of interference are used by the link-level simulator to produce BER performance statistics. At both levels, we have strived to use realistic assumptions. As a benchmark, we have considered the MRC-based SUD and the PIC-SD.

The Erlang capacity of the network is found to grow linearly with the number of receive antennas for both MRC-based SUD and our new DLISR MUD despite the existence of interference. Significant increases of capacity are achieved with DLISR which offers capacity gains over MRC-based SUD of at least 3 dB for QPSK (480 kbit/s) and about 6.5–8.1 dB when 16 QAM (960 kbit/s) is employed. A 4×4 MIMO system can support 17 Erlangs of 480 kbit/s traffic per sector corresponding to a spectral efficiency of 4 bits/s/Hz. DLISR-H-BC always achieves best performance outperforming DLISR-R-SD by about 0.3–0.5 dB. DLISR-R-SD outperforms PIC-SD by 0.5–0.9 dB, hence illustrating the advantage of linearly constrained beamforming (DLISR-R-SD) compared to subtraction (PIC-SD). DLISR-H-FC generally achieves the least gain over MRC, but it also possesses the simplest structure.

Our DLISR solutions have low complexity when DACCA is employed in UTRAN. The gains cited herein are achieved at a complexity of about 1.6 Gops, which is only about 4 times that of SUD, and close to the complexity of PIC-SD. The realistic assumptions of our study suggest that our solution is low risk. The new DLISR MUD is therefore a serious candidate for DL MUD in CDMA-based MIMO and SISO systems.

ACKNOWLEDGMENT

This work was supported by the Bell/Nortel/NSERC Industrial Research Chair in Personal Communications and the Natural Sciences and Engineering Research Council of Canada (NSERC) Research Grants Program.

REFERENCES

[1] F. Adachi, M. Sawahashi, and H. Suda, "Wideband DS-CDMA for next-generation mobile communication systems," *IEEE Communications Magazine*, vol. 36, no. 9, pp. 56–69, 1998.

[2] R. Prasad and T. Ojanper, "An overview of CDMA evolution toward wideband CDMA," *IEEE Communications Surveys*, vol. 1, no. 1, pp. 2–29, 1998.

[3] Third Generation Partnership Project, "Ultra high speed downlink packet access (HSDPA) (Release 5)," 3GPP Tech. Spec. (TS) 25.308, v. 5.2.0, Technical Specification Group Radio Access Network, March 2002.

[4] Third Generation Partnership Project, "Physical layer aspects of UTRA high speed downlink packet access (Release 4)," 3GPP Tech. Rep. (TR) 25.848, v. 4.0.0, Technical Specification Group Radio Access Network, March 2001.

[5] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[6] S. Verdu, *Multiuser Detection*, Cambridge University Press, Cambridge, UK, 1998.

[7] A. Duel-Hallen, J. Holtzman, and Z. Zvonar, "Multiuser detection for CDMA systems," *IEEE Personal Communications*, vol. 2, no. 2, pp. 46–58, 1995.

[8] S. Moshavi, "Multi-user detection for DS-CDMA communications," *IEEE Communications Magazine*, vol. 34, no. 10, pp. 124–136, 1996.

[9] S. Catreux, P. F. Driessen, and L. J. Greenstein, "Attainable throughput of an interference-limited multiple-input multiple-output (MIMO) cellular system," *IEEE Trans. Communications*, vol. 49, no. 8, pp. 1307–1311, 2001.

[10] S. Catreux, P. F. Driessen, and L. J. Greenstein, "Simulation results for an interference-limited multiple-input multiple-output cellular system," *IEEE Communications Letters*, vol. 4, no. 11, pp. 334–336, 2000.

[11] R. Lupas and S. Verdu, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 123–136, 1989.

[12] R. Lupas and S. Verdu, "Near-far resistance of multiuser detectors in asynchronous channels," *IEEE Trans. Communications*, vol. 38, no. 4, pp. 496–508, 1990.

[13] K. S. Schneider, "Optimum detection of code division multiplexed signals," *IEEE Trans. on Aerospace and Electronics Systems*, vol. 15, pp. 181–185, 1979.

[14] Z. Xie, R. T. Short, and C. K. Rushforth, "A family of suboptimum detectors for coherent multiuser communications," *IEEE Journal on Selected Areas in Communications*, vol. 8, no. 4, pp. 683–690, 1990.

[15] U. Madhow and M. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Communications*, vol. 42, no. 12, pp. 3178–3188, 1994.

[16] R. Kohno, H. Imai, M. Hatori, and S. Pasupathy, "Combination of an adaptive array antenna and a canceller of interference for direct-sequence spread-spectrum multiple-access system," *IEEE Journal on Selected Areas in Communications*, vol. 8, no. 4, pp. 675–682, 1990.

[17] M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code-division multiple-access communications," *IEEE Trans. Communications*, vol. 38, no. 4, pp. 509–519, 1990.

[18] P. Patel and J. Holtzman, "Analysis of a simple successive interference cancellation scheme in a DS/CDMA system," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 5, pp. 796–807, 1994.

[19] C. S. Wijting, T. Ojanperä, M. J. Juntti, K. Kansanen, and R. Prasad, "Groupwise serial multiuser detectors for multirate DS-CDMA," in *IEEE 49th Vehicular Technology Conference*, vol. 1, pp. 836–840, Houston, Tex, USA, July 1999.

[20] D. Guo, L. K. Rasmussen, S. Sun, and T. J. Lim, "A matrix-algebraic approach to linear parallel interference cancellation in CDMA," *IEEE Trans. Communications*, vol. 48, no. 1, pp. 152–161, 2000.

[21] L. K. Rasmussen, T. J. Lim, and A. L. Johansson, "A matrix-algebraic approach to successive interference cancellation in CDMA," *IEEE Trans. Communications*, vol. 48, no. 1, pp. 145–151, 2000.

[22] S. Affes and P. Mermelstein, "A new receiver structure for asynchronous CDMA: STAR—the spatio-temporal array-receiver," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1411–1422, 1998.

[23] H. Hansen, S. Affes, and P. Mermelstein, "A beamformer for CDMA with enhanced near-far resistance," in *Proc. IEEE International Conference on Communications*, vol. 3, pp. 1583–1587, Vancouver, British Columbia, Canada, June 1999.

[24] A. J. Viterbi, *CDMA Principles of Spread Spectrum Communication*, vol. 1 of *Addison-Wesley Wireless Communication Series*, Addison-Wesley, 1997.

[25] G. E. Bottomley, T. Ottosson, and Y.-P. E. Wang, "A generalized RAKE receiver for interference suppression," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 8, pp. 1536–1545, 2000.

[26] G. Kutz and A. Chass, "On the performance of a practical downlink CDMA generalized RAKE receiver," in *IEEE 56th Vehicular Technology Conference*, vol. 3, pp. 1352–1356, Vancouver, British Columbia, Canada, September 2002.

[27] G. Kutz and A. Chass, "Low complexity implementation of a downlink CDMA generalized RAKE receiver," in *IEEE 56th Vehicular Technology Conference*, vol. 3, pp. 1357–1361, Vancouver, British Columbia, Canada, September 2002.

[28] M. F. Madkour, S.C. Gupta, and Y.-P. E. Wang, "Successive interference cancellation algorithms for downlink W-CDMA communications," *IEEE Trans. on Wireless Communications*, vol. 1, no. 1, pp. 169–177, 2002.

[29] Third Generation Partnership Project, "Spreading and modulation (FDD) (Release 5)," 3GPP Tech. Spec. (TS) 25.213, v. 5.2.0, Technical Specification Group Radio Access Network, September 2002.

[30] S. Affes, H. Hansen, and P. Mermelstein, "Interference subspace rejection: a framework for multiuser detection in wideband CDMA," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 2, pp. 287–302, 2002.

[31] S. Affes, K. Lajnef, K. Cheikhrouhou, and P. Mermelstein, "Adaptive MIMO-diversity selection with closed-loop power control over wireless CDMA rayleigh-fading channels," in *Proc. 7th IEEE International Symposium on Signal Processing and Its Applications*, vol. 1, pp. 41–44, Paris, France, July 2003.

[32] W. C. Jakes Jr., *Microwave Mobile Communications*, John Wiley & Sons, New York, NY, USA, 1974.

[33] T. Minn and K.-Y. Siu, "Dynamic assignment of orthogonal variable-spreading-factor codes in W-CDMA," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 8, pp. 1429–1440, 2000.

[34] Y.-C. Tseng, C.-M. Chao, and S.-L. Wu, "Code placement and replacement strategies for wideband CDMA OVSF code tree management," in *Proc. IEEE Global Telecommunications Conference*, vol. 1, pp. 562–566, San Antonio, Tex, USA, November 2001.

[35] H. Hansen, S. Affes, and P. Mermelstein, "Downlink MIMO multiuser detection with interference subspace rejection," in *Proc. IEEE Global Telecommunications Conference*, San Francisco, Calif, USA, December 2003.

[36] S. Affes, H. Hansen, and P. Mermelstein, "Near-far resistant single-user channel identification by interference subspace rejection in wideband CDMA," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, pp. 54–57, Taoyuan, Taiwan, March 2001.

[37] K. Cheikhrouhou, S. Affes, and P. Mermelstein, "Impact of synchronization on performance of enhanced array-receivers in wideband CDMA networks," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 12, pp. 2462–2476, 2001.

[38] Y. Okumura, E. Ohmori, T. Kawano, and K. Fukuda, "Field strength and its variability in VHF and UHF land-mobile radio service," *Review of the Electrical Communication Laboratory*, vol. 16, no. 9-10, pp. 825–873, 1968.

[39] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Trans. Vehicular Technology*, vol. VT-29, no. 3, pp. 317–325, 1980.

[40] B. Smida, S. Affes, and P. Mermelstein, "Frequency and time synchronization for the CDMA array-receiver STAR with interference subspace rejection," in *IEEE Vehicular Technology Conference*, vol. 6, pp. 3290–3295, Orlando, Fla, USA, October 2003.

Henrik Hansen was born in Copenhagen, Denmark, on April 19, 1971. In 1997, he received the M.S. degree with communication as specific field of engineering and in March 2001, he received the Ph.D. degree in electrical engineering, both from the Technical University of Denmark, Denmark. During his Ph.D. study, he was a Visiting Researcher at INRS-Telecommunications, Montréal, Quebec, Canada. The topic of the Ph.D. study was digital demodulation principles for CDMA systems, with special focus on multiuser detection. Since January 2001, he has been employed as a Radio-Network Planner at Ericsson, Denmark. He is primarily involved in the planning of WCDMA radio networks with focus on the dimensioning of the air interface. He is also active in various research projects within wireless technologies. His research interests include digital modulation/demodulation principles, multiuser detection, interference cancellation, adaptive antennas, space-time processing, coding, and radio-network planning and optimization.

Sofiène Affes received the "Diplôme d'Ingénieur" in electrical engineering in 1992, and the Ph.D. degree with honors in signal processing in 1995, both from the École Nationale Supérieure des Télécommunications, Paris, France. He has been since then with INRS-Telecommunications, University of Quebec, Montreal, Canada, as a Research Associate from 1995 till 1997, then as an Assistant Professor till 2000. Currently, he is an Associate Professor in the Personal Communications Group. His research interests are in wireless communications, statistical signal and array processing, synchronization, multiuser detection, and MIMO systems. In 1997, he participated in the major program in personal and mobile communications of the Canadian Institute for Telecommunications Research. From 1998 to 2001, he has been leading the radio-design and signal processing activities of the Bell/Nortel/NSERC Industrial Research Chair in Personal Communications at INRS-Telecommunications. Professor Affes is the Corecipient of the 2002 Prize for Research Excellence of INRS and currently holds a Canada Research Chair in High-Speed Wireless Communications.

Paul Mermelstein received the B. Eng. degree in engineering physics from McGill University, Montreal, Canada, in 1959 and the S.M., E.E., and D.S. degrees in electrical engineering from Massachusetts Institute of Technology, Cambridge, Mass, in 1960, 1963, and 1964, respectively. From 1964 to 1973, he was a member of the technical staff in the Speech and Communications Research Department, Bell Laboratories, Murray Hill, NJ, carrying out research studies in speech generation and recognition. From 1973 to 1977, he was with Haskins Laboratories, New Haven, Conn, conducting research in speech perception and recognition. He moved to Bell Northern Research, later Nortel Networks, in 1977 to lead a speech communications research group. In 1994, he moved to INRS-Telecommunications, University of Quebec, as Professor and the Bell/Nortel/NSERC Industrial Research Chair in Personal Communications. From 1994 to 2000, he was the leader of the major program in personal and mobile communications of the Canadian Institute for Telecommunications Research. He is a Fellow of the IEEE and the Acoustical Society of America and served as Editor for Speech Communications of the IEEE Transactions on Communications.

EURASIP Journal on Applied Signal Processing 2004:5, 727–739

Maximum Likelihood Turbo Iterative Channel Estimation for Space-Time Coded Systems and Its Application to Radio Transmission in Subway Tunnels

Miguel González-López
Departamento de Electrónica y Sistemas, Universidade da Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain
Email: miguelgl@udc.es

Joaquín Míguez
Departamento de Electrónica y Sistemas, Universidade da Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain
Email: jmiguez@udc.es

Luis Castedo
Departamento de Electrónica y Sistemas, Universidade da Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain
Email: luis@udc.es

Received 31 December 2002; Revised 31 July 2003

This paper presents a novel channel estimation technique for space-time coded (STC) systems. It is based on applying the maximum likelihood (ML) principle not only over a known pilot sequence but also over the unknown symbols in a data frame. The resulting channel estimator gathers both the deterministic information corresponding to the pilot sequence and the statistical information, in terms of a posteriori probabilities, about the unknown symbols. The method is suitable for Turbo equalization schemes where those probabilities are computed with more and more precision at each iteration. Since the ML channel estimation problem does not have a closed-form solution, we employ the expectation-maximization (EM) algorithm in order to iteratively compute the ML estimate. The proposed channel estimator is first derived for a general time-dispersive MIMO channel and then is particularized to a realistic scenario consisting of a transmission system based on the global system mobile (GSM) standard performing in a subway tunnel. In this latter case, the channel is nondispersive but there exists controlled ISI introduced by the Gaussian minimum shift keying (GMSK) modulation format used in GSM. We demonstrate, using experimentally measured channels, that the training sequence length can be reduced from 26 bits as in the GSM standard to only 5 bits, thus achieving a 14% improvement in system throughput.

Keywords and phrases: STC, turbo equalization, turbo channel estimation, maximum likelihood channel estimation, GSM, subway tunnels.

1. INTRODUCTION

Recently, the so-called Turbo codes [1, 2, 3] have revealed themselves as a very powerful coding technique able to approach the Shannon limit in AWGN channels. A Turbo code is made up of two component codes (block or convolutional) parallely or serially concatenated via an interleaver. This simple coding scheme produces very long codewords, so each source information bit is highly spread through the transmitted coded sequence. At reception, optimum maximum likelihood (ML) decoding can be carried out by considering the hypertrellis associated with the concatenation of the two component codes. Obviously, such a decoding approach becomes impractical in most situations. The key idea behind Turbo coding is to overcome this problem by employing a suboptimal, but very powerful, decoding scheme termed *iterative maximum a posteriori* (MAP) decoding [3, 4]. Basically, the method relies on independently decoding each of the component codes and exchanging in an iterative fashion the statistical information, that is, the a posteriori probabilities about symbols, obtained in each decoding module.

The same decoding principle has also been successfully applied, under the term *Turbo equalization* [5], to effectively compensate the ISI induced by the channel and/or the

modulation scheme. This technique exploits the fact that ISI can be viewed as a form of rate-1, nonrecursive coding. So, whatever coding scheme is used, if an interleaver is located prior to the channel, the overall effect of coding and ISI can be treated as a concatenated code and therefore, iterative MAP decoding can be applied. Luschi et al. [6] present an in-depth review of this technique and further improvements can be found in [7, 8, 9, 10]. In general, iterative MAP processing can be applied to a variety of situations where the overall system can be viewed as a concatenation of modules whose input/output relationship can be described as a (hidden) Markov chain. Several works have appeared in the last years exploiting this idea. For instance, Görtz [11], Garcia-Frias and Villasenor [12], and Guyader et al. [13] worked on the problem of joint source-channel decoding and Zhang and Burr [14] addressed the problem of symbol timing recovery.

In practical receivers, where the channel impulse response has to be estimated, it is convenient to have channel estimators capable of benefiting from the high performance of Turbo equalizers [15, 16, 17]. Moreover, second- and third-generation mobile standards consider the transmission of pilot sequences known by the receiver for channel estimation purposes. In the *global system mobile* (GSM) standard, this sequence is 26 bits long, which represents 17.6% of the total frame length (148 bits) [18]. Such a long training sequence is necessary if classical estimation techniques, such as least squares (LS), are used. Employing more refined channel estimators, such as the one presented in this paper, we can dramatically decrease the necessary length of the training sequence and therefore increase the overall system throughput. In [19], an ML-based channel estimator is presented where the ML principle is applied not only to the pilot sequence, but also to the whole data frame. Since the involved optimization problem had no analytical solution, the expectation-maximization (EM) algorithm [20] was used for iteratively obtaining the solution.

Also, wireless communications research has been very influenced by the discovery of the potentials of communicating through multiple-input multiple-output (MIMO) channels, which can be carried out using antenna diversity not only at reception, as classical space-diversity techniques have been doing, but also at transmission. MIMO techniques have the advantage to provide high data rate wireless services at no extra bandwidth expansion or power consumption. Telatar [21] calculated the capacity associated with a MIMO channel that in certain cases grows linearly with the number of antennas [22]. More recent progress in information theoretical properties of multiantenna channel can be found in [23].

Although MIMO channel capacity can be really high, it can only be successfully exploited by proper coding and modulation schemes. The term space-time Coding (STC) [24, 25] has been adopted for such techniques. Special efforts have been made in code design [24, 26] and several decoding approaches have been developed for these codes. In both fields, the Turbo principle has been applied in profusion. Turbo ST codes designs can be found in [27, 28, 29] and various Turbo decoding schemes are exposed in [30, 31].

As in single-antenna systems, practical ST receivers must perform the operation of channel estimation. Having efficient and robust estimators is crucial to guarantee that the system performance degradation due to the channel estimation error is minimized. In this paper, we present a novel channel estimation technique that gathers both the deterministic information corresponding to the pilot sequence and the statistical information, in terms of a posteriori probabilities, about the unknown symbols. The method is suitable for Turbo equalization schemes where those probabilities are computed with more and more precision at each iteration. We derive the channel estimator for general MIMO time-dispersive channels and analyze its performance in a multiple-antenna communication system based on the GSM standard operating inside subway tunnels.

The main motivation for developing a multiple-antenna GSM-based communication system is the following. GSM is, by far, the most widely deployed radio-communication system. Since 1993, its radio interface (GSM-R) has been adopted by the European railway digital radio-communication systems. Due to the conservative nature of its market, it is expected that railway radio-communication systems will employ GSM-R for the long-term future. For this reason, when subway operators wish to deploy advanced, high data rate, digital services for security or entertainment purposes, it is very likely that they will prefer to increase the capacity of the existing GSM-R system rather than switch to another radio standard. STC and Turbo equalization are very promising ways of achieving this capacity growth [32]. In this specific application, we will show that the proposed iterative MLMIMO channel estimation method has large benefits over traditional channel estimation approaches.

The rest of the paper is organized as follows. Section 2 presents the signal model and Section 3 describes the Turbo equalization scheme for STC systems. Next, in Section 4, we derive the ML channel estimator for a general time-dispersive MIMO channel. Since direct application of the ML principle leads to an optimization problem without closed-form solution, the EM algorithm is applied for computing the actual value of the solution, resulting in the so-called ML-EM estimator. The application of the proposed channel estimator to a STC GSM-based system operating in subway tunnels is detailed in Section 5. Section 6 presents the results of computer experiments for both the general case and experimental measurements of subway tunnel MIMO channels. Finally, Section 7 is devoted to the conclusions.

2. SIGNAL MODEL

We consider the transmitter signal model corresponding to an STC system shown in Figure 1. The original bit sequence $u(k)$ feeds an ST encoder whose output is a sequence of vectors $\mathbf{c}(k) = [c_1(k)\ c_2(k)\ \cdots\ c_N(k)]^T$, with N being the number of transmitting antennas. The specific spatio-temporal structure of the sequence of vectors $\mathbf{c}(k)$ depends on the particular STC technique employed. Any of the several STC methods that have been proposed in the literature could be used in our scheme. However, we have focused on ST

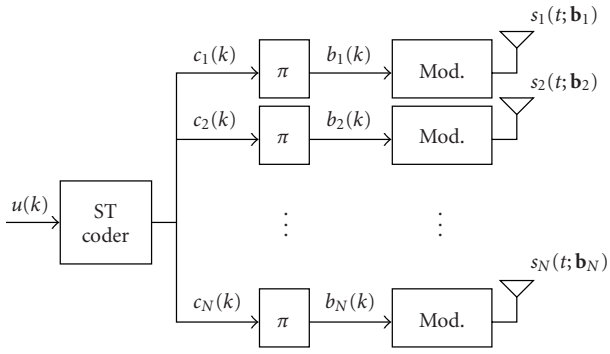


Figure 1: Transmitter model.

trellis codes [24, 25] to elaborate our simulation results. Each component of $\mathbf{c}(k)$ is independently interleaved to produce a new symbol vector $\mathbf{b}(k) = [b_1(k)\ \ b_2(k)\ \ \cdots\ \ b_N(k)]^T$ and these are the symbols that are afterwards modulated (waveform encoded) to yield the signals $s_i(t;\mathbf{b}_i)\ i = 1, 2, \ldots, N$ that will be transmitted along the radio channel. Without loss of generality, we will assume that the modulation format is linear and that the channel suffers from time-dispersive multipath fading with memory length m. It is well known that at reception, matched-filtering and symbol-rate sampling can be used to obtain a set of sufficient statistics for the detection of the transmitted symbols. Using vector notation, this set of statistics will be grouped in vectors $\mathbf{x}(k) = [x_1(k)\ \ x_2(k)\ \ \cdots\ \ x_L(k)]^T$, $k = 0, 1, \ldots, K-1$, where L is the number of receiving antennas and K is the number of total transmitted symbol vectors in a data frame. Elaborating the signal model, it can be easily shown that the sufficient statistics $\mathbf{x}(k)$ can be expressed as

$$\mathbf{x}(k) = \mathbf{H}\mathbf{z}(k) + \mathbf{v}(k), \tag{1}$$

where matrix $\mathbf{H} = [H(m-1)\ \ H(m-2)\ \ \cdots\ \ H(0)]$ represents the overall dispersive MIMO channel with memory length m. Each submatrix

$$H(i) = \begin{bmatrix} h_{11}(i) & h_{12}(i) & \cdots & h_{1N}(i) \\ h_{21}(i) & h_{22}(i) & \cdots & h_{2N}(i) \\ \vdots & \vdots & \ddots & \vdots \\ h_{L1}(i) & h_{L2}(i) & \cdots & h_{LN}(i) \end{bmatrix} \tag{2}$$

contains the fading coefficients that affect the symbol vector $\mathbf{b}(k-i)$. Vector $\mathbf{z}(k)$ results from stacking the source vectors $\mathbf{b}(k)$, that is,

$$\mathbf{z}(k) = [\mathbf{b}^T(k-m+1)\ \ \mathbf{b}^T(k-m+2)\ \ \cdots\ \ \mathbf{b}^T(k)^T]. \tag{3}$$

Finally, the noise component $\mathbf{v}(k)$ is a vector of mutually independent complex-valued, circularly symmetric Gaussian random processes, that is, the real and imaginary parts are zero-mean, mutually independent Gaussian random processes having the same variance. We will also assume that the noise is temporally white with variance σ_v^2.

3. ST TURBO DETECTION

Figure 2 shows the block diagram of an ST Turbo detector. The MAP equalizer [4] computes $L[\mathbf{b}(k)|\tilde{\mathbf{x}}]$ which are the *a posteriori* log-probabilities of the input symbols $\mathbf{b}(k)$ based on the available observations $\tilde{\mathbf{x}} = [\mathbf{x}^T(0)\ \ \mathbf{x}^T(1)\ \ \cdots\ \ \mathbf{x}(K-1)]^T$. Due to its time-dispersive nature, it is convenient to represent our MIMO channel by means of a finite-state machine (FSM) having $2^{N(m-1)}$ states. This FSM has 2^N transitions per state which implies that there is a total number of 2^{Nm} transitions between two time instants. Let $e_k = (s_{k-1}, \mathbf{b}(k), \mathbf{s}(k), s_k)$ be one of the 2^{Nm} possible transitions at time k of this FSM. This transition depends on four parameters: the incoming state s_{k-1}, the outgoing state s_k, the input symbol vector $\mathbf{b}(k)$, and the output symbol vector without noise $\mathbf{s}(k) = \mathbf{H}\mathbf{z}(k)$. It is important to point out that the incoming state is determined by the $m-1$ previous symbol vectors, that is, $s_{k-1} = (\mathbf{b}(k-m+1), \mathbf{b}(k-m+2), \ldots, \mathbf{b}(k-1))$. On the other hand, the outgoing state is a function of the previous state and the current input symbols, that is, $s_k = f_{\text{next}}(s_{k-1}, \mathbf{b}(k))$. For a better description of the MAP equalizer, we are going to introduce the notation $\mathbf{b}(k) = L_{\text{in}}(e_k)$ and $\mathbf{s}(k) = L_{\text{out}}(e_k)$ to represent the input and output symbol vectors associated to the transition e_k, respectively. Note that the output vector does not depend on the outgoing state s_k, so we will slightly change our notation and write

$$\begin{aligned} \mathbf{s}(k) &= L_{\text{out}}(e_k) = L_{\text{out}}(s_{k-1}, \mathbf{b}(k)) \\ &= L_{\text{out}}(\mathbf{z}(k)) = \mathbf{H}\mathbf{z}(k). \end{aligned} \tag{4}$$

The *a posteriori* log-probabilities $L[\mathbf{b}(k)|\tilde{\mathbf{x}}]$ can be recursively computed by means of the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm [3, 4] which is summarized in the sequel. The first stage when computing the *a posteriori* log-probabilities is noting that

$$L[\mathbf{b}(k)|\tilde{\mathbf{x}}] = L[\mathbf{b}(k), \tilde{\mathbf{x}}] + h_b, \tag{5}$$

where h_b is the constant that makes $P[\mathbf{b}(k)|\tilde{\mathbf{x}}]$ a probability mass function and

$$L[\mathbf{b}(k), \tilde{\mathbf{x}}] = \log \sum_{e_k : L_{\text{in}}(e_k) = \mathbf{b}(k)} \exp L[e_k, \tilde{\mathbf{x}}] \tag{6}$$

is the joint log-probability of the transition e_k and the set of available observations $\tilde{\mathbf{x}}$. This joint log-probability can be expressed as

$$L[e_k, \tilde{\mathbf{x}}] = \alpha_{k-1}[s_{k-1}] + \gamma_k[e_k] + \beta_k[s_k], \tag{7}$$

where

$$\begin{aligned} \alpha_k[s] &= L[s_{k-1}, \tilde{\mathbf{x}}_k^-], \\ \gamma_k[e_k] &= L[\mathbf{b}(k)] + L[\mathbf{x}(k)|\mathbf{s}(k)], \\ \beta_k[s] &= L[\tilde{\mathbf{x}}_k^+|s_k], \end{aligned} \tag{8}$$

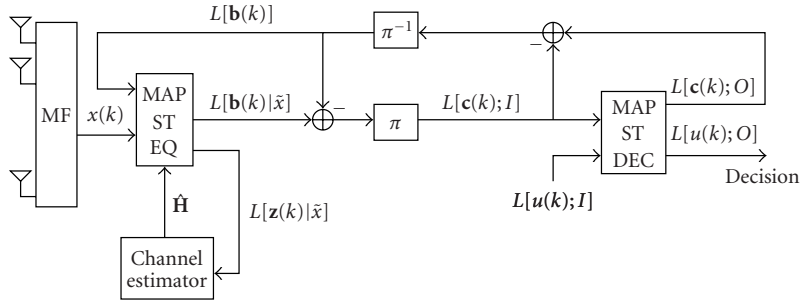


Figure 2: Receiver model.

with

$$L[\mathbf{x}(k)|\mathbf{s}(k)] = -\frac{1}{\sigma_v^2}||\mathbf{x}(k) - \mathbf{H}\mathbf{z}(k)||^2, \tag{9}$$

$$\tilde{\mathbf{x}}_k^- = [\mathbf{x}^T(0) \;\; \mathbf{x}^T(1) \;\; \cdots \;\; \mathbf{x}^T(k-1)], \tag{10}$$

$$\tilde{\mathbf{x}}_k^+ = [\mathbf{x}^T(k+1) \;\; \mathbf{x}^T(k+2) \;\; \cdots \;\; \mathbf{x}^T(K-1)]. \tag{11}$$

Note that the noise variance σ_v^2 is needed in (9). Our simulation results assume this parameter as known. However, it could be estimated and, in particular, it can be considered as another parameter to be estimated by the ML estimator described in Section 4, as shown in [33], for the case of a decision feedback-equalizer (DFE) instead of a MAP detector. The computation of the quantities $\alpha_k[s]$, $\gamma_k[e_k]$, and $\beta_k[s]$ can be carried out recursively by first performing a forward recursion

$$\alpha_{k-1}[s_{k-1}] = \log \sum_{\substack{\mathbf{b}(k),s_{k-2}:\\ f_{\text{next}}(s_{k-2},\mathbf{b}(k-1))=s_{k-1}}} \exp\Big\{\alpha_{k-2}[s_{k-2}] + L[\mathbf{b}(k-1)] + L[\mathbf{x}(k)|\mathbf{s}(k)]\Big\} \tag{12}$$

with initial values $\alpha_0[s=0] = 0$ and $\alpha_0[s \neq 0] = -\infty$, and then proceeding with a backward recursion

$$\beta_k[s_k] = \log \sum_{\substack{\mathbf{b}(k+1),s_{k+1}:\\ f_{\text{next}}(s_k,\mathbf{b}(n+1))=s_{k+1}}} \exp\Big\{\beta_{k+1}[s_{k+1}] + L[\mathbf{b}(k+1)] + L[\mathbf{x}(k+1)|\mathbf{s}(k+1)]\Big\} \tag{13}$$

using as initial values $\beta_{K-1}[s = s_{K-1}] = 0$ and $\beta_{K-1}[s \neq s_{K-1}] = -\infty$.

Similarly, the decoder has to compute the *a posteriori* log-probabilities of the original symbols $L[u(k);O]$ from their a priori log-probabilities $L[u(k);I] = \log(0.5)$ and the a priori log-probabilities $L[\mathbf{c}(k);I]$ which come from the detector. Again, the BCJR algorithm applies [3, 4]. It also computes the *a posteriori* log-probabilities of the transmitted symbols $L[\mathbf{c}(k);O]$ using

$$L[\mathbf{c}(k);O] = \log \sum_{e_k:L_{\text{out}}(e_k)=\mathbf{c}(k)} \exp\{\alpha_{k-1}[s_{k-1}] + \gamma_k[s_k] + \beta_k[s_k]\}, \tag{14}$$

where $L[\mathbf{c}(k);I]$ is utilized as branch metric. These computed log-probabilities are then fed back to the detector to act as the *a priori* log-probabilities $L[\mathbf{b}(k)]$. As reflected in Figure 2, note that it is always necessary to subtract the *a priori* component from the computed log-probabilities before forwarding them to the other module in order to avoid statistical dependence with the results of the previous iteration.

4. MAXIMUM LIKELIHOOD CHANNEL ESTIMATION

Channel estimation is often mandatory when practically implementing ST detection strategies, unless we deal with some kind of blind processing techniques. In this section, we will present a novel channel estimation method that will enable us to take full advantage from the Turbo detection scheme presented in the Section 3.

When developing our channel estimation approach, we will exploit the fact that transmitted data frames in most practical systems contain a deterministic known pilot sequence of length M for the purpose of estimating the channel at reception. For instance, in GSM, this sequence is $M = 26$ bits long [18]. Let $\tilde{\mathbf{b}}_f = [\tilde{\mathbf{b}}_t^T \;\; \tilde{\mathbf{b}}^T]^T$ denote the overall data frame, which includes $\tilde{\mathbf{b}}_t = [\mathbf{b}_t^T(0) \;\; \mathbf{b}_t^T(1) \;\; \cdots \;\; \mathbf{b}_t^T(M-1)]^T$ as the training sequence and $\tilde{\mathbf{b}} = [\mathbf{b}^T(M) \;\; \mathbf{b}^T(M+1) \;\; \cdots \;\; \mathbf{b}^T(K-1)]^T$ as the information sequence. Analogously, $\tilde{\mathbf{x}}_f = [\tilde{\mathbf{x}}_t^T \;\; \tilde{\mathbf{x}}^T]^T$ are the observations corresponding to one data frame, where $\tilde{\mathbf{x}}_t = [\mathbf{x}_t^T(0) \;\; \mathbf{x}_t^T(1) \;\; \cdots \;\; \mathbf{x}_t^T(M-1)]^T$ represents the pilot sequence and $\tilde{\mathbf{x}} = [\mathbf{x}(M) \;\; \mathbf{x}(M+1) \;\; \cdots \;\; \mathbf{x}(K-1)]^T$ corresponds to the information sequence. The ML estimator is thus given by

$$\hat{\mathbf{H}} = \arg\max_{\mathbf{H}} f_{\tilde{\mathbf{x}}|\tilde{\mathbf{b}}_t;\mathbf{H}}(\tilde{\mathbf{x}}), \tag{15}$$

where $f_{\tilde{\mathbf{x}}_t|\tilde{\mathbf{b}}_t;\mathbf{H}}$ is the probability density function (pdf) of the observations conditioned on the available information (the training sequence $\mathbf{b}_t$) and the parameters to be estimated

(the channel matrix $\mathbf{H}$). Although, this is a problem without closed-form solution, the EM algorithm [20] can be employed to iteratively solve (15). The EM algorithm relies on defining a so-called "complete data" set formed by the observable variables and by additional unobservable variables. At each iteration of the algorithm, a more refined estimate is computed by averaging the log-likelihood of the complete data set with respect to the pdf of the unobservable variables conditioned on the available set of observations. Using the EM terminology, we define the union of the observations (which are the observable variables) and the transmitted bit sequence (which are the unobservable variables) $\tilde{\mathbf{x}}_e = [\tilde{\mathbf{b}}_f^T \ \tilde{\mathbf{x}}_f^T]^T$ as the complete data set, whereas the observations $\tilde{\mathbf{x}}_f$ are the incomplete data set. The relationship between $\tilde{\mathbf{x}}_e$ and $\tilde{\mathbf{x}}_f$ must be given by a noninvertible linear transformation, that is, $\tilde{\mathbf{x}}_f = \mathbf{T}\tilde{\mathbf{x}}_e$. It can be easily seen that in our case, this transformation is given by $\mathbf{T} = [\mathbf{0}_{L(M+K)\times N(M+K)}\mathbf{I}_{L(M+K)}]$. With these definitions in mind, the estimate of the channel at the $i+1$th iteration is obtained by solving

$$\hat{\mathbf{H}}_{i+1} = \arg\max_{\mathbf{H}} E_{\tilde{\mathbf{x}}_e|\tilde{\mathbf{x}}_f,\tilde{\mathbf{b}}_t;\hat{\mathbf{H}}_i}\{\log f_{\tilde{\mathbf{x}}_e|\tilde{\mathbf{b}}_t;\mathbf{H}}(\tilde{\mathbf{x}}_e)\}, \tag{16}$$

where $E_f\{\cdot\}$ denotes the expectation operator with respect to the pdf $f(x)$. Expanding the previous expression, we have

$$\begin{aligned}
\hat{\mathbf{H}}_{i+1} &= \arg\max_{\mathbf{H}} E_{\tilde{\mathbf{b}}|\tilde{\mathbf{x}};\hat{\mathbf{H}}_i}\{\log[f_{\tilde{\mathbf{x}}_f|\tilde{\mathbf{b}}_f;\mathbf{H}}(\tilde{\mathbf{x}}_f) f_{\tilde{\mathbf{b}}}(\tilde{\mathbf{b}})]\} \\
&= \arg\max_{\mathbf{H}} E_{\tilde{\mathbf{b}}|\tilde{\mathbf{x}};\hat{\mathbf{H}}_i}\{\log[f_{\tilde{\mathbf{x}}_t|\tilde{\mathbf{b}}_t;\mathbf{H}}(\tilde{\mathbf{x}}_t) f_{\tilde{\mathbf{x}}|\tilde{\mathbf{b}};\mathbf{H}}(\tilde{\mathbf{x}})]\} \\
&= \arg\max_{\mathbf{H}} \log f_{\tilde{\mathbf{x}}_t|\tilde{\mathbf{b}}_t;\mathbf{H}}(\tilde{\mathbf{x}}_t) + E_{\tilde{\mathbf{b}}|\tilde{\mathbf{x}};\hat{\mathbf{H}}_i}\{\log f_{\tilde{\mathbf{x}}|\tilde{\mathbf{b}};\hat{\mathbf{H}}}(\tilde{\mathbf{x}})\} \\
&= \arg\min_{\mathbf{H}} \sum_{k=0}^{M-1} ||\mathbf{x}_t(k) - \mathbf{H}\mathbf{z}_t(k)||^2 \\
&\quad + E_{\tilde{\mathbf{b}}|\tilde{\mathbf{x}};\hat{\mathbf{H}}_i}\left\{\sum_{k=M}^{K-1} ||\mathbf{x}(k) - \mathbf{H}\mathbf{z}(k)||^2\right\},
\end{aligned} \tag{17}$$

where the last equality follows from the fact that, as far as we assume AWGN, the pdf of the observations conditioned on the transmitted symbols $f_{\tilde{\mathbf{x}}|\tilde{\mathbf{b}};\hat{\mathbf{H}}_i}$ is Gaussian. This leads to the following quadratic optimization problem:

$$\begin{aligned}
\hat{\mathbf{H}}_{i+1} = \arg\min_{\mathbf{H}} &\sum_{k=0}^{M-1} ||\mathbf{x}_t(k) - \mathbf{H}\mathbf{z}_t(k)||^2 \\
&+ \sum_{k=M}^{K-1} E_{\mathbf{z}(k)|\tilde{\mathbf{x}};\hat{\mathbf{H}}_i}\left\{||\mathbf{x}(k) - \mathbf{H}\mathbf{z}(k)||^2\right\}
\end{aligned} \tag{18}$$

with the closed-form solution[1]

$$\hat{\mathbf{H}}_{i+1} = (\mathbf{R}_{xz,t} + \mathbf{R}_{xz}) \times (\mathbf{R}_{z,t} + \mathbf{R}_z)^{-1}, \tag{19}$$

where

$$\mathbf{R}_{xz,t} = \sum_{k=0}^{M-1} \mathbf{x}_t(k)\mathbf{z}_t^H(k), \tag{20}$$

$$\mathbf{R}_{z,t} = \sum_{k=0}^{M-1} \mathbf{z}_t(k)\mathbf{z}_t^H(k), \tag{21}$$

$$\mathbf{R}_{xz} = \sum_{k=M}^{K-1} E_{\mathbf{z}(k)|\tilde{\mathbf{x}};\hat{\mathbf{H}}_i}\{\mathbf{x}(k)\mathbf{z}^H(k)\}, \tag{22}$$

$$\mathbf{R}_z = \sum_{k=M}^{K-1} E_{\mathbf{z}(k)|\tilde{\mathbf{x}};\hat{\mathbf{H}}_i}\{\mathbf{z}(k)\mathbf{z}^H(k)\}. \tag{23}$$

Note that for computing (22) and (23), it is necessary to know the probability mass function $p_{\mathbf{z}(k)|\tilde{\mathbf{x}};\hat{\mathbf{H}}_i}$. Towards this aim, we take benefit from the Turbo equalization process because

$$L[\mathbf{z}(k)|\tilde{\mathbf{x}};\hat{\mathbf{H}}_i] = L[\mathbf{z}(k),\tilde{\mathbf{x}};\hat{\mathbf{H}}_i] + h_z = L[e_k,\tilde{\mathbf{x}}] + h_z, \tag{24}$$

where h_z is the constant that makes $p_{\mathbf{z}(k)|\tilde{\mathbf{x}};\hat{\mathbf{H}}_i}$ a probability mass function and $L[e_k,\tilde{\mathbf{x}}]$ is the joint log-probability of the transition e_k and the set of available observations. Notice that this quantity has already been computed in the Turbo equalization process (see (7)). This fact makes the proposed channel estimator very suitable to be used within a Turbo equalization structure.

5. APPLICATION TO AN STC SYSTEM FOR SUBWAY ENVIRONMENTS

We focus now on the application of the ML-EM channel estimator described in Section 4 to an STC GSM-like system for underground railway transportation systems. Some practical considerations follow. In subway tunnel environments, propagation conditions result in flat multipath fading because its delay spread is small when compared to the GSM symbol period [35]. Nevertheless, the modulation employed by the GSM standard, Gaussian minimum shift keying (GMSK), induces controlled ISI and thus Turbo ST Equalization can be employed for the purpose of joint demodulating and decoding. In addition, experimental measurements [36] have revealed that in this environment, there exist strong spatial correlations between subchannels. These spatial correlations will be taken into account when evaluating the receivers' performance in the following section because we will use, in the computer simulations, experimental measurements of MIMO channel impulse responses obtained in subway tunnels. These field measurements have been carried out in the framework of the European project "ESCORT" [37]. We will show how the proposed channel estimator allows to reduce the necessary length of the training sequence from 26 bits in the GSM standard up to only 5 bits, while performance is maintained very close to the optimum (i.e., the bit error rate (BER) obtained when the channel is perfectly known at reception) which clearly implies a very high gain in the overall system throughput.

[1] Since the expectation operator is linear, the derivation leading to (19) follows, step by step, the usual optimization procedure to find the LS estimate of a linear system given a set of noisy observations (see, e.g., [34]). Such a procedure includes the calculation of the gradient with respect to the system coefficients and then solving for the points where the gradient vanishes. Hence, solving (17) is tedious, since derivatives have to be computed for the coefficients in matrix $\mathbf{H}$, but conceptually straightforward.

Figure 1 can be useful again for modeling the STC transmitter under consideration (for the sake of clarity, we refer the reader to Appendix A for a detailed description). This model can be summarized as follows. Each component of $\mathbf{b}(k)$ is independently modulated using the GMSK modulation format. GMSK is a partial response continuous phase modulation (CMP) signal and thus a nonlinear modulation format. Nevertheless, it can be expressed in terms of its Laurent expansion [38, 39, 40] as the sum of 2^{p-1} PAM signals, where p is the memory induced by the modulation. For the GMSK format in the GSM standard, $p = 3$ but the first PAM component contains 99.63% of the total GMSK signal energy [39, 40], so we can approximate the signal radiated by the ith antenna as

$$s_i(t;\mathbf{b_i}) \approx \sqrt{\frac{2E_b}{T}} \sum_{k=-\infty}^{\infty} a_i(k)h(t-kT), \tag{25}$$

where E_b is the bit energy, T the symbol period, $a_i(k) = ja_i(k-1)b_i(k)$ are the transmitted symbols which belong to a QPSK constellation, $\mathbf{b}_i = \{b_i(k)\}_{k=-\infty}^{\infty}$ is the bit sequence to be modulated, and $h(t)$ is a pulse waveform that spans along the interval $[0, pT]$, where p is the memory of the modulation. It is demonstrated in [38] that the transmitted symbols $a_i(k)$ are uncorrelated and have unit variance. In order to simplify the detection process at the receiver, we will assume that a differential precoder is employed prior to modulation, that is, $d_i(k) = b_i(k-1)b_i(k)$ because we have then $a_i(k) = ja_i(k-1)d_i(k) = j^k b_i(k)$.

Considering that the transmission channel inside subway tunnels suffers from flat multipath fading [35], the signal received at the lth antenna is

$$y_l(t) = \sum_{i=1}^{N} h_{li}s_i(t;\mathbf{b}_i) + n_l(t), \tag{26}$$

where h_{li} is the fading observed between the ith transmitting antenna and the lth receiving antenna and $n_l(t)$ is a continuous-time complex-valued white Gaussian process with power spectral density $N_0/2$.

The received signals $y_l(t)$ are passed through a bank of filters matched to the pulse waveform $h(t)$ and sampled at the symbol rate in order to obtain a set of sufficient statistics for the detection of the transmitted symbols. Because $h(t)$ does not satisfy the zero-ISI condition, a discrete-time whitening filter [41, 42] is located after sampling. In addition, the rotation j^k induced by the GMSK modulation is compensated by multiplying the received signal by j^{-k}, resulting in the following expression for the observations:

$$\begin{aligned} x_l(k) &= \sum_{i=1}^{N} h_{li} \sum_{m=0}^{p-1} f(m)b_i(k-m) + v_l(k) \\ &= \sum_{i=1}^{N} h_{li}s_i(k) + v_l(k), \end{aligned} \tag{27}$$

where $v_l(k)$ represents the complex-valued AWGN with variance σ_v^2 and $f(m) = [0.8053, -0.5853j, -0.0704]$ is the equivalent discrete-time impulse response that takes into account the transmitting, receiving, and whitening filters, and the derotation operation. Using vector notation, the output of the whitening filters after the derotation can be expressed as

$$\mathbf{x}(k) = \mathcal{H}\mathbf{s}(k) + \mathbf{v}(k), \tag{28}$$

where $\mathbf{x}(k) = [x_1(k) \;\; x_2(k) \;\; \cdots \;\; x_L(k)]^T$ and

$$\mathcal{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N} \\ h_{21} & h_{22} & \cdots & h_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{L1} & h_{L2} & \cdots & h_{LN} \end{bmatrix}. \tag{29}$$

Equation (28) can be rewritten in the form of (1) as

$$\begin{aligned} \mathbf{x}(k) &= \begin{bmatrix} f(0)\mathcal{H} & f(1)\mathcal{H} & f(2)\mathcal{H} \end{bmatrix} \begin{bmatrix} \mathbf{b}(k-2) \\ \mathbf{b}(k-1) \\ \mathbf{b}(k) \end{bmatrix} + \mathbf{v}(k) \\ &\equiv \mathbf{H}\mathbf{z}(k) + \mathbf{v}(k). \end{aligned} \tag{30}$$

However, this signal model for the observations does not emphasize that the ISI comes from the GMSK modulation format instead of the time-dispersion of the multipath channel. As a consequence, we prefer to rewrite (28) as

$$\mathbf{x}(k) = \mathcal{H}\mathbf{B}(k)\mathbf{f} + \mathbf{v}(k), \tag{31}$$

where

$$\begin{aligned} \mathbf{B}(k) &= \begin{bmatrix} \mathbf{b}(k) & \mathbf{b}(k-1) & \mathbf{b}(k-2) \end{bmatrix}, \\ \mathbf{f} &= [0.8053, -0.5853j, -0.0704]^T. \end{aligned} \tag{32}$$

5.1. ML channel estimation for STC GSM-like systems with flat fading

Estimating the channel according to (30) and directly applying the method described in the previous section is highly inefficient because we have to estimate an unnecessarily large number of parameters. In addition, this way we do not take into account the knowledge at reception of the controlled ISI introduced by the modulator, given by $f(m)$. Equation (31) is preferable because it enables us to formulate the estimation of only the unknown channel coefficients h_{li}, as it is explained in the sequel. Again, we assume that the transmitted data frames contain a known pilot sequence of length M. The ML estimator of the channel is given by

$$\widehat{\mathcal{H}} = \arg\max_{\mathcal{H}} f_{\tilde{\mathbf{x}}|\tilde{\mathbf{b}}_i;\mathcal{H}}(\tilde{\mathbf{x}}). \tag{33}$$

This is a problem without closed-form solution, so we will apply the EM algorithm in a similar way to the general case explored in Section 4. We define the complete and incomplete data sets as $\tilde{\mathbf{x}}_e = [\tilde{\mathbf{b}}_f^T \;\; \tilde{\mathbf{x}}_f^T]^T$ and $\tilde{\mathbf{x}}_f$, respectively. Both sets are related through the linear transformation $\tilde{\mathbf{x}}_f = \mathbf{T}\tilde{\mathbf{x}}_e$, where $\mathbf{T} = [\mathbf{0}_{L(M+K)\times N(M+K)}\mathbf{I}_{L(M+K)}]$. Using the latter definitions, the $i+1$th estimate of the channel is computed using

the EM method as

$$\widehat{\mathcal{H}}_{i+1} = \arg\max_{\mathcal{H}} E_{\tilde{\mathbf{x}}_e|\tilde{\mathbf{x}}_f,\tilde{\mathbf{b}}_t;\widehat{\mathcal{H}}_i}\{\log f_{\tilde{\mathbf{x}}_e|\tilde{\mathbf{b}}_t;\mathcal{H}}(\tilde{\mathbf{x}}_e)\}. \tag{34}$$

Making similar manipulations to those made for the time-dispersive MIMO channel, we arrive at the following optimization problem:

$$\begin{aligned}\widehat{\mathcal{H}}_{i+1} = \arg\min_{\mathcal{H}} &\sum_{k=0}^{M-1} ||\mathbf{x}_t(k) - \mathcal{H}\mathbf{B}_t(k)\mathbf{f}||^2 \\ &+ \sum_{k=M}^{K-1} E_{\mathbf{b}(k)|\tilde{\mathbf{x}};\widehat{\mathcal{H}}_i}\left\{||\mathbf{x}(k) - \mathcal{H}\mathbf{B}(k)\mathbf{f}||^2\right\}\end{aligned} \tag{35}$$

which is also a quadratic optimization problem whose solution is

$$\widehat{\mathcal{H}}_{i+1} = \left(\mathbf{R}_{xb,t} + \mathbf{R}_{xb}\right) \times \left(\mathbf{R}_{b,t} + \mathbf{R}_b\right)^{-1}, \tag{36}$$

where

$$\begin{aligned}\mathbf{R}_{xb,t} &= \sum_{k=0}^{M-1} \mathbf{x}_t(k)\left(\mathbf{B}_t(k)\mathbf{f}\right)^H, \\ \mathbf{R}_{b,t} &= \sum_{k=0}^{M-1} \left(\mathbf{B}_t(k)\mathbf{f}\right)\left(\mathbf{B}_t(k)\mathbf{f}\right)^H, \\ \mathbf{R}_{xb} &= \sum_{k=M}^{K-1} E_{\mathbf{b}(k)|\tilde{\mathbf{x}};\widehat{\mathcal{H}}_i}\left\{\mathbf{x}(k)\left(\mathbf{B}(k)\mathbf{f}\right)^H\right\}, \\ \mathbf{R}_b &= \sum_{k=M}^{K-1} E_{\mathbf{b}(k)|\tilde{\mathbf{x}};\widehat{\mathcal{H}}_i}\left\{\left(\mathbf{B}(k)\mathbf{f}\right)\left(\mathbf{B}(k)\mathbf{f}\right)^H\right\}.\end{aligned} \tag{37}$$

Here we need to average with respect to the pdf $f_{\mathbf{B}(k)|\tilde{\mathbf{x}};\widehat{\mathcal{H}}_i}$. Again, we take benefit from the Turbo equalization process because

$$L[\mathbf{B}(k)|\tilde{\mathbf{x}};\widehat{\mathcal{H}}_i] = L[\mathbf{B}(k),\tilde{\mathbf{x}};\widehat{\mathcal{H}}_i] + h_B = L[e_k,\tilde{\mathbf{x}}] + h_B, \tag{38}$$

where h_B is the constant that makes $p_{\mathbf{B}(k)|\tilde{\mathbf{x}};\widehat{\mathcal{H}}_i}$ a probability mass function and $L[e_k,\tilde{\mathbf{x}}]$ is a quantity already computed in the Turbo equalization process.

6. SIMULATION RESULTS

6.1. Rayleigh MIMO channel

Computer simulations were carried out to illustrate the performance of the proposed channel estimator. Figure 3 plots the BER after decoding obtained for a 2×2 STC system over a nondispersive channel. Data are transmitted in blocks of 218 bits out of which the pilot sequence occupies $M = 10$ bits. The performance curves for both the LS method and when the channel is perfectly known are also shown for comparison. Note that there is no iteration gain when the channel is known because there is no ISI and, therefore, no "inner coding" for the Turbo processing. Nevertheless, this is not true when the ML-EM channel estimator is used because the channel is reestimated at each iteration of the Turbo equalization process. The ST encoder is a rate 1/2 full diversity convolutional binary code with generating matrix

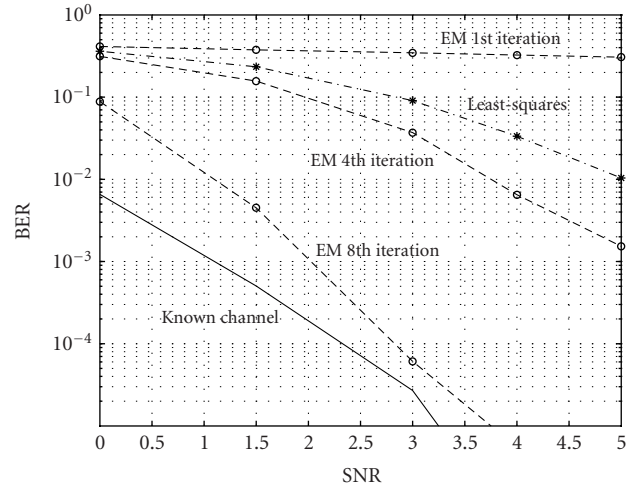


Figure 3: Performance results for ST coded data over a nondispersive channel.

$\mathbf{G} = [46, 72]$ in octal representation [26]. The independent interleavers are 20800 bits long each. The modulation format is BPSK and each channel coefficient is modeled as a zero-mean, complex-valued, circularly invariant Gaussian random process. Consequently, their magnitudes are Rayleigh distributed. We have also assumed that the channel coefficients are both temporally and spatially independent, having variance $\sigma_h^2 = 1/2$ per complex dimension. The signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = \frac{E\left\{\left(\mathbf{Hz}(k)\right)^H\left(\mathbf{Hz}(k)\right)\right\}}{E\{\mathbf{v}^H(k)\mathbf{v}(k)\}} = \frac{\text{Tr}\{\mathbf{HH}^H\}}{L\sigma_v^2}, \tag{39}$$

where $\text{Tr}\{\cdot\}$ denotes the trace operator. The channel changes at each transmitted block. Figure 3 shows that, even if its result for the first iteration is very poor, the ML-EM channel estimator outperforms the classical LS method from the fourth iteration.

The bad performance obtained by the ML-EM estimator at the first iteration comes from the fact that the Turbo equalizer is using an uninformative initial estimate of the channel. Specifically, (19) can be viewed as an LS estimator, where the correlation matrices $\mathbf{R}_{xz,t}$ and $\mathbf{R}_{z,t}$ have been modified by the addition of the matrices $\mathbf{R}_{xz}$ and $\mathbf{R}_z$, respectively. In the first iteration, these matrices are computed by assuming that $p_{\mathbf{z}(k)|\tilde{\mathbf{x}};\hat{\mathbf{H}}_i}$ is a uniform probability mass function (therefore, independent of the initial channel estimate $\widehat{\mathbf{H}}_0$) in (22) and (23). This results in a degradation of the pure LS estimator and a very high symbol error rate (SER) after decoding. Such a high SER (around 0.4) can never lead the Turbo equalization process to convergence. However, in our case, convergence is achieved because, in the next iterations, a substantial improvement is obtained in channel estimation from the EM algorithm (not from the Turbo structure itself). Notice that one iteration of the EM algorithm (19) is performed only after one complete equalization and decoding step. Anyway, once the channel estimate is good enough for the Turbo

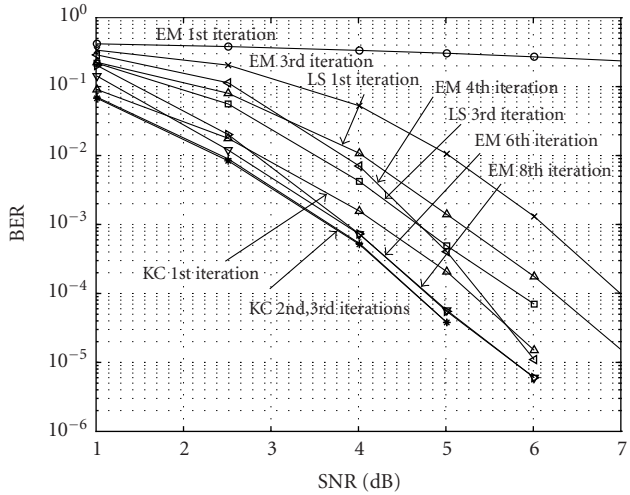


Figure 4: Performance results for ST coded data over a dispersive channel with memory $m = 2$.

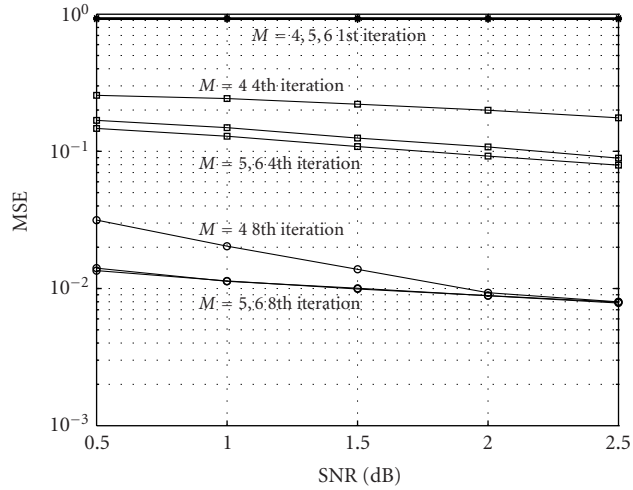


Figure 5: MSE for several lengths of the training sequence.

equalization structure to lie in its convergence region, both the EM algorithm and the Turbo iterative process help in reducing the error rate. Figure 3 also shows that at the eighth iteration, the performance is very close to the optimum, that is, known channel case. Only 0.5 dB separates the two curves at a BER of 10^{-4}.

Figure 4 shows the results (BER after decoding) obtained when a time-dispersive MIMO channel with memory $m = 2$ is considered. The simulation parameters are the same as in Figure 3. In particular, note that, again, each channel coefficient has variance $\sigma_h^2 = 1/2$ per complex dimension. It is apparent that at the fourth iteration, the ML-EM estimator performs very similar to the LS method, which does not improve significantly through the iterations. At the eighth iteration, the performance of the ML-EM estimator is again very close to the known channel case.

6.2. GSM-based transmission over subway tunnel MIMO channels

The performance of the proposed GSM-based transmission system with a Turbo STC receiver in subway tunnel environments has also been tested through computer simulations. The channel matrices $\mathcal{H}$ result from experimental measurements (carried out within the framework of the European project "ESCORT") of the MIMO channel impulse response present in a subway tunnel. The experimental setup consisted of four transmitting antennas, each one having a 12 dBi gain, located at the station platform, and four patch antennas located behind the train windscreen. The complex impulse responses were measured with a channel sounder having a bandwidth of 35 MHz by switching successively the antennas and stopping the train approximately each 2 m. From the whole set of 4×4 measured subchannels, only those corresponding to the furthest antennas were picked up for constructing a 2×2 system. In [35], it was demonstrated that the mean capacity of the measured channel is less than the capacity of Rayleigh fading channels, this difference being more remarkable in the case of a 4×4 system.

The ability of our channel estimation technique to combine the deterministic information of the pilot symbols and the statistical information from the unknown symbols, thanks to the ST Turbo detector, enables us to considerably reduce the size of the training sequence in GSM systems. Indeed, by means of computer simulations, we have determined the minimum length of the training sequence for the considered GSM-based MIMO system. Figure 5 shows the channel estimation mean square error (MSE) for several values of the training sequence length ($M = 4$, 5, and 6 bits). The channel code is the same as in the previous simulations. The interleaver size is 20800 bits and the frame length is 148, as established in the GSM standard. There is a significant difference in the estimation error between using $M = 4$ bits and $M = 5$ bits, whereas the gap between $M = 5$ and $M = 6$ is very small. This points out that $M = 5$ bits is the minimum length for the training sequence. This assumption can also be corroborated in Figure 6, where the SER at the output of the decoder is plotted versus the required SNR.

Next, we compare the results obtained with the proposed estimator using a training sequence of $M = 5$ bits and those obtained with classical LS using a training sequence of $M = 26$ bits (the length standardized in GSM). The results obtained when the receiver perfectly knows the channel are also plotted for comparison. As it is shown in Figure 7, the proposed method (ML-EM) with $M = 5$ bits performs better than the LS with $M = 26$ bits beyond the sixth iteration, achieving a performance very close to the known channel case beyond the seventh iteration.

7. CONCLUSIONS

In this paper, we propose a novel ML-based time-dispersive MIMO channel estimator for STC systems that employ

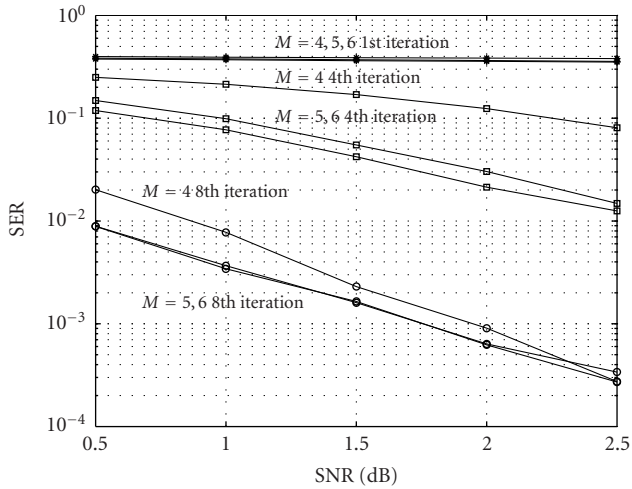


Figure 6: SER versus SNR at the output of the decoder for several lengths of the training sequence.

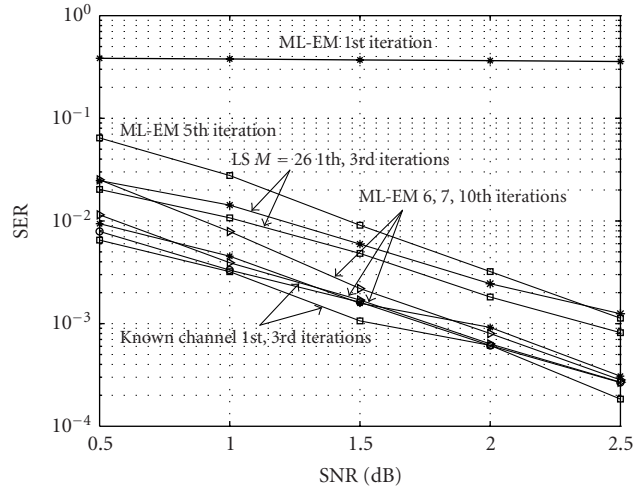


Figure 7: Performance comparison between ML-EM ($M = 5$ bits), LS ($M = 26$ bits), and known channel.

Turbo ST receivers. We formulate the ML estimation problem that takes into account the deterministic symbols corresponding to the training sequence and the statistics of the unknown symbols. These statistics can be obtained and successively refined if an ST Turbo equalizer is used at reception. This full exploitation of all the available statistical information at reception renders an extremely powerful channel estimation technique that outperforms conventional approaches based only on the training sequence. Since the involved optimization problem has no closed-form solution, the EM algorithm is employed in order to iteratively obtain the solution. The main limitation of our approach is that the computational complexity of the channel estimator grows exponentially with the number of transmitting antennas and the channel memory size, hence it is only practical for a moderate size of the transmitter antenna array. Note, however, that this complexity is inherent to the problem of optimal detection and estimation in MIMO systems.

The method has been particularized for a realistic scenario in which an STC system based on the GSM standard transmits along railway subway tunnels. Simulation results show how our channel estimation technique enables us to diminish the training sequence length up to only 5 bits, instead of the 26 bits considered in the GSM standard, thus achieving a 14% increase in the system throughput.

APPENDICES

A. SIGNAL MODEL OF AN STC GSM SYSTEM

The transmitter model depicted in Figure 1 is valid for an STC GSM system. The signal radiated by ith antenna is given by [38, 40]

$$s_i(t;\mathbf{b}_i) = \sqrt{\frac{2E_b}{T}} \exp\left[j\pi \sum_{k=-\infty}^{\infty} b(k) q(t-kT) \right], \quad (A.1)$$

where E_b is the bit energy, T the symbol period, $\mathbf{b}_i = \{b_i(k)\}_{k=-\infty}^{\infty}$ the bit sequence to be modulated, and

$$q(t) = \int_{-\infty}^{t} g(\tau) d\tau, \quad (A.2)$$

where $g(t)$ is the convolution between a Gaussian-shaped pulse and a rectangular-shaped pulse centered at the origin [43, 44], that is,

$$g(t) = u(t) * \operatorname{rect}\left(\frac{t}{T}\right), \quad (A.3)$$

where

$$\operatorname{rect}\left(\frac{t}{T}\right) = \begin{cases} \dfrac{1}{2T}, & |t| \le \dfrac{T}{2}, \\ 0, & \text{otherwise}, \end{cases}$$
$$u(t) = \frac{1}{\sqrt{2\pi}\sigma_u} \exp\left\{ -\frac{1}{2}\left(\frac{t}{\sigma_u}\right)^2 \right\}, \quad (A.4)$$

with

$$\sigma_u = \frac{\sqrt{\log 2}}{2\pi B}, \quad (A.5)$$

where B is the 3 dB bandwidth of $u(t)$. It is possible to derive a closed-form expression for $g(t)$ given by [38, 40]

$$g(t) = \frac{1}{2T}\left[Q\left(\frac{t - T/2}{\sigma_u}\right) - Q\left(\frac{t + T/2}{\sigma_u}\right) \right], \quad (A.6)$$

where

$$Q(t) = \frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{-\tau^2/2} d\tau \quad (A.7)$$

is the Gaussian complementary error function. With the aim

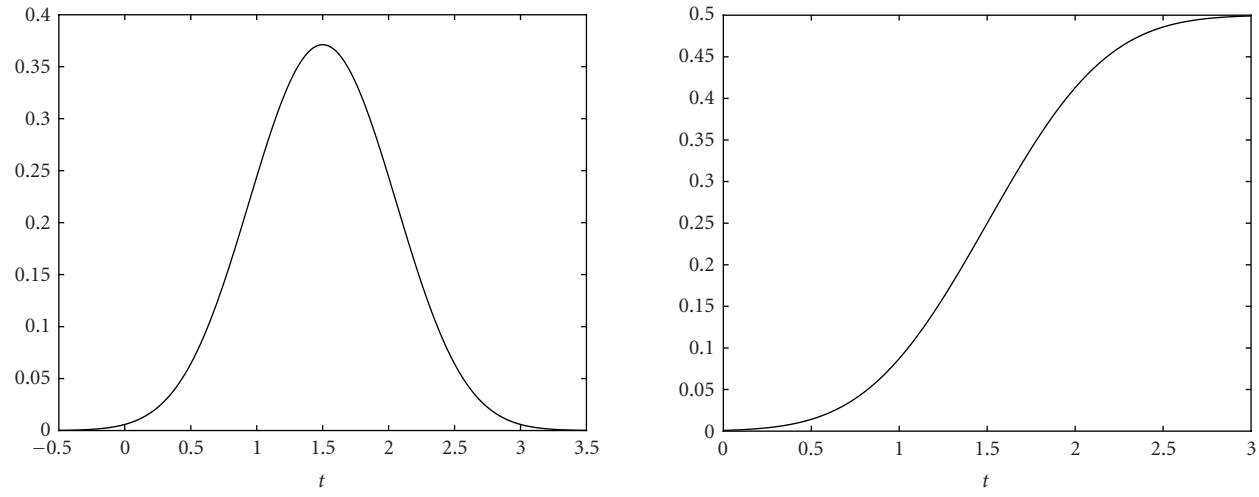


FIGURE 8: (a) Shifted GMSK pulse, $g(t - 1.5T)$, for $p = 3$. (b) GMSK phase pulse, $q(t)$.

of simplifying subsequent analysis, we redefine $g(t) \equiv g(t - p/2T)$, so it is limited to the interval $[0, pT]$, where p is the number of symbol periods where the signal has significant values. For GSM ($B = 0.3$), a value of $p = 3$ is reasonable [40], as it can be verified in Figure 8, that plot the properly shifted versions of $g(t)$ and $q(t)$ when $B = 0.3$.

Since GMSK is a partial response CPM, it can be expressed in terms of its Laurent expansion [38, 39, 40], formed by the sum of 2^{p-1} PAM signals, where p is the memory induced by the modulation. Since in GSM, the first PAM component contains 99.63% of the total GMSK signal energy [39, 40], we can approximate the signal radiated by the ith antenna by

$$s_i(t; \mathbf{b_i}) \approx \sqrt{\frac{2E_b}{T}} \sum_{k=-\infty}^{\infty} a_i(k) h(t - kT), \tag{A.8}$$

where $a_i(k) = ja_i(k-1)b_i(k)$ are the transmitted symbols, which belong to a QPSK constellation, are uncorrelated and have unit variance [38]. In order to simplify the detection process at the receiver, we will assume that a differential precoder is employed prior to modulation, that is, $d_i(k) = b_i(k-1)b_i(k)$ because then we have $a_i(k) = ja_i(k-1)d_i(k) = j^k b_i(k)$. The pulse waveform $h(t)$ is equal to $C(t - 3T)C(t - 2T)C(t - T)$, where $C(t) = \cos(\pi q(|t|))$. Figure 9a shows that it takes significant values over the interval $[0.5T, 3.5T]$ because the actual and the linearized GMSK waveforms are shifted by half a symbol period.

In order to detect the transmitted symbols, $s_i(t; \mathbf{b}_i)$ is passed through a filter matched to the pulse waveform $h(t)$ and then sampled at the symbol rate. The output of the matched filter is given by

$$\begin{aligned} r_i(t) &= a_i(t) * h(t) * h^*(-t) + n(t) * h^*(-t) \\ &= a_i(t) * R_h(t) + g(t), \end{aligned} \tag{A.9}$$

where

$$a_i(t) = \sqrt{\frac{2E_b}{T}} \sum_{k=-\infty}^{\infty} a_i(k)\delta(t - kT) \tag{A.10}$$

and $R_h(t)$ (see Figure 9b) denotes the autocorrelation function of $h(t)$. After sampling, we have

$$r_i(k) \equiv r_i(t = kT) = a_i(k) * R_h(k) + g(k), \tag{A.11}$$

where the autocorrelation function of $g(k)$ is $R_g(k) = (N_0/2)R_h(k)$. Clearly, the noise $g(k)$ is colored because $h(t)$ does not satisfy the zero-ISI condition. Since it is more comfortable to perform detection assuming white noise, a discrete-time whitening filter [41, 42] is located after sampling

$$W(z) = \frac{1}{F^*(z^{-1})}, \tag{A.12}$$

where $F^*(z^{-1})$ comes from the factorization of the autocorrelation function $R_h(k) = F(z)F^*(z^{-1})$. This expression for the whitening filter leads to an overall system response given by $F(z)$. In Appendix B, we demonstrate that the maximum phase $F(z)$ polynomial is given by

$$\begin{aligned} F(z) &= \sqrt{\frac{r_2}{\rho_1\rho_2}}\left(1 - \rho_1 z^{-1}\right)\left(1 - \rho_2 z^{-1}\right) \\ &= 0.8053 + 0.5853z^{-1} + 0.0704z^{-2}, \end{aligned} \tag{A.13}$$

where $\rho_1 = -0.1522$, $\rho_2 = -0.5746$, and $r_2 = R_h(-2)$.

In addition, the rotation j^k induced by the GMSK is compensated by multiplying the received signal by j^{-k}, resulting

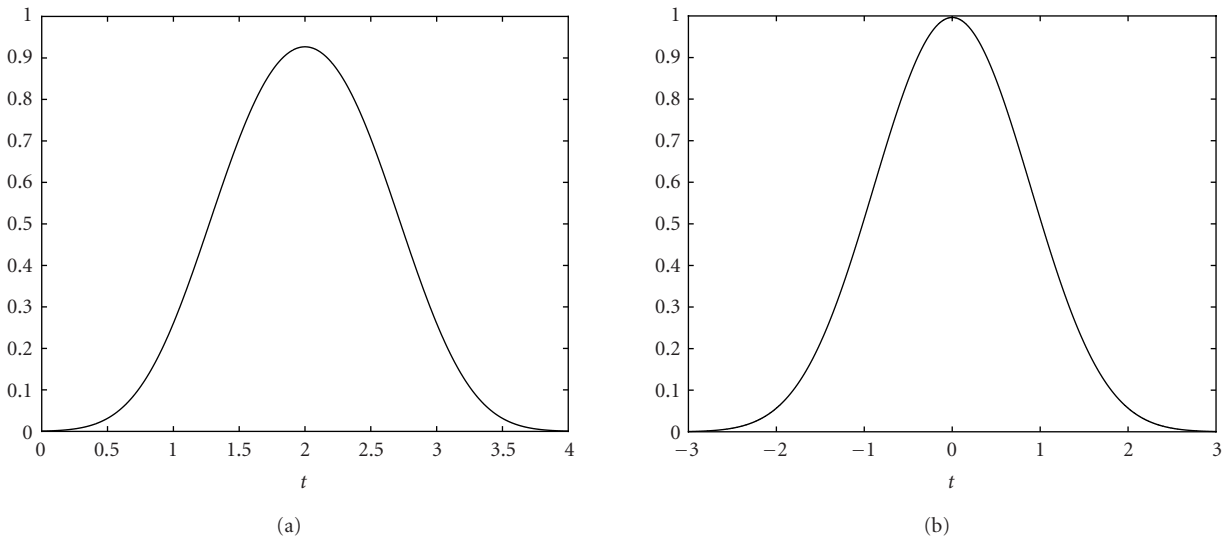


Figure 9: (a) Pulse shape, $h(t)$. (b) Autocorrelation function, $R_h(t)$.

in the following expression for the observations:

$$\begin{aligned} x_l(k) &= \sum_{i=1}^{N} h_{li} \sum_{m=0}^{p-1} f(m) b_i(k-l) + v_l(k) \\ &= \sum_{i=1}^{N} h_{li} s_i(k) + v_l(k), \end{aligned} \tag{A.14}$$

where $v_l(k)$ is AWGN with variance σ_v^2 and $f(m) = [0.8053 \quad -0.5853j \quad -0.0704]$ is the equivalent discrete-time impulse response that takes into account the transmitting, receiving, and whitening filters, and the derotation operation. Using vector notation, the output of the whitening filters after the derotation can be expressed as in (28), where $\mathbf{x}(k) = [x_1(k) \; x_2(k) \; \cdots \; x_L(k)]^T$, $\mathbf{s}(k) = [s_1(k) \; s_2(k) \; \cdots \; s_N(k)]^T$, and is as in (29).

B. COMPUTATION OF THE DISCRETE-TIME WHITENING FILTER

First, it is important to note that there are 2^p choices of $F(z)$ that satisfy the desired factorization $R_h(z) = F(z)F^*(z^{-1})$. The different choices yield filters $1/F^*(z^{-1})$ that have the same magnitude but different phase response. One possible choice is to select $F^*(z^{-1})$ so that it is a minimum phase, that is, with all its roots inside the unit circle. In this way, $1/F^*(z^{-1})$ is a realizable causal and stable discrete system. The problem of this selection is that the overall impulse response $F(z)$ will be the maximum phase and anticausal, and the resulting ISI will be difficult to compensate. To overcome this limitation, we choose $F^*(z^{-1})$ to be the maximum phase and thus the whitening filter $1/F^*(z^{-1})$ will be stable only if it is considered anticausal. Nevertheless, anticausal filters can be implemented if a sufficient large delay is introduced. The advantage of this approach is that now the overall impulse response $F(z)$ is causal and minimum phase.

Considering that $R_h(t)$ takes significant values only over the interval $[-2, 2]$ (see Figure 9b), we have

$$\begin{aligned} R_h(k) &= \{r_{-2}, r_{-1}, r_0, r_1, r_2\} = \{r_2^*, r_1^*, r_0, r_1, r_2\} \\ &= \{0.0567, 0.5127, 0.9963, 0.5127, 0.0567\}. \end{aligned} \tag{B.1}$$

As mentioned before, note that $h(t)$ contains 99.63% of the actual GMSK total energy because $R_h(0) = 0.9963$. The Z-transform of $R_h(k)$ is

$$R_h(z) = r_{-2}z^2 + r_{-1}z + r_0 + r_1 z^{-1} + r_2 z^{-2} \tag{B.2}$$

that we can express as

$$R_h(z) = r_2 \left(\frac{z-1}{\rho_1^*}\right)\left(\frac{z-1}{\rho_2^*}\right)(1-\rho_1 z^{-1})(1-\rho_2 z^{-1}). \tag{B.3}$$

Forcing $|\rho_1|, |\rho_2| \leq 1$ in order that the resulting whitening filter exists and be stable, we have $\rho_1 = -0.1522$ and $\rho_2 = -0.5746$. Taking into account that ρ_1 and ρ_2 are real valued, we arrive at

$$F(z) = \sqrt{\frac{r_2}{\rho_1 \rho_2}}(1-\rho_1 z^{-1})(1-\rho_2 z^{-1}) \tag{B.4}$$

and thus the whitening filter is given by

$$W(z) = \frac{1}{F^*(z^{-1})} = \frac{\sqrt{\rho_1 \rho_2 / r_2}}{(1-\rho_1 z)(1-\rho_2 z)} \tag{B.5}$$

whose inverse Z-transform is

$$w(k) = \{w_k\}_{k=-\infty}^{0} = \left\{ \frac{\sqrt{\rho_1 \rho_2 / r_2}}{\rho_2 - \rho_1} \left(\rho_2^{-k+1} - \rho_1^{-k+1}\right) \right\}. \tag{B.6}$$

Since $\{|w_k|\}_{k=0}^{-\infty}$ is a strictly decreasing series, we can consider only the first significant w_k coefficients. Taking into account

that $|w_{-20}| < 10^{-4}$, we can implement $w(k)$ as an anticausal FIR filter:

$$\begin{aligned} w(k) &\approx \{w_{-19}, w_{-18}, \ldots, w_{-1}, w_0\} \\ &= \{0.0001, -0.0001, 0.0002, -0.0004, 0.0007, \\ &\quad -0.0013, 0.0022, -0.0038, 0.0066, -0.0115, \\ &\quad 0.0201, -0.0349, 0.0608, -0.1058, 0.1839, \\ &\quad 0.3189, 0.5473, -0.9025, 1.2417\}. \end{aligned} \tag{B.7}$$

ACKNOWLEDGMENT

This work has been supported by the European Commission under Contract no. IST-1999-20006 (ESCORT project) and by Ministerio de Ciencia y Tecnología of Spain and FEDER funds from the European Union under Grant no. TIC2001-0751-C04-01.

REFERENCES

[1] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Trans. Communications*, vol. 44, no. 10, pp. 1261–1271, 1996.

[2] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 429–445, 1996.

[3] C. Heegard and S. B. Wicker, *Turbo Coding*, Kluwer Academic Publishers, Boston, Mass, USA, 1999.

[4] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "A soft-input soft-output APP module for iterative decoding of concatenated codes," *IEEE Communications Letters*, vol. 1, no. 1, pp. 22–24, 1997.

[5] C. Douillard, A. Picart, P. Didier, M. Jézéquel, C. Berrou, and A. Glavieux, "Iterative correction of intersymbol interference: turbo-equalization," *European Transactions on Telecommunications*, vol. 6, no. 5, pp. 507–511, 1995.

[6] C. Luschi, M. Sandell, P. Strauch, et al., "Advanced signal-processing algorithms for energy-efficient wireless communications," *Proceedings of the IEEE*, vol. 88, no. 10, pp. 1633–1650, 2000.

[7] X. Wang and R. Chen, "Blind turbo equalization in Gaussian and impulsive noise," *IEEE Trans. Vehicular Technology*, vol. 50, no. 4, pp. 1092–1105, 2001.

[8] C. Laot, A. Glavieux, and J. Labat, "Turbo equalization: adaptive equalization and channel decoding jointly optimized," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1744–1752, 2001.

[9] Z. Yang and X. Wang, "Turbo equalization for GMSK signaling over multipath channels based on the Gibbs sampler," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1753–1763, 2001.

[10] B. L. Yeap, T. H. Liew, J. Hámorský, and L. Hanzo, "Comparative study of turbo equalization schemes using convolutional, convolutional turbo, and block-turbo codes," *IEEE Transactions on Wireless Communications*, vol. 1, no. 2, pp. 266–273, 2002.

[11] N. Görtz, "On the iterative approximation of optimal joint source-channel decoding," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1662–1670, 2001.

[12] J. Garcia-Frias and J. D. Villasenor, "Joint turbo decoding and estimation of hidden Markov sources," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1671–1679, 2001.

[13] A. Guyader, E. Fabre, C. Guillemot, and M. Robert, "Joint source-channel turbo decoding of entropy-coded sources," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1680–1696, 2001.

[14] L. Zhang and A. Burr, "APPA symbol timing recovery scheme for turbo-codes," in *Proc. 13th IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications Conference*, Lisbon, Portugal, September 2002.

[15] M. C. Valenti and B. D. Woerner, "Iterative channel estimation and decoding of pilot symbol assisted turbo codes over flat-fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1697–1705, 2001.

[16] C. Komninakis and R. D. Wesel, "Joint iterative channel estimation and decoding in flat correlated Rayleigh fading," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1706–1717, 2001.

[17] K.-D. Kammeyer, V. Kühn, and T. Petermann, "Blind and nonblind turbo estimation for fast fading GSM channels," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1718–1728, 2001.

[18] A. Mehrotra, *GSM System Engineering*, Artech House, Boston, Mass, USA, 1997.

[19] A. O. Berthet, B. S. Ünal, and R. Visoz, "Iterative decoding of convolutionally encoded signals over multipath Rayleigh fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1729–1743, 2001.

[20] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, John Wiley & Sons, New York, NY, USA, 1997.

[21] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[22] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[23] A. Grant, "Rayleigh fading multi-antenna channels," *EURASIP Journal on Applied Signal Processing*, vol. 2002, no. 3, pp. 316–329, 2002.

[24] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.

[25] A. F. Naguib, N. Seshadri, and A. R. Calderbank, "Increasing data rate over wireless channels," *IEEE Signal Processing Magazine*, vol. 17, no. 3, pp. 76–92, 2000.

[26] A. R. Hammons Jr. and H. El Gamal, "On the theory of space-time codes for PSK modulation," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 524–542, 2000.

[27] X. Lin and R. S. Blum, "Improved space-time codes using serial concatenation," *IEEE Communications Letters*, vol. 4, no. 7, pp. 221–223, 2000.

[28] H.-J. Su and E. Geraniotis, "Space-time turbo codes with full antenna diversity," *IEEE Trans. Communications*, vol. 49, no. 1, pp. 47–57, 2001.

[29] Y. Liu, M. P. Fitz, and O. Y. Takeshita, "Full rate space-time turbo codes," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 5, pp. 969–980, 2001.

[30] S. Lek, "Turbo space-time processing to improve wireless channel capacity," *IEEE Trans. Communications*, vol. 48, no. 8, pp. 1347–1359, 2000.

[31] A. Stefanov and T. M. Duman, "Turbo-coded modulation for systems with transmit and receive antenna diversity over block fading channels: system model, decoding approaches, and practical considerations," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 5, pp. 958–968, 2001.

[32] M. González-López, A. Dapena, and L. Castedo, "MAP space-time receivers for GSM in subway tunnel environments," in *Proc. 11th European Signal Processing Conference*, Toulouse, France, September 2002.
[33] M. González-López, J. Míguez, and L. Castedo, "Decision feedback Turbo equalization for space-time coded systems," in *Proc. IEEE 28th Int. Conf. Acoustics, Speech, Signal Processing*, Hong Kong, China, April 2003.
[34] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1992.
[35] M. González-López, A. Dapena, and L. Castedo, "Space-time coding for GSM systems in subway tunnel environments," in *Proc. IEEE 27th Int. Conf. Acoustics, Speech, Signal Processing*, Orlando, Fla, USA, May 2002.
[36] J. Baudet, M. González-López, D. Degardin, et al., "Performance of space time coding in subway tunnel environments," in *Proc. IEE Technical Seminar on MIMO Communication Systems: from Concept to Implementation*, pp. 2/1–2/6, London, UK, December 2001.
[37] ESCORT, "Enhanced diversity and space-time coding for metro and railway transmission," Final Tech. Rep. D 6021, France, 2002.
[38] U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*, Plenum Press, New York, NY, USA, 1997.
[39] N. Al-Dhahir and G. Saulnier, "A high-performance reduced-complexity GMSK demodulator," *IEEE Trans. Communications*, vol. 46, no. 11, pp. 1409–1412, 1998.
[40] N. Al-Dhahir and G. Saulnier, "A high-performance reduced-complexity GMSK demodulator," Tech. Rep. 96CRD107, GE Global Research, Niskayuna, NY, USA, 1996.
[41] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 3rd edition, 1995.
[42] J. Kurzweil, *An Introduction to Digital Communications*, John Wiley & Sons, New York, NY, USA, 2000.
[43] J. M. H. Rábanos, *Comunicaciones Móviles GSM*, Fundación Airtel Móvil, Madrid, Spain, 1999.
[44] J. D. Laster, *Robust GMSK Demodulation using demodulator diversity and BER estimation*, Ph.D. thesis, Virginia Polytechnic Institute and State University, Blacksburg, Va, USA, 1997.

Miguel González-López was born in Santiago de Compostela, Spain, in 1977. He received his Ingeniero en Informática (M.S.) degree from Universidade da Coruña in 2000, where he is currently working to obtain his Ph.D. degree. His research interests include the application of the Turbo principle to channel estimation/equalization and coding on graphs, with special focus on their generalization to MIMO systems and their implementation issues.

Joaquín Míguez was born in Ferrol, Galicia, Spain, in 1974. He obtained his Licenciado en Informática (M.S.) and Doctor en Informática (Ph.D.) degrees from Universidade da Coruña, Spain, in 1997 and 2000, respectively. Late in 2000, he joined the Departamento de Electrónica y Sistemas, Universidade da Coruña, where he became an Associate Professor in July 2003. From April 2001 through December 2001, he was a Visiting Scholar in the Department of Electrical and Computer Engineering, the State University of New York at Stony Brook. His research interests are in the field of statistical signal processing with emphasis on the topics of Bayesian analysis, sequential Monte Carlo methods, adaptive filtering, stochastic optimization, and their applications to multiuser communications, smart antenna systems, target tracking, and vehicle positioning and navigation.

Luis Castedo was born in Santiago de Compostela, Spain, in 1966. He received his Ingeniero de Telecomunicación (M.S.) and Doctor Ingeniero de Telecomunicación (Ph.D.) degrees, both from Universidad Politécnica de Madrid (UPM), Spain, in 1990 and 1993, respectively. From 1990 to 1994, he was with the Departamento de Señales, Sistemas y Radiocomunicación at the UPM, where he worked in array processing applied to digital communications. During the academic year 1991/92, he was a Visiting Scholar at the University of Southern California, USA. In 1994, he joined the Departamento de Electrónica y Sistemas at Universidad da Coruña, Spain, where he is currently a Professor and teaches courses in signal processing, digital communications, and linear control systems. His research interests include adaptive filtering and signal processing methods for space and code diversity exploitation in communication systems.

EURASIP Journal on Applied Signal Processing 2004:5, 740–750

Space-Time Chip Equalization for Maximum Diversity Space-Time Block Coded DS-CDMA Downlink Transmission

Geert Leus
Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, Mekelweg 4, 2628CD Delft, The Netherlands
Email: leus@cas.et.tudelft.nl

Frederik Petré
Wireless Research, Interuniversity Micro-Electronics Center (IMEC), Kapeldreef 75, 3001 Leuven, Belgium
Email: petre@imec.be

Marc Moonen
Department of Electrical Engineering (ESAT), Katholieke Universiteit Leuven (K.U.Leuven), Kasteelpark Arenberg 10, 3001 Leuven, Belgium
Email: moonen@esat.kuleuven.ac.be

Received 24 December 2002; Revised 4 August 2003

In the downlink of DS-CDMA, frequency-selectivity destroys the orthogonality of the user signals and introduces multiuser interference (MUI). Space-time chip equalization is an efficient tool to restore the orthogonality of the user signals and suppress the MUI. Furthermore, multiple-input multiple-output (MIMO) communication techniques can result in a significant increase in capacity. This paper focuses on space-time block coding (STBC) techniques, and aims at combining STBC techniques with the original single-antenna DS-CDMA downlink scheme. This results into the so-called space-time block coded DS-CDMA downlink schemes, many of which have been presented in the past. We focus on a new scheme that enables both the maximum multiantenna diversity and the maximum multipath diversity. Although this maximum diversity can only be collected by maximum likelihood (ML) detection, we pursue suboptimal detection by means of space-time chip equalization, which lowers the computational complexity significantly. To design the space-time chip equalizers, we also propose efficient pilot-based methods. Simulation results show improved performance over the space-time RAKE receiver for the space-time block coded DS-CDMA downlink schemes that have been proposed for the UMTS and IS-2000 W-CDMA standards.

Keywords and phrases: downlink CDMA, space-time block coding, space-time chip equalization.

1. INTRODUCTION

Direct sequence code division multiple access (DS-CDMA) has emerged as the predominant multiple access technique for 3G cellular systems. In the downlink of DS-CDMA, orthogonal user signals are transmitted from the base station. All these signals are distorted by the same channel when propagating to the desired mobile station. Hence, when this channel is frequency-selective, the orthogonality of the user signals is destroyed and severe multiuser interference (MUI) is introduced. Space-time chip equalization can then restore the orthogonality of the user signals and suppress the MUI [1, 2, 3, 4].

Multiple-input multiple-output (MIMO) systems, on the other hand, have recently been shown to realize a significant increase in capacity for rich scattering environments [5, 6, 7]. Both space division multiplexing (SDM) [8, 9] and space-time coding (STC) [10, 11, 12] are popular MIMO communication techniques. SDM techniques mainly aim at an increase in throughput by transmitting different data streams from the different transmit antennas. However, SDM typically requires as many receive as transmit antennas, which seriously impairs a cost-efficient implementation at the mobile station. STC techniques, on the other hand, mainly aim at an increase in performance by introducing spatial and temporal correlation in the transmitted data streams. As opposed to SDM, STC supports any number of receive antennas, and thus enables a cost-efficient implementation at the mobile station. In this perspective, space-time block coding (STBC) techniques, introduced in [11] for two transmit antennas and

later generalized in [12] for any number of transmit antennas, are particularly appealing because they facilitate maximum likelihood (ML) detection with simple linear processing. However, these STBC techniques have originally been developed for signaling over frequency-flat channels, and do not enable the maximum multiantenna and multipath diversity present in frequency-selective channels. Therefore, improved STBC techniques have recently been developed for signaling over frequency-selective channels [13, 14, 15]. The STBC technique proposed in [13] enables the maximum multiantenna diversity, and although it is presented as a technique that provides the maximum multipath diversity, it is not possible to prove it without any proper discussion on how to treat the edge effects at the beginning and the end of a burst. If the edge effects are handled by a cyclic prefix as in [14], maximum multipath diversity is not guaranteed. On the other hand, if the edge effects are handled by a zero postfix as in [15], maximum multipath diversity is guaranteed.

Up till now, research on STBC techniques has mainly focused on single-user communication links. In this paper, we aim at combining STBC techniques with the original single-antenna DS-CDMA downlink scheme, resulting into so-called space-time block coded DS-CDMA downlink schemes. As an example, we mention the space-time block coded DS-CDMA downlink schemes that have been proposed for the UMTS and IS-2000 W-CDMA standards, both special cases of the so-called space-time spreading scheme presented in [16], which consists of a mixture of the original single-antenna DS-CDMA downlink scheme and the STBC technique of [12]. However, this scheme does not enable the maximum multiantenna and multipath diversity present in frequency-selective channels. A second example is the space-time block coded DS-CDMA downlink scheme presented in [17], which consists of the original single-antenna DS-CDMA downlink scheme followed by the STBC technique of [14]. However, this scheme only enables the maximum multiantenna diversity but not the maximum multipath diversity (due to the fact that maximum multipath diversity is not provided by the STBC technique of [14]). Therefore, in this paper, we consider the space-time block coded DS-CDMA downlink scheme that consists of the original single-antenna DS-CDMA downlink scheme followed by the STBC technique of [15]. This scheme enables both the maximum multiantenna diversity and the maximum multipath diversity (due to the fact that maximum multipath diversity is provided by the STBC technique of [15]). Although this maximum diversity can only be collected by ML detection, we pursue suboptimal detection by means of space-time chip equalization, which lowers the computational complexity significantly. Note that this suboptimal detection technique can also be applied to the STBC technique of [15] on its own, without combining it with the original single-antenna DS-CDMA downlink scheme.

Assuming there are J transmit antennas, the straightforward way to implement space-time chip equalization is to apply J space-time chip equalizers to recover the J transmitted space-time block coded multiuser chip sequences, then to apply space-time decoding to recover J subsequences of the original multiuser chip sequence, and finally, to perform simple despreading. Since this comes down to an equalization problem with J sources, we need $J + 1$ chip rate sampled outputs at each mobile station for a finite-length zero-forcing (ZF) solution to exist (i.e., $J + 1$ receive antennas if the antennas are sampled at chip rate). However, we will show that the space-time chip equalization and space-time decoding operations can be swapped, which allows us to first apply space-time decoding, then to apply J space-time chip equalizers to recover J subsequences of the original multiuser chip sequence, and finally, to perform simple despreading. Since this comes down to J equalization problems with only one source, we need only two chip rate sampled outputs at each mobile station for a finite-length ZF solution to exist (i.e., two receive antennas if the antennas are sampled at chip rate). To design the space-time chip equalizers, we finally propose efficient pilot-based methods.

In Section 2, we discuss the transceiver design of the proposed space-time block coded DS-CDMA system. We distinguish between the transmitter design, the channel model, and the receiver design, where the latter is based on space-time chip equalization. In Section 3, we then propose two pilot-based methods for practical space-time chip equalizer design. We show some simulation results in Section 4. In Section 5, we finally draw our conclusions.

Notation

We use upper (lower) bold face letters to denote matrices (vectors). Superscripts $*$, T, and H represent conjugate, transpose, and Hermitian, respectively. Further, $\lfloor \cdot \rfloor$ represents the flooring operation, and $\mathcal{E}\{\cdot\}$ represents the expectation operation. We denote the $N \times N$ identity matrix as $\mathbf{I}_N$ and the $M \times N$ all-zero matrix as $\mathbf{0}_{M\times N}$. Next, $[\mathbf{A}]_{m,n}$ denotes the entry at position (m, n) of the matrix $\mathbf{A}$. Finally, $\mathrm{diag}\{\mathbf{a}\}$ represents the diagonal matrix with the vector $\mathbf{a}$ on the diagonal.

2. TRANSCEIVER DESIGN

We consider the downlink of a space-time block coded DS-CDMA system. We assume the base station is equipped with J transmit antennas, and the mobile station is equipped with M receive antennas. In the following, we discuss the transmitter design, the channel model, and the receiver design.

2.1. Transmitter design

At the base station, a space-time block coded DS-CDMA downlink scheme transforms $\{s_u[k]\}_{u=1}^{U}$ and $s_p[k]$, where $s_u[k]$ is the uth user's data symbol sequence and $s_p[k]$ is the pilot symbol sequence, into J space-time block coded multiuser chip sequences $\{u_j[n]\}_{j=1}^{J}$.

We consider the space-time block coded DS-CDMA downlink scheme that consists of the original single-antenna CDMA downlink transmission scheme followed by the STBC technique of [15]. This scheme enables both the maximum multiantenna diversity and the maximum multipath diversity. For simplicity, we will focus on the case of $J = 2$ transmit antennas. Extensions to more than two transmit antennas

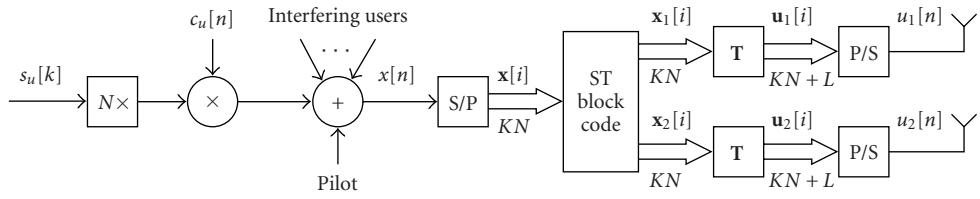


Figure 1: Proposed space-time block coded DS-CDMA downlink scheme.

($J > 2$) are straightforward and can be developed following the design rules presented in [18].

Figure 1 depicts the proposed space-time block coded DS-CDMA downlink scheme ($N\times$ repeats each sample N times, whereas "S/P" and "P/S" represent a serial-to-parallel and parallel-to-serial conversion, respectively). First, the original multiuser chip sequence $x[n]$ is constructed:

$$x[n] := \sum_{u=1}^{U} s_u[\lfloor n/N \rfloor] c_u[n] + s_p[\lfloor n/N \rfloor] c_p[n], \tag{1}$$

where $c_u[n]$ is the uth user's code sequence and $c_p[n]$ is the pilot code sequence. We assume that both $c_u[n]$ and $c_p[n]$ are normalized and consist of a multiplication of a user/pilot specific orthogonal Walsh-Hadamard spreading code of length N and a base-station specific long scrambling code. Note that the above pilot insertion technique is similar to the so-called common pilot channel (CPICH) [19] in forthcoming 3G systems. Second, the original multiuser chip sequence $x[n]$ is serial-to-parallel converted into the $1 \times KN$ multiuser chip block sequence $\mathbf{x}[i]$:

$$\mathbf{x}[i] := [x[iKN], \ldots, x[(i+1)KN - 1]]. \tag{2}$$

Third, the multiuser chip block sequence $\mathbf{x}[i]$ is transformed into the two $1 \times KN$ block sequences $\mathbf{x}_1[i]$ and $\mathbf{x}_2[i]$:

$$\begin{bmatrix} \mathbf{x}_1[2i] & \mathbf{x}_1[2i+1] \\ \mathbf{x}_2[2i] & \mathbf{x}_2[2i+1] \end{bmatrix} := \begin{bmatrix} \mathbf{x}[2i] & -\mathbf{x}^*[2i+1]\mathbf{P}_{KN} \\ \mathbf{x}[2i+1] & \mathbf{x}^*[2i]\mathbf{P}_{KN} \end{bmatrix}, \tag{3}$$

where $\mathbf{P}_N$ is an $N \times N$ permutation matrix that performs a reversal of the entries, that is, $[\mathbf{P}_N]_{n,n'} = \delta[n + n' - N - 1]$. Fourth, we add a zero postfix of length L to each block of the block sequence $\mathbf{x}_j[i]$, resulting into the $1 \times (KN + L)$ block sequence $\mathbf{u}_j[i]$: $\mathbf{u}_j[i] := \mathbf{x}_j[i]\mathbf{T}$, where $\mathbf{T}$ is the $KN \times (KN+L)$ zero postfix insertion matrix: $\mathbf{T} := [\mathbf{I}_{KN}, \mathbf{0}_{KN\times L}]$. Finally, the block sequence $\mathbf{u}_j[i]$ is parallel-to-serial converted into the space-time block coded multiuser chip sequence $u_j[n]$:

$$[u_j[i(KN+L)], \ldots, u_j[(i+1)(KN+L) - 1]] := \mathbf{u}_j[i], \tag{4}$$

which is transmitted at the jth transmit antenna with rate $1/T_c$ (the chip rate).

2.2. Channel model

Assuming the mth receive antenna is sampled at the chip rate, the received sequence at the mth receive antenna can be written as

$$y_m[n] = \sum_{j=1}^{2} \sum_{l=0}^{L} h_{m,j}[l] u_j[n-l] + e_m[n], \tag{5}$$

where $e_m[n]$ is the additive noise at the mth receive antenna and $h_{m,j}[l]$ is the channel from the jth transmit antenna to the mth receive antenna, including transmit and receive filters. We assume that $h_{m,j}[l]$ is FIR with order $L_{j,m}$ and that L is a known upper bound on $\max_{j,m}\{L_{j,m}\}$. Note that L was also chosen as the zero postfix length in Section 2.1.

2.3. Receiver design

A first option is to serial-to-parallel convert the received sequence $y_m[n]$ into the $1 \times (KN + L)$ received block sequence $\mathbf{y}_m[i]$:

$$\mathbf{y}_m[i] := [y_m[i(KN+L)], \ldots, y_m[(i+1)(KN+L) - 1]], \tag{6}$$

then to apply space-time decoding and Viterbi equalization as in [18], and finally, to perform simple despreading. This detection technique is overall ML, but leads to a very large computational complexity. That is why we pursue suboptimal detection by means of space-time chip equalization, which lowers the computational complexity significantly. Note that this suboptimal detection technique can also be applied to the STBC technique of [15] on its own, without combining it with the original single-antenna DS-CDMA downlink scheme.

We first introduce some new notation. Defining the $M \times 1$ vector

$$\mathbf{y}[n] := [y_1[n], \ldots, y_M[n]]^T, \tag{7}$$

we can write

$$\mathbf{y}[n] = \sum_{j=1}^{2} \sum_{l=0}^{L} \mathbf{h}_j[l] u_j[n-l] + \mathbf{e}[n], \tag{8}$$

where $\mathbf{e}[n]$ is similarly defined as $\mathbf{y}[n]$, and

$$\mathbf{h}_j[l] := [h_{1,j}[l], \ldots, h_{M,j}[l]]^T. \tag{9}$$

Further, defining the $(Q+1)M \times KN$ matrix

$$\mathbf{Y}[i] := \begin{bmatrix} \mathbf{y}[i(KN+L)] & \cdots & \mathbf{y}[i(KN+L)+KN-1] \\ \vdots & \vdots & \vdots \\ \mathbf{y}[i(KN+L)+Q] & \cdots & \mathbf{y}[i(KN+L)+KN-1+Q] \end{bmatrix}, \tag{10}$$

we can write

$$\mathbf{Y}[i] = \sum_{j=1}^{2} \boldsymbol{\mathcal{H}}_j \mathbf{U}_j[i] + \mathbf{E}[i], \tag{11}$$

where $\mathbf{E}[i]$ is similarly defined as $\mathbf{Y}[i]$,

$$\boldsymbol{\mathcal{H}}_j := \begin{bmatrix} \mathbf{h}_j[L] & \cdots & \mathbf{h}_j[0] & \mathbf{0}_{M\times 1} & \cdots & \mathbf{0}_{M\times 1} \\ \mathbf{0}_{M\times 1} & \mathbf{h}_j[L] & \cdots & \mathbf{h}_j[0] & \cdots & \mathbf{0}_{M\times 1} \\ \vdots & \vdots & \ddots & & \ddots & \vdots \\ \mathbf{0}_{M\times 1} & \mathbf{0}_{M\times 1} & \cdots & \mathbf{h}_j[L] & \cdots & \mathbf{h}_j[0] \end{bmatrix},$$

$$\mathbf{U}_j[i] := \begin{bmatrix} u_j[i(KN+L)-L] & \cdots & u_j[i(KN+L)-L+KN-1] \\ \vdots & \vdots & \vdots \\ u_j[i(KN+L)+Q] & \cdots & u_j[i(N+L)+Q+KN-1] \end{bmatrix}. \tag{12}$$

The parameter Q basically represents the order of the adopted space-time chip equalizer. This equalizer order Q is usually chosen to be close to the channel order L. For the sake of conciseness, we assume $Q = L$. However, the proposed results can easily be extended to other values of the equalizer order Q.

Choosing $Q = L$, it is clear from the zero postfix insertion that $\mathbf{U}_j[i]$ can be expressed as

$$\mathbf{U}_j[i] = \mathcal{T}(\mathbf{x}_j[i]) := \begin{bmatrix} \mathbf{x}_j[i]\mathbf{J}_{KN}^{(-L)} \\ \vdots \\ \mathbf{x}_j[i]\mathbf{J}_{KN}^{(L)} \end{bmatrix}, \tag{13}$$

with $\mathbf{J}_N^{(l)}$ the $N \times N$ shift matrix with $[\mathbf{J}_N^{(l)}]_{n,n'} = \delta[n-n'-l]$ (note that $\mathbf{J}_N^{(0)} = \mathbf{I}_N$).

To proceed, the straightforward way is to apply two space-time chip equalizers on $\mathbf{Y}[i]$ to recover $\mathbf{x}_1[i]$ and $\mathbf{x}_2[i]$, then to apply space-time decoding to recover $\mathbf{x}[2i]$ and $\mathbf{x}[2i+1]$, and finally, to perform simple despreading. Since this comes down to an equalization problem with two sources, we need three chip rate sampled receive antennas at each mobile station for a finite-length ZF solution to exist (for $J > 2$ transmit antennas, we need $J+1$ chip rate sampled receive antennas at each mobile station). However, we will show that the space-time chip equalization and space-time decoding operations can be swapped, which allows us to first apply space-time decoding on $\mathbf{Y}[2i]$ and $\mathbf{Y}[2i+1]$, then to apply two space-time chip equalizers to recover $\mathbf{x}[2i]$ and $\mathbf{x}[2i+1]$, and finally, to perform simple despreading. Since this comes down to two equalization problems with only one source, we need only two chip rate sampled receive antennas at each mobile station for a finite-length ZF solution to exist (even for $J > 2$ transmit antennas, we need only two chip rate sampled receive antennas at each mobile station). The latter option clearly has more degrees of freedom to tackle the equalization problem, and therefore leads to a better performance. This option is explained in more detail next.

2.3.1. Space-time decoding

Using (11) and (13), we can write $\mathbf{Y}[2i]$ and $\mathbf{Y}[2i+1]$ as

$$\begin{aligned} \mathbf{Y}[2i] &= \boldsymbol{\mathcal{H}}_1 \mathcal{T}(\mathbf{x}_1[2i]) + \boldsymbol{\mathcal{H}}_2 \mathcal{T}(\mathbf{x}_2[2i]) + \mathbf{E}[2i], \\ \mathbf{Y}[2i+1] &= \boldsymbol{\mathcal{H}}_1 \mathcal{T}(\mathbf{x}_1[2i+1]) + \boldsymbol{\mathcal{H}}_2 \mathcal{T}(\mathbf{x}_2[2i+1]) \\ &\quad + \mathbf{E}[2i+1]. \end{aligned} \tag{14}$$

Since $\mathbf{x}_1[2i+1] = -\mathbf{x}_2^*[2i]\mathbf{P}_{KN}$ (see (3)), we can derive from (13) that

$$\begin{aligned} \mathcal{T}(\mathbf{x}_1[2i+1]) &= \begin{bmatrix} \mathbf{x}_1[2i+1]\mathbf{J}_{KN}^{(-L)} \\ \vdots \\ \mathbf{x}_1[2i+1]\mathbf{J}_{KN}^{(L)} \end{bmatrix} \\ &= -\begin{bmatrix} \mathbf{x}_2^*[2i]\mathbf{P}_{KN}\mathbf{J}_{KN}^{(-L)} \\ \vdots \\ \mathbf{x}_2^*[2i]\mathbf{P}_{KN}\mathbf{J}_{KN}^{(L)} \end{bmatrix} \\ &= -\begin{bmatrix} \mathbf{x}_2^*[2i]\mathbf{J}_{KN}^{(L)} \\ \vdots \\ \mathbf{x}_2^*[2i]\mathbf{J}_{KN}^{(-L)} \end{bmatrix} \mathbf{P}_{KN} \\ &= -\mathbf{P}_{2L+1}\begin{bmatrix} \mathbf{x}_2^*[2i]\mathbf{J}_{KN}^{(-L)} \\ \mathbf{x}_2^*[2i]\mathbf{J}_{KN}^{(L)} \end{bmatrix} \mathbf{P}_{KN} \\ &= -\mathbf{P}_{2L+1}\mathcal{T}^*(\mathbf{x}_2[2i])\mathbf{P}_{KN}. \end{aligned} \tag{15}$$

Similarly, since $\mathbf{x}_2[2i+1] = \mathbf{x}_1^*[2i]\mathbf{P}_{KN}$ (see (3)), we can derive from (13) that

$$\mathcal{T}(\mathbf{x}_2[2i+1]) = \mathbf{P}_{2L+1}\mathcal{T}^*(\mathbf{x}_1[2i])\mathbf{P}_{KN}. \tag{16}$$

Conjugating $\mathbf{Y}[2i+1]$ and multiplying it to the right-hand side with $\mathbf{P}_{KN}$, we then arrive at

$$\begin{aligned} &\mathbf{Y}^*[2i+1]\mathbf{P}_{KN} \\ &= \boldsymbol{\mathcal{H}}_1^*\mathcal{T}^*(\mathbf{x}_1[2i+1])\mathbf{P}_{KN} + \boldsymbol{\mathcal{H}}_2^*\mathcal{T}^*(\mathbf{x}_2[2i+1])\mathbf{P}_{KN} \\ &\quad + \mathbf{E}^*[2i+1]\mathbf{P}_{KN} \\ &= -\boldsymbol{\mathcal{H}}_1^*\mathbf{P}_{2L+1}\mathcal{T}(\mathbf{x}_2[2i]) + \boldsymbol{\mathcal{H}}_2^*\mathbf{P}_{2L+1}\mathcal{T}(\mathbf{x}_1[2i]) \\ &\quad + \mathbf{E}^*[2i+1]\mathbf{P}_{KN}, \end{aligned} \tag{17}$$

where the second equality is due to (15) and (16). Stacking $\mathbf{Y}[2i]$ and $\mathbf{Y}^*[2i+1]\mathbf{P}_{KN}$:

$$\bar{\mathbf{Y}}[i] := \begin{bmatrix} \mathbf{Y}[2i] \\ \mathbf{Y}^*[2i+1]\mathbf{P}_{KN} \end{bmatrix}, \tag{18}$$

and using the fact that $\mathbf{x}_1[2i] = \mathbf{x}[2i]$ and $\mathbf{x}_2[2i] = \mathbf{x}[2i+1]$ (see (3)), we finally obtain

$$\bar{\mathbf{Y}}[i] = \boldsymbol{\mathcal{H}}\bar{\mathbf{X}}[i] + \bar{\mathbf{E}}[i], \tag{19}$$

where $\bar{\mathbf{E}}[i]$ is similarly defined as $\bar{\mathbf{Y}}[i]$,

$$\begin{aligned} \boldsymbol{\mathcal{H}} &:= \begin{bmatrix} \boldsymbol{\mathcal{H}}_1 & \boldsymbol{\mathcal{H}}_2 \\ \boldsymbol{\mathcal{H}}_2^* \mathbf{P}_{2L+1} & -\boldsymbol{\mathcal{H}}_1^* \mathbf{P}_{2L+1} \end{bmatrix}, \\ \bar{\mathbf{X}}[i] &:= \begin{bmatrix} \mathcal{T}(\mathbf{x}[2i]) \\ \mathcal{T}(\mathbf{x}[2i+1]) \end{bmatrix}. \end{aligned} \tag{20}$$

2.3.2. Space-time chip equalization

We now apply two space-time chip equalizers on $\bar{\mathbf{Y}}[i]$: $\mathbf{f}_e$ and $\mathbf{f}_o$. The $1 \times 2(L+1)M$ space-time chip equalizer $\mathbf{f}_e$ is designed to extract the even multiuser chip block $\mathbf{x}[2i]$, whereas the $1 \times 2(L+1)M$ space-time chip equalizer $\mathbf{f}_o$ is designed to extract the odd multiuser chip block $\mathbf{x}[2i+1]$:

$$\hat{\mathbf{x}}[2i] = \mathbf{f}_e\bar{\mathbf{Y}}[i], \qquad \hat{\mathbf{x}}[2i+1] = \mathbf{f}_o\bar{\mathbf{Y}}[i]. \tag{21}$$

Note that $\mathbf{x}[2i]$ and $\mathbf{x}[2i+1]$ are two distinct rows of $\bar{\mathbf{X}}[i]$.

A first possibility is to apply two ZF space-time chip equalizers, completely eliminating the interchip interference (ICI) at the expense of potentially excessive noise enhancement:

$$\begin{aligned} \mathbf{f}_e &= \mathbf{i}_e\left(\boldsymbol{\mathcal{H}}^H\mathbf{R}_e^{-1}\boldsymbol{\mathcal{H}}\right)^{-1}\boldsymbol{\mathcal{H}}^H\mathbf{R}_e^{-1}, \\ \mathbf{f}_o &= \mathbf{i}_o\left(\boldsymbol{\mathcal{H}}^H\mathbf{R}_e^{-1}\boldsymbol{\mathcal{H}}\right)^{-1}\boldsymbol{\mathcal{H}}^H\mathbf{R}_e^{-1}, \end{aligned} \tag{22}$$

where $\mathbf{i}_e$ is a $1 \times (4L+2)$ unit vector with a one in the $(L+1)$th position, $\mathbf{i}_o$ is a $1 \times (4L+2)$ unit vector with a one in the $(3L+2)$th position, and $\mathbf{R}_e := 1/(KN)\mathcal{E}\{\bar{\mathbf{E}}[i]\bar{\mathbf{E}}^H[i]\}$. A second possibility is to apply two minimum mean-squared error (MMSE) space-time chip equalizers, balancing ICI elimination with noise enhancement:

$$\begin{aligned} \mathbf{f}_e &= \mathbf{i}_e\left(\boldsymbol{\mathcal{H}}^H\mathbf{R}_e^{-1}\boldsymbol{\mathcal{H}} + \mathbf{R}_x^{-1}\right)^{-1}\boldsymbol{\mathcal{H}}^H\mathbf{R}_e^{-1}, \\ \mathbf{f}_o &= \mathbf{i}_o\left(\boldsymbol{\mathcal{H}}^H\mathbf{R}_e^{-1}\boldsymbol{\mathcal{H}} + \mathbf{R}_x^{-1}\right)^{-1}\boldsymbol{\mathcal{H}}^H\mathbf{R}_e^{-1}, \end{aligned} \tag{23}$$

where $\mathbf{R}_x := 1/(KN)\mathcal{E}\{\bar{\mathbf{X}}[i]\bar{\mathbf{X}}^H[i]\}$.

Assuming the additive noise sequences $\{e_m[n]\}_{m=1}^M$ are mutually uncorrelated and white with variance σ_e^2, we can write $\mathbf{R}_e = \sigma_e^2\mathbf{I}_{2(L+1)M}$. Furthermore, assuming the data symbol sequences $\{s_u[n]\}_{u=1}^U$ are mutually uncorrelated and white with variance σ_s^2, the original multiuser chip sequence $x[n]$ is white with variance $\sigma_x^2 = \sigma_s^2 J/N$ (justified by the long scrambling code), and we can write $\mathbf{R}_x = \sigma_x^2\,\mathrm{diag}\{[\mathbf{r}_x, \mathbf{r}_x]\} = \sigma_s^2 J/N\,\mathrm{diag}\{[\mathbf{r}_x, \mathbf{r}_x]\}$, where $\mathbf{r}_x = [(KN-L)/(KN), \ldots, (KN-1)/(KN), 1, (KN-1)/(KN), \ldots, (KN-L)/(KN)]$.

2.3.3. Despreading

We define the $1 \times KU$ multiuser data symbol block $\mathbf{s}[i]$ as

$$\mathbf{s}[i] := [\mathbf{s}_1[i], \ldots, \mathbf{s}_U[i]], \tag{24}$$

where $\mathbf{s}_u[i]$ is the uth user's $1 \times K$ data symbol block given by

$$\mathbf{s}_u[i] := [s_u[iK], \ldots, s_u[(i+1)K-1]]. \tag{25}$$

Note that the $1 \times K$ pilot symbol block $\mathbf{s}_p[i]$ is similarly defined as $\mathbf{s}_u[i]$. We further define the multiuser code matrix $\mathbf{C}[i]$ as

$$\mathbf{C}[i] := [\mathbf{C}_1[i]^T, \ldots, \mathbf{C}_U[i]^T]^T, \tag{26}$$

where $\mathbf{C}_u[i]$ is the uth user's code matrix given by

$$\mathbf{C}_u[i] := \begin{bmatrix} \mathbf{c}_u[iK] & & \\ & \ddots & \\ & & \mathbf{c}_u[(i+1)K-1] \end{bmatrix}, \tag{27}$$

with $\mathbf{c}_u[k] := [c_u[kN], \ldots, c_u[(k+1)N-1]]$. Note that the pilot code matrix $\mathbf{C}_p[i]$ is similarly defined as $\mathbf{C}_u[i]$. It is then clear from (1) that the multiuser chip block $\mathbf{x}[i]$ can be expressed as

$$\begin{aligned} \mathbf{x}[i] &= \sum_{u=1}^{U} \mathbf{s}_u[i]\mathbf{C}_u[i] + \mathbf{s}_p[i]\mathbf{C}_p[i] \\ &= \mathbf{s}[i]\mathbf{C}[i] + \mathbf{s}_p[i]\mathbf{C}_p[i]. \end{aligned} \tag{28}$$

Hence, by despreading the multiuser chip block $\mathbf{x}[i]$ with the uth user's code matrix $\mathbf{C}_u[i]$, we obtain

$$\mathbf{s}_u[i] = \mathbf{x}[i]\mathbf{C}_u^H[i] \tag{29}$$

because $\mathbf{C}_p[i]\mathbf{C}_u^H[i] = \mathbf{0}_{K\times K}$, $\mathbf{C}_{u'}[i]\mathbf{C}_u^H[i] = \mathbf{0}_{K\times K}$ for $u \neq u'$, and $\mathbf{C}_u[i]\mathbf{C}_u^H[i] = \mathbf{I}_K$. Therefore, once $\mathbf{x}[i]$ has been estimated, we can find an estimate for $\mathbf{s}_u[i]$ by simple despreading:

$$\hat{\mathbf{s}}_u[i] = \hat{\mathbf{x}}[i]\mathbf{C}_u^H[i]. \tag{30}$$

Plugging (30) into (21), we thus obtain

$$\begin{aligned} \hat{\mathbf{s}}_u[2i] &= \mathbf{f}_e\bar{\mathbf{Y}}[i]\mathbf{C}_u^H[2i], \\ \hat{\mathbf{s}}_u[2i+1] &= \mathbf{f}_o\bar{\mathbf{Y}}[i]\mathbf{C}_u^H[2i+1]. \end{aligned} \tag{31}$$

From these equations, it is also clear that the order of equalization and despreading can be reversed. In other words, we can first despread $\bar{\mathbf{Y}}[i]$ with $\mathbf{C}_u[2i]$ and $\mathbf{C}_u[2i+1]$, and then perform space-time chip equalization on both results.

3. PRACTICAL SPACE-TIME CHIP EQUALIZER DESIGN

In this section, we focus on practical space-time chip equalizer design. In [20, 21], we have developed two pilot-based space-time chip equalizer design methods for the original single-antenna DS-CDMA downlink scheme: a *training-based* method and a *semiblind* method. In this section, these two methods are appropriately modified and applied to the

proposed space-time coded DS-CDMA downlink scheme. We consider a burst of $2I$ data symbol blocks.

The goal of the *training-based* method is to compute the uth user's even and odd data symbol blocks $\{\mathbf{s}_u[2i]\}_{i=1}^{I}$ and $\{\mathbf{s}_u[2i+1]\}_{i=1}^{I}$ from $\{\bar{\mathbf{Y}}[i]\}_{i=1}^{I}$, based on the even and odd pilot symbol blocks $\{\mathbf{s}_p[2i]\}_{i=1}^{I}$ and $\{\mathbf{s}_p[2i+1]\}_{i=1}^{I}$, the even and odd pilot code matrices $\{\mathbf{C}_p[2i]\}_{i=1}^{I}$ and $\{\mathbf{C}_p[2i+1]\}_{i=1}^{I}$, and the uth user's even and odd code matrices $\{\mathbf{C}_u[2i]\}_{i=1}^{I}$ and $\{\mathbf{C}_u[2i+1]\}_{i=1}^{I}$.

The goal of the *semiblind* method is to compute the uth user's even and odd data symbol blocks $\{\mathbf{s}_u[2i]\}_{i=1}^{I}$ and $\{\mathbf{s}_u[2i+1]\}_{i=1}^{I}$ from $\{\bar{\mathbf{Y}}[i]\}_{i=1}^{I}$, based on the even and odd pilot symbol blocks $\{\mathbf{s}_p[2i]\}_{i=1}^{I}$ and $\{\mathbf{s}_p[2i+1]\}_{i=1}^{I}$, the even and odd pilot code matrices $\{\mathbf{C}_p[2i]\}_{i=1}^{I}$ and $\{\mathbf{C}_p[2i+1]\}_{i=1}^{I}$, and the even and odd multiuser code matrices $\{\mathbf{C}[2i]\}_{i=1}^{I}$ and $\{\mathbf{C}[2i+1]\}_{i=1}^{I}$. Note that the semiblind method requires the knowledge of the active codes. This knowledge can be obtained by means of a limited feedback from the base station to the mobile station (only the indices of the active codes have to be fed back). However, this knowledge can also be obtained by first adopting the training-based method to design a space-time chip equalizer, and then comparing for each code the energy obtained after equalization and despreading with some threshold in order to decide whether this code is active or not.

For the sake of conciseness, we will only focus on block implementations. These block implementations might look rather complex, but they form the basis for practical low-complexity adaptive implementations, which can be derived in a similar fashion as done in [20, 21].

For the sake of simplicity, we make the following assumptions:

(A1) the matrix $\boldsymbol{\mathcal{H}}$ has full column rank $4L+2$;

(A2) the matrices $\bar{\mathbf{X}}[2i]$ and $\bar{\mathbf{X}}[2i+1]$ have full row rank $4L+2$ for all $i \in \{1,\ldots,I\}$.

The first assumption requires that $2(L+1)(M-1) \geq 2L$, which means we need only $M \geq 2$ receive antennas at each mobile station (even for $J > 2$ transmit antennas, we need only $M \geq 2$ receive antennas at each mobile station). The second assumption requires that $4L+2 \leq KN$. Note that these assumptions are not really necessary for the proposed methods to work. The only true requirement is that $\mathbf{x}[2i]$ and $\mathbf{x}[2i+1]$ belong to the row space of $\bar{\mathbf{Y}}[i]$ for all $i \in \{1,\ldots,I\}$. Assumptions (A1) and (A2) are sufficient but not necessary conditions for this. However, they considerably simplify the analysis.

Assume no noise is present. Because of assumption (A1), the row space of $\bar{\mathbf{Y}}[i]$ equals the row space of $\bar{\mathbf{X}}[i]$. Hence, there exist two $1 \times 2(L+1)M$ space-time chip equalizers $\mathbf{f}_e$ and $\mathbf{f}_o$, for which

$$\begin{aligned} \mathbf{f}_e\bar{\mathbf{Y}}[i] - \mathbf{x}[2i] &= \mathbf{0}_{1\times KN}, \\ \mathbf{f}_o\bar{\mathbf{Y}}[i] - \mathbf{x}[2i+1] &= \mathbf{0}_{1\times KN}. \end{aligned} \tag{32}$$

Because of assumption (A2), these two space-time chip equalizers $\mathbf{f}_e$ and $\mathbf{f}_o$ are ZF. By using (28), we then obtain

$$\begin{aligned} \mathbf{f}_e\bar{\mathbf{Y}}[i] - \mathbf{s}[2i]\mathbf{C}[2i] - \mathbf{s}_p[2i]\mathbf{C}_p[2i] &= \mathbf{0}_{1\times KN}, \\ \mathbf{f}_o\bar{\mathbf{Y}}[i] - \mathbf{s}[2i+1]\mathbf{C}[2i+1] - \mathbf{s}_p[2i+1]\mathbf{C}_p[2i+1] &= \mathbf{0}_{1\times KN}. \end{aligned} \tag{33}$$

3.1. Training-based method

By despreading (33) with the even and odd pilot code matrices $\mathbf{C}_p[2i]$ and $\mathbf{C}_p[2i+1]$, we obtain

$$\begin{aligned} \mathbf{f}_e\bar{\mathbf{Y}}[i]\mathbf{C}_p^H[2i] - \mathbf{s}_p[2i] &= \mathbf{0}_{1\times K}, \\ \mathbf{f}_o\bar{\mathbf{Y}}[i]\mathbf{C}_p^H[2i+1] - \mathbf{s}_p[2i+1] &= \mathbf{0}_{1\times K} \end{aligned} \tag{34}$$

because $\mathbf{C}[i]\mathbf{C}_p^H[i] = \mathbf{0}_{K\times K}$ and $\mathbf{C}_p[i]\mathbf{C}_p^H[i] = \mathbf{I}_K$. The training-based method solves (34) for $\mathbf{f}_e$ and $\mathbf{f}_o$ for all $i \in \{1,\ldots,I\}$. In the noisy case, this leads to the following least squares (LS) problems:

$$\begin{aligned} &\min_{\mathbf{f}_e}\left\{\sum_{i=1}^{I}\left\|\mathbf{f}_e\bar{\mathbf{Y}}[i]\mathbf{C}_p^H[2i] - \mathbf{s}_p[2i]\right\|^2\right\}, \\ &\min_{\mathbf{f}_o}\left\{\sum_{i=1}^{I}\left\|\mathbf{f}_o\bar{\mathbf{Y}}[i]\mathbf{C}_p^H[2i+1] - \mathbf{s}_p[2i+1]\right\|^2\right\}, \end{aligned} \tag{35}$$

which can be interpreted as follows. The space-time decoded output matrix $\bar{\mathbf{Y}}[i]$ is first equalized with the even and odd space-time chip equalizers $\mathbf{f}_e$ and $\mathbf{f}_o$, and then despread with the even and odd pilot code matrices $\mathbf{C}_p[2i]$ and $\mathbf{C}_p[2i+1]$. The resulting even and odd vectors $\mathbf{f}_e\bar{\mathbf{Y}}[i]\mathbf{C}_p^H[2i]$ and $\mathbf{f}_o\bar{\mathbf{Y}}[i]\mathbf{C}_p^H[2i+1]$ should then be as close as possible in an LS sense to the even and odd pilot symbol blocks $\mathbf{s}_p[2i]$ and $\mathbf{s}_p[2i+1]$ for all $i \in \{1,\ldots,I\}$. The solutions of (35) can be written as

$$\begin{aligned} \hat{\mathbf{f}}_e &= \left(\sum_{i=1}^{I}\mathbf{s}_p[2i]\mathbf{C}_p[2i]\bar{\mathbf{Y}}^H[i]\right) \\ &\quad\times\left(\sum_{i=1}^{I}\bar{\mathbf{Y}}[i]\mathbf{C}_p^H[2i]\mathbf{C}_p[2i]\bar{\mathbf{Y}}^H[i]\right)^{-1}, \\ \hat{\mathbf{f}}_o &= \left(\sum_{i=1}^{I}\mathbf{s}_p[2i+1]\mathbf{C}_p[2i+1]\bar{\mathbf{Y}}^H[i]\right) \\ &\quad\times\left(\sum_{i=1}^{I}\bar{\mathbf{Y}}[i]\mathbf{C}_p^H[2i+1]\mathbf{C}_p[2i+1]\bar{\mathbf{Y}}^H[i]\right)^{-1}. \end{aligned} \tag{36}$$

The obtained space-time chip equalizers $\hat{\mathbf{f}}_e$ and $\hat{\mathbf{f}}_o$ are subsequently used to estimate the uth user's even and odd data symbol blocks $\mathbf{s}_u[2i]$ and $\mathbf{s}_u[2i+1]$ for all $i \in \{1,\ldots,I\}$:

$$\begin{aligned} \hat{\mathbf{s}}_u[2i] &= \hat{\mathbf{f}}_e\bar{\mathbf{Y}}[i]\mathbf{C}_u^H[2i], \\ \hat{\mathbf{s}}_u[2i+1] &= \hat{\mathbf{f}}_o\bar{\mathbf{Y}}[i]\mathbf{C}_u^H[2i+1]. \end{aligned} \tag{37}$$

These soft estimates are fed into a decision device that determines the nearest constellation point.

3.2. Semiblind method

The semiblind method directly solves (33) for $(\mathbf{f}_e, \mathbf{s}[2i])$ and $(\mathbf{f}_o, \mathbf{s}[2i+1])$ for all $i \in \{1,\ldots,I\}$. In the noisy case, this leads to the following LS problems:

$$\begin{aligned}
&\min_{(\mathbf{f}_e,\{\mathbf{s}[2i]\}_{i=1}^{I})} \left\{ \sum_{i=1}^{I} \left|\left|\mathbf{f}_e\bar{\mathbf{Y}}[i] - \mathbf{s}[2i]\mathbf{C}[2i] - \mathbf{s}_p[2i]\mathbf{C}_p[2i]\right|\right|^2 \right\},\\
&\min_{(\mathbf{f}_o,\{\mathbf{s}[2i+1]\}_{i=1}^{I})} \left\{ \sum_{i=1}^{I} \left|\left|\mathbf{f}_o\bar{\mathbf{Y}}[i] - \mathbf{s}[2i+1]\mathbf{C}[2i+1] \right.\right.\right.\\
&\qquad\qquad \left.\left.\left. - \mathbf{s}_p[2i+1]\mathbf{C}_p[2i+1]\right|\right|^2 \right\}.
\end{aligned} \tag{38}$$

Since we are interested in $\mathbf{f}_e$ and $\mathbf{f}_o$, we can first solve (38) for $\mathbf{s}[2i]$ and $\mathbf{s}[2i+1]$ for all $i \in \{1,\ldots,I\}$, which results into

$$\begin{aligned}
\hat{\mathbf{s}}[2i] &= \mathbf{f}_e\bar{\mathbf{Y}}[i]\mathbf{C}^H[2i],\\
\hat{\mathbf{s}}[2i+1] &= \mathbf{f}_o\bar{\mathbf{Y}}[i]\mathbf{C}^H[2i+1]
\end{aligned} \tag{39}$$

because $\mathbf{C}[i]\mathbf{C}_p^H[i] = \mathbf{0}_{K\times K}$ and $\mathbf{C}_p[i]\mathbf{C}_p^H[i] = \mathbf{I}_K$. Substituting $\hat{\mathbf{s}}[2i]$ and $\hat{\mathbf{s}}[2i+1]$ in (38) leads to the following LS problems:

$$\begin{aligned}
&\min_{\mathbf{f}_e} \left\{ \sum_{i=1}^{I} \left|\left|\mathbf{f}_e\bar{\mathbf{Y}}[i]\left(\mathbf{I}_{KN} - \mathbf{C}^H[2i]\mathbf{C}[2i]\right) - \mathbf{s}_p[2i]\mathbf{C}_p[2i]\right|\right|^2 \right\},\\
&\min_{\mathbf{f}_o} \left\{ \sum_{i=1}^{I} \left|\left|\mathbf{f}_o\bar{\mathbf{Y}}[i]\left(\mathbf{I}_{KN} - \mathbf{C}^H[2i+1]\mathbf{C}[2i+1]\right) \right.\right.\right.\\
&\qquad\qquad \left.\left.\left. - \mathbf{s}_p[2i+1]\mathbf{C}_p[2i+1]\right|\right|^2 \right\},
\end{aligned} \tag{40}$$

which can be interpreted as follows. The space-time decoded output matrix $\bar{\mathbf{Y}}[i]$ is first equalized with the even and odd space-time chip equalizers $\mathbf{f}_e$ and $\mathbf{f}_o$ and then projected on the orthogonal complement of the subspace spanned by the even and odd multiuser code matrices $\mathbf{C}[2i]$ and $\mathbf{C}[2i+1]$. The resulting even and odd vectors $\mathbf{f}_e\bar{\mathbf{Y}}[i](\mathbf{I}_{KN} - \mathbf{C}^H[2i]\mathbf{C}[2i])$ and $\mathbf{f}_o\bar{\mathbf{Y}}[i](\mathbf{I}_{KN} - \mathbf{C}^H[2i+1]\mathbf{C}[2i+1])$ should then be as close as possible in an LS sense to the even and odd pilot chip blocks $\mathbf{s}_p[2i]\mathbf{C}_p[2i]$ and $\mathbf{s}_p[2i+1]\mathbf{C}_p[2i+1]$ for all $i \in \{1,\ldots,I\}$. The solutions of (40) can be written as

$$\begin{aligned}
\hat{\mathbf{f}}_e &= \left(\sum_{i=1}^{I} \mathbf{s}_p[2i]\mathbf{C}_p[2i]\bar{\mathbf{Y}}^H[i]\right)\\
&\quad\times \left(\sum_{i=1}^{I} \bar{\mathbf{Y}}[i]\left(\mathbf{I}_{KN} - \mathbf{C}^H[2i]\mathbf{C}[2i]\right)\bar{\mathbf{Y}}^H[i]\right)^{-1},\\
\hat{\mathbf{f}}_o &= \left(\sum_{i=1}^{I} \mathbf{s}_p[2i+1]\mathbf{C}_p[2i+1]\bar{\mathbf{Y}}^H[i]\right)\\
&\quad\times \left(\sum_{i=1}^{I} \bar{\mathbf{Y}}[i]\left(\mathbf{I}_{KN} - \mathbf{C}^H[2i+1]\mathbf{C}[2i+1]\right)\bar{\mathbf{Y}}^H[i]\right)^{-1}.
\end{aligned} \tag{41}$$

The obtained space-time chip equalizers $\hat{\mathbf{f}}_e$ and $\hat{\mathbf{f}}_o$ are subsequently used to estimate the uth user's even and odd data symbol blocks $\mathbf{s}_u[2i]$ and $\mathbf{s}_u[2i+1]$ for all $i \in \{1,\ldots,I\}$:

$$\begin{aligned}
\hat{\mathbf{s}}_u[2i] &= \hat{\mathbf{f}}_e\bar{\mathbf{Y}}[i]\mathbf{C}_u^H[2i],\\
\hat{\mathbf{s}}_u[2i+1] &= \hat{\mathbf{f}}_o\bar{\mathbf{Y}}[i]\mathbf{C}_u^H[2i+1].
\end{aligned} \tag{42}$$

These soft estimates are fed into a decision device that determines the nearest constellation point.

With some algebraic manipulations, it is easy to prove that (40) is equivalent to

$$\begin{aligned}
&\min_{\mathbf{f}_e} \left\{ \sum_{i=1}^{I} \left|\left|\mathbf{f}_e\bar{\mathbf{Y}}[i]\mathbf{C}_p^H[2i] - \mathbf{s}_p[2i]\right|\right|^2 \right.\\
&\qquad \left. + \left|\left|\mathbf{f}_e\bar{\mathbf{Y}}[i]\left(\mathbf{I}_{KN} - \mathbf{C}^H[2i]\mathbf{C}[2i] - \mathbf{C}_p^H[2i]\mathbf{C}_p[2i]\right)\right|\right|^2 \right\},\\
&\min_{\mathbf{f}_o} \left\{ \sum_{i=1}^{I} \left|\left|\mathbf{f}_o\bar{\mathbf{Y}}[i]\mathbf{C}_p^H[2i+1] - \mathbf{s}_p[2i+1]\right|\right|^2 \right.\\
&\qquad + \left|\left|\mathbf{f}_o\bar{\mathbf{Y}}[i]\left(\mathbf{I}_{KN} - \mathbf{C}^H[2i+1]\mathbf{C}[2i+1] \right.\right.\right.\\
&\qquad\qquad \left.\left.\left. - \mathbf{C}_p^H[2i+1]\mathbf{C}_p[2i+1]\right)\right|\right|^2 \right\}.
\end{aligned} \tag{43}$$

This shows that (40) naturally decouples into a training-based part and a blind part (hence the name *semiblind*). The training-based part corresponds to (35). The blind part can be interpreted as follows. The space-time decoded output matrix $\bar{\mathbf{Y}}[i]$ is first equalized with the even and odd space-time chip equalizers $\mathbf{f}_e$ and $\mathbf{f}_o$ and then projected on the orthogonal complement of the subspace spanned by the even and odd multiuser code matrices $\mathbf{C}[2i]$ and $\mathbf{C}[2i+1]$ and the even and odd pilot code matrices $\mathbf{C}_p[2i]$ and $\mathbf{C}_p[2i+1]$. The resulting even and odd vectors $\mathbf{f}_e\bar{\mathbf{Y}}[i](\mathbf{I}_{KN} - \mathbf{C}^H[2i]\mathbf{C}[2i] - \mathbf{C}_p^H[2i]\mathbf{C}_p[2i])$ and $\mathbf{f}_o\bar{\mathbf{Y}}[i](\mathbf{I}_{KN} - \mathbf{C}^H[2i+1]\mathbf{C}[2i+1] - \mathbf{C}_p^H[2i+1]\mathbf{C}_p[2i+1])$ should then be as small as possible in an LS sense for all $i \in \{1,\ldots,I\}$. Note that when the user load increases, the orthogonal complement of the subspace spanned by the even and odd multiuser code matrices $\mathbf{C}[2i]$ and $\mathbf{C}[2i+1]$ and the even and odd pilot code matrices $\mathbf{C}_p[2i]$ and $\mathbf{C}_p[2i+1]$ decreases in dimension. As a result, the information that the blind part contributes to the training-based part diminishes, and the semiblind method converges to the training-based method. In the extreme case when the system is fully loaded, that is, $N = U - 1$, the orthogonal complement of the subspace spanned by the even and odd multiuser code matrices $\mathbf{C}[2i]$ and $\mathbf{C}[2i+1]$ and the even and odd pilot code matrices $\mathbf{C}_p[2i]$ and $\mathbf{C}_p[2i+1]$ is empty, that is, $\mathbf{I}_{KN} - \mathbf{C}^H[2i]\mathbf{C}[2i] - \mathbf{C}_p^H[2i]\mathbf{C}_p[2i] = \mathbf{0}_{KN\times KN}$ and $\mathbf{I}_{KN} - \mathbf{C}^H[2i+1]\mathbf{C}[2i+1] - \mathbf{C}_p^H[2i+1]\mathbf{C}_p[2i+1] = \mathbf{0}_{KN\times KN}$. Hence, the blind part does not contribute any additional information to the training-based part, and the semiblind method reduces to the training based method, that is, (43) reduces to (35).

4. SIMULATION RESULTS

In this section, we compare the proposed space-time chip equalizer for the proposed space-time coded downlink CDMA transmission scheme with the space-time RAKE receiver for the space-time spreading scheme, which encompasses the space-time coded downlink CDMA transmission schemes that have been proposed for the UMTS and IS-2000 W-CDMA standards [16]. We do not consider channel codes when comparing the above transceivers. Otherwise, it will not be very clear whether a performance gain is due to the transceiver or the channel code. Moreover, the influence of channel codes on performance has been studied extensively in literature. In W-CDMA, the target coded BER typically is 10^{-6}, which boils down to an uncoded BER of 10^{-2} with a convolutional code of rate 1/2, constraint length 7, and soft decision Viterbi [22]. Therefore, we compare the different transceivers at an uncoded BER of 10^{-2} in the sequel.

We consider a downlink CDMA system with a spreading factor of $N = 32$, $J = 2$ transmit antennas at the base station, and $M = 2$ receive antennas at each mobile station. We assume that all channels are independent. We further assume that each channel $h_{j,m}[n]$ is FIR with order $L_{j,m} = 3$ and has independent Rayleigh fading channel taps of equal variance σ_h^2. Note that the bandwidth efficiency of the proposed space-time coded downlink CDMA transmission scheme is $\epsilon_1 = KU/(KN + L)$, whereas the bandwidth efficiency of the space-time spreading scheme is $\epsilon_2 = U/N$. Hence, in order to make a fair comparison between the two systems, their spectral efficiencies should be comparable. We therefore take $K = 5$ and $L = 3$ for the proposed space-time coded downlink CDMA transmission scheme, which results into $\epsilon_1/\epsilon_2 \approx 0.98$. We assume QPSK modulated data symbols, and define the signal-to-noise ratio (SNR) as the received bit energy over the noise power:

$$\begin{aligned}\text{SNR} &= \frac{\sigma_s^2/2 \sum_{j=1}^{2} \sum_{l=0}^{L} \mathcal{E}\left\{||\mathbf{h}_j[l]||^2\right\}}{\sigma_e^2} \\ &= \frac{2(L+1)\sigma_s^2\sigma_h^2}{\sigma_e^2}.\end{aligned} \tag{44}$$

Two test cases are investigated.

Test case 1

We first assume that the pilot enables us to obtain perfect channel knowledge at the receiver. We then compare the proposed MMSE space-time chip equalizer for the proposed space-time coded downlink CDMA transmission scheme with the MMSE space-time RAKE receiver for the space-time spreading scheme (see [23, 24]), which is different from the matched space-time RAKE receiver for the space-time spreading scheme (see [16]) because it uses an MMSE filter instead of a matched filter to combine the finger outputs. It has been shown in [23, 24] that for the space-time spreading scheme, the MMSE space-time RAKE receiver significantly outperforms the matched space-time RAKE receiver. Figures 2, 3, and 4 compare the performance of the two transceivers

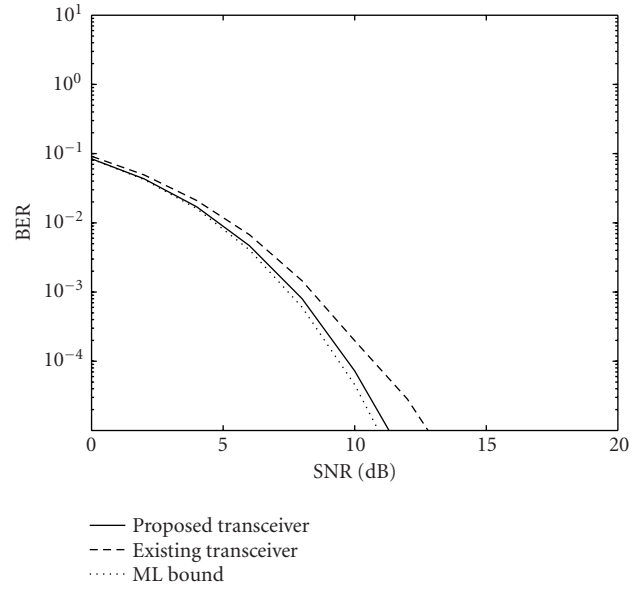


Figure 2: Performance comparison for $U = 1$.

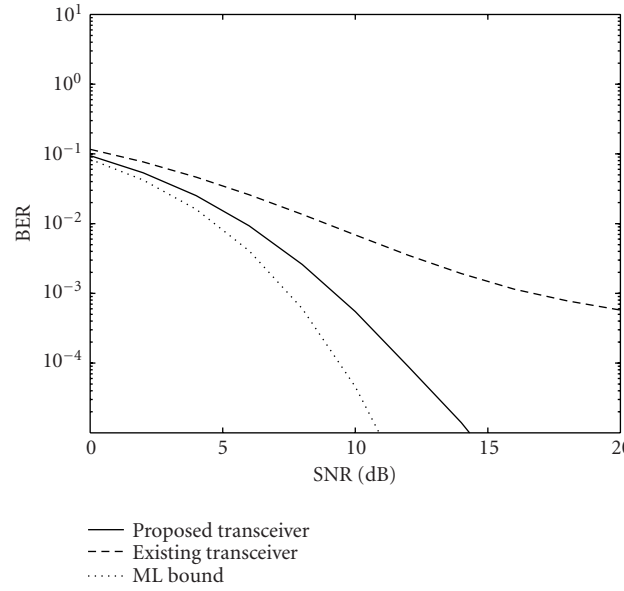


Figure 3: Performance comparison for $U = 15$.

for $U = 1$, $U = 15$, and $U = 31$ users, respectively. The performance results are averaged over 1000 random channel realizations, where for each channel realization, we consider 10 random data and noise realizations corresponding to $I = 10$ (100 data symbols per user). Also shown is the theoretical performance of $\sum_{j,m}(L_{j,m} + 1) = 16$-fold diversity over Rayleigh fading channels [22].

First of all, we see that the proposed transceiver comes close to extracting the maximum diversity at low-to-medium

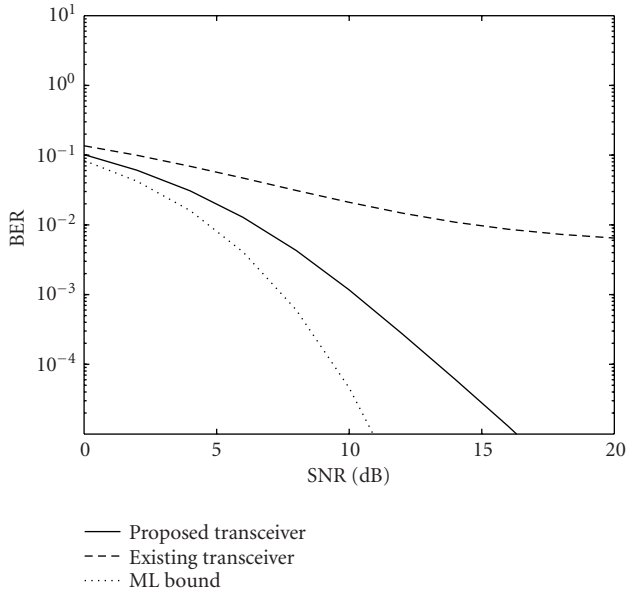


Figure 4: Performance comparison for $U = 31$.

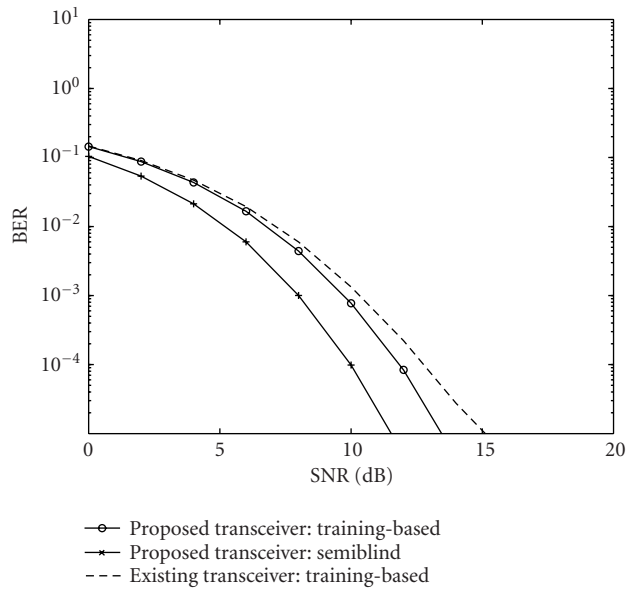


Figure 5: Performance of pilot-based methods for $U = 1$.

user loads. More specifically, at a BER of 10^{-2}, the proposed transceiver incurs a 0.1, 1, and 1.8 dB loss compared to the theoretical ML bound for $U = 1$, $U = 15$, and $U = 31$ users, respectively. The existing transceiver, on the other hand, performs poorly at medium-to-high user loads. At a BER of 10^{-2}, it incurs a 0.5, 3, and 8.2 dB performance loss compared to the proposed transceiver for $U = 1$, $U = 15$, and $U = 31$ users, respectively. The existing transceiver is not capable of completely suppressing the MUI at high SNR. This results into a flooring of the BER at high SNR. Note that the flooring level increases with the number of users U.

Test case 2

We now investigate the performance of the pilot-based methods. Note that for the space-time spreading scheme, it is easy to derive a training-based method to estimate the combining filter of the space-time RAKE receiver based on the knowledge of the pilot. The performance results are again averaged over 1000 random channel realizations, where for each channel realization, we consider 10 random data and noise realizations corresponding to $I = 10$ (100 data symbols per user). Figures 5, 6, and 7 compare the performance of the different methods for $U = 1$, $U = 15$, and $U = 31$ users, respectively.

First of all, we observe that the difference between the training-based method and the semiblind method for the proposed transceiver decreases with an increasing user load, as indicated in Section 3.2. Next, we observe that the training-based method for the existing transceiver performs much worse than the training-based and semiblind methods for the proposed transceiver at medium-to-high user loads. Finally, note that for the proposed transceiver, the MMSE performance discussed in test case 1 can be viewed as the convergence point of the training-based and semiblind methods as I goes to infinity. Comparing the figures of test case 2 with the figures of test case 1, we observe that for $I = 10$, the training-based method is still far from the MMSE performance, whereas the semiblind method is already very close to the MMSE performance. Hence, as I increases, the semiblind method converges faster to the MMSE performance than the training-based method.

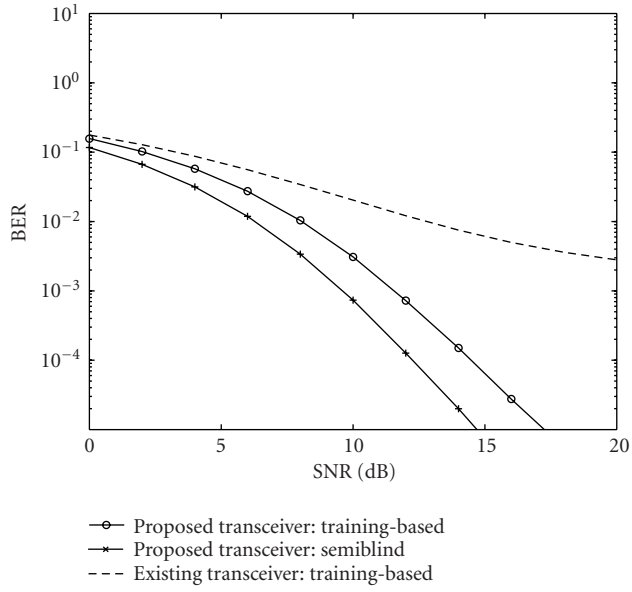


Figure 6: Performance of pilot-based methods for $U = 15$.

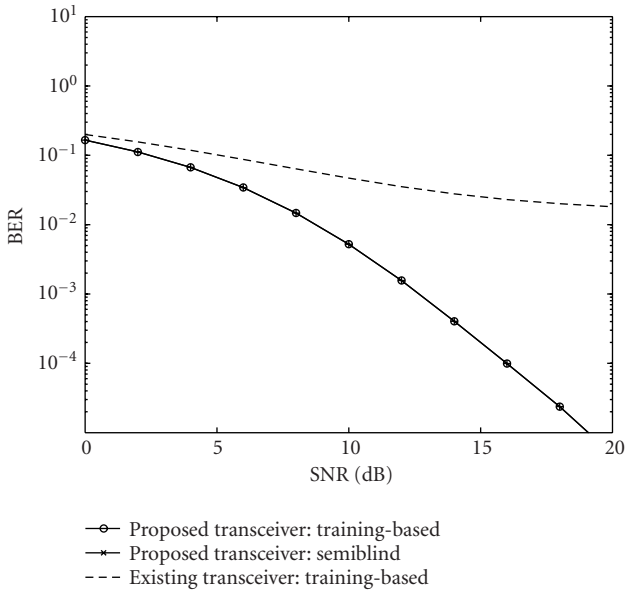


Figure 7: Performance of pilot-based methods for $U = 31$.

5. CONCLUSIONS

We have aimed at combining STBC techniques with the original single-antenna DS-CDMA downlink scheme, resulting into the so-called space-time block coded DS-CDMA downlink schemes. Many space-time block coded DS-CDMA downlink transmission schemes can be considered. We have focussed on a new scheme that enables both the maximum multiantenna diversity and the maximum multipath diversity. Although this maximum diversity can only be collected by ML detection, we have pursued suboptimal detection by means of space-time chip equalization, which lowers the computational complexity significantly. To design the space-time chip equalizers, we have also proposed efficient pilot-based methods. Simulation results have shown improved performance over the space-time RAKE receiver for the space-time block coded DS-CDMA downlink schemes that have been proposed for the UMTS and IS-2000 W-CDMA standards.

ACKNOWLEDGMENTS

This research work was carried out in the frame of the Belgian State's Interuniversity Poles of Attraction Programme (2002–2007): IAP P5/22 ("Dynamical Systems and Control: Computation, Identification, and Modelling") and P5/11 ("Mobile Multimedia Communication Systems and Networks"); the Concerted Research Action GOA-MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology) of the Flemish Government; and Research Project FWO no. G.0196.02 ("Design of Efficient Communication Techniques for Wireless Time-Dispersive Multiuser MIMO Systems"). Part of this work appeared in the proceedings of the International Conference on Communications (ICC), New York city, NY, April-May 2002. During this research work, Geert Leus was a Postdoctoral Fellow of the Fund for Scientific Research - Flanders (FWO - Vlaanderen), and Frederik Petré was a Research Assistant of the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT).

REFERENCES

[1] A. Klein, "Data detection algorithms specially designed for the downlink of CDMA mobile radio systems," in *Proc. IEEE Vehicular Technology Conference*, vol. 1, pp. 203–207, Phoenix, Ariz, USA, May 1997.

[2] I. Ghauri and D. Slock, "Linear receivers for the DS-CDMA downlink exploiting orthogonality of spreading sequences," in *Proc. 32nd Asilomar Conf. on Signals, Systems, and Computers*, vol. 1, pp. 650–654, Pacific Grove, Calif, USA, November 1998.

[3] T. P. Krauss, W. J. Hillery, and M. D. Zoltowski, "Downlink specific linear equalization for frequency selective CDMA cellular systems," *Journal of VLSI Signal Processing*, vol. 30, no. 3, pp. 143–161, 2002.

[4] C. D. Frank, E. Visotsky, and U. Madhow, "Adaptive interference suppression for the downlink of a direct sequence CDMA system with long spreading sequences," *Journal of VLSI Signal Processing*, vol. 30, no. 1–3, pp. 273–291, 2002.

[5] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[6] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communications," *IEEE Trans. Communications*, vol. 46, no. 3, pp. 357–366, 1998.

[7] D. Gesbert, H. Bolcskei, D. Gore, and A. Paulraj, "MIMO wireless channels: capacity and performance prediction," in *IEEE Global Telecommunications Conference, 2000*, vol. 2, pp. 1083–1088, San Francisco, Calif, USA, December 2000.

[8] A. Paulraj and T. Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception (DTDR)," U.S. Patent 5345599, Stanford University, Stanford, Calif, USA, September 1994.

[9] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[10] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.

[11] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.

[12] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.

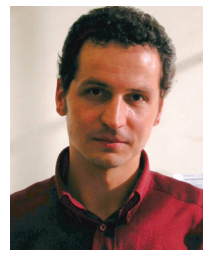
[13] Lindskog E. and A. Paulraj, "A transmit diversity scheme for channels with intersymbol interference," in *Proc. IEEE Conference on International Communications*, vol. 1, pp. 307–311, New Orleans, La, USA, June 2000.

[14] N. Al-Dhahir, "Single-carrier frequency-domain equalization for space-time block-coded transmissions over frequency-selective fading channels," *IEEE Communications Letters*, vol. 5, no. 7, pp. 304–306, 2001.

[15] S. Zhou and G. B. Giannakis, "Space-time coding with maximum diversity gains over frequency-selective fading channels," *IEEE Signal Processing Letters*, vol. 8, no. 10, pp. 269–272, 2001.
[16] B. Hochwald, T. L. Marzetta, and C. B. Papadias, "A transmitter diversity scheme for wideband CDMA systems based on space-time spreading," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 1, pp. 48–60, 2001.
[17] S. Barbarossa, G. Scutari, and A. Swami, "MUI-free CDMA systems incorporating space-time coding and channel shortening," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 2213–2216, Orlando, Fla, USA, May 2002.
[18] S. Zhou and G. B. Giannakis, "Single-carrier space-time block coded transmissions over frequency-selective fading channels," *IEEE Transactions on Information Theory*, vol. 49, no. 1, pp. 164–179, 2003.
[19] H. Holma and A. Toskala, Eds., *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*, John Wiley & Sons, New York, NY, USA, 2001.
[20] F. Petré, G. Leus, M. Engels, M. Moonen, and H. De Man, "Space-time chip equalization for WCDMA forward link with code-multiplexed pilot and soft handover," in *IEEE Global Telecommunications Conference, 2001*, vol. 1, pp. 280–284, San Antonio, Tex, USA, November 2001.
[21] F. Petré, G. Leus, L. Deneire, M. Engels, and M. Moonen, "Adaptive space-time chip-level equalization for WCDMA downlink with code-multiplexed pilot and soft handover," in *Proc. IEEE Conference on International Communications*, pp. 1635–1639, New York City, NY, USA, April 2002.
[22] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 3rd edition, 1995.
[23] M. Lenardi, A. Medles, and D. T. M. Slock, "Comparison of downlink transmit diversity schemes for RAKE and SINR maximizing receivers," in *Proc. IEEE Conference on International Communications*, pp. 1679–1683, New Orleans, La, USA, June 2001.
[24] C. D. Frank, "MMSE reception of DS-CDMA with open-loop transmit diversity," in *Proc. 2nd International Conference on 3G Mobile Communication Technologies, 2001*, pp. 156–160, London, UK, March 2001.

Geert Leus was born in Leuven, Belgium, in 1973. He received his Electrical Engineering degree and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Belgium, in June 1996 and May 2000, respectively. He has been a Research Assistant and a Postdoctoral Fellow of the Fund for Scientific Research - Flanders, Belgium, from October 1996 till September 2003. During that period, Geert Leus was affiliated with the Electrical Engineering Department, the Katholieke Universiteit Leuven, Belgium. Currently, Geert Leus is an Assistant Professor at the Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, The Netherlands. During the summer of 1998, he visited Stanford University, and from March 2001 till May 2002, he was a Visiting Researcher and Lecturer at the University of Minnesota. His research interests are in the area of signal processing for communications. Geert Leus received a 2002 IEEE Signal Processing Society Young Author Best Paper Award. He is a member of the IEEE Signal Processing for Communications Technical Committee, and an Associate Editor for the IEEE Transactions on Wireless Communications and the IEEE Signal Processing Letters.

Frederik Petré was born in Tienen, Belgium, 1974. He received the Electrical Engineering degree and the Ph.D. in Applied Sciences from the Katholieke Universiteit Leuven, Leuven, Belgium, in July 1997 and December 2003, respectively. In September 1997, he joined the Design Technology for Integrated Information and Communication Systems (DESICS) division, Interuniversity Micro-Electronics Center (IMEC), Leuven, Belgium. Within the Digital Broadband Terminals (DBATE) group of DESICS, he first performed predoctoral research on wireline transceiver design for twisted pair, coaxial cable, and powerline communications. During the fall of 1998, he visited the Information Systems Laboratory (ISL), Stanford University, California, USA, working on OFDM-based powerline communications. In January 1999, he joined the Wireless Systems (WISE) group of DESICS as a Ph.D. Researcher, funded by the Institute for Scientific and Technological Research in Flanders (IWT). Since January 2004, he is a Senior Scientist within the Wireless Research group of DESICS. He is investigating the baseband signal processing algorithms and architectures for future wireless communication systems, like third generation (3G) and fourth generation (4G) cellular networks, and wireless local area networks (WLANs). His main research interests are modulation theory, multiple access schemes, channel estimation and equalization, smart antenna, and MIMO techniques. He is a member of the ProRISC technical program committee and the IEEE Benelux Section on Communications and Vehicular Technology (CVT). He is a member of the Executive Board and Project Leader of the Reconfigurable Radio project of the Network of Excellence in Wireless Communications (NEWCOM), established under the sixth framework of the European Commission.



Marc Moonen received the Electrical Engineering degree and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1986 and 1990, respectively. Since 2000, he has been an Associate Professor at the Electrical Engineering Department, Katholieke Universiteit Leuven, where he is currently heading a research team of sixteen Ph.D. candidates and postdocs working in the area of signal processing for digital communications, wireless communications, DSL, and audio signal processing. He received the 1994 KU Leuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award (with Piet Vandaele), and was a 1997 "Laureate of the Belgium Royal Academy of Science." He was the Chairman of the IEEE Benelux Signal Processing Chapter (1998–2002), and is currently a EURASIP AdCom Member (European Association for Signal, Speech, and Image Processing, 2000). He is the Editor-in-Chief for the EURASIP Journal on Applied Signal Processing (2003), and a member of the editorial board of Integration, the VLSI Journal, IEEE Transactions on Circuits and Systems II, and IEEE Signal Processing Magazine.

EURASIP Journal on Applied Signal Processing 2004:5, 751–761

Joint Power Control and Blind Beamforming over Wireless Networks: A Cross Layer Approach

Zhu Han
Department of Electrical and Computer Engineering and Institute for Systems Research, University of Maryland, College Park, MD 20742, USA
Email: hanzhu@glue.umd.edu

Farrokh R. Farrokhi
Department of Electrical and Computer Engineering and Institute for Systems Research, University of Maryland, College Park, MD 20742, USA
Email: frfarrokhi@yahoo.com

K. J. Ray Liu
Department of Electrical and Computer Engineering and Institute for Systems Research, University of Maryland, College Park, MD 20742, USA
Email: kjrliu@umd.edu

Received 6 January 2003; Revised 9 September 2003

Traditional joint power control and beamforming achieve the targeted signal-to-interference-noise ratio (SINR) at the receivers by assuming the knowledge of the measurements of channel parameters and SINR. Blind beamforming is an effective technique for beamforming and channel estimation without the need of training sequences, thus not consuming extra bandwidth. In this paper, we propose a novel joint power control and blind beamforming algorithm that reformulates the power control problem in such a way that it does not need any prior knowledge and additional measurements in the physical layer. In contrast to the traditional schemes that optimize SINR and, as a result, minimize bit error rate (BER), our proposed algorithm achieves the desired BER by adjusting a quantity available from blind beamforming. By sending this quantity to the transmitter through a feedback channel, the transmit power is iteratively updated in a distributed manner in the wireless networks with cochannel interferences (CCIS). Our proposed algorithm is more robust to estimation errors. We have shown in both analysis and simulation that our algorithm converges to the desired solution. In addition, a Cramer-Rao lower bound (CRB) is derived to compare with the performance of our proposed joint power control and blind beamforming system.

Keywords and phrases: array signal processing, power control, distributed control, direction of arrival estimation.

1. INTRODUCTION

Over the past few decades, wireless communications and networking have witnessed an unprecedented growth, and have become pervasive much sooner than anyone could have imagined. One of the major challenges for the system design is the limited available radio frequency spectrum. Channel reuse is a common method to increase the wireless system capacity by reusing the same channel beyond some distance. However this introduces cochannel interference (CCI) that degrades the link quality. Two promising approaches to combat CCI are power control and antenna array processing. Power control is one direct approach toward minimizing CCI. The transmit powers are constantly adjusted. They are increased if the signal-to-interference-noise-ratios (SINRs) at the receivers are low and are decreased if the SINRs are high. Such a process improves the quality of weak links and reduces the unnecessary transmit powers. Antenna array processing techniques such as beamforming can be applied to receive and transmit multiple signals that are separated in space. Hence, multiple cochannel users can be supported in each cell to increase the capacity by exploring the space diversity.

Many works have been reported in the literature for employing power control and beamforming to reduce CCI. Traditional beamformers such as minimum mean square error (MMSE) and minimum variance distortion response (MVDR) methods have been commonly employed [1]. In [2, 3], general frameworks for power control are constructed. Beamforming is a physical layer technique that can greatly

increase receivers' SINR by using the signal processing algorithms, while power control is a media access control layer technique that can effectively control users' transmit powers to share the channels. Many joint power control and beamforming algorithms are proposed in [4, 5, 6, 7, 8]. Most of the existing works assume the availability of prior channel information and measurement of SINR.

As a majority of communication systems often struggle with the limited bandwidth constraint, it is desirable for the receiver with multiple antennas to steer to the desired direction and to estimate the transmit signals without consuming much channel bandwidth. By eliminating the training sequence overhead, used for estimation, and maximizing the channel capacity for information transmission, blind estimation and beamforming [9, 10, 11, 12, 13, 14, 15, 16] offer a bandwidth efficient solution to signal separation and estimation. Its importance also lies in the practical need for some communication receivers to equalize unknown channels without the assistance and the expense of training sequences.

Current methods of joint power control and beamforming [4, 5, 6, 7, 8] assume perfect measurement of channel parameters and SINR at the receivers, which is very difficult to obtain in practice. Blind beamforming can estimate and separate, without the use of training sequences, the transmitted signals that suffer from the channel distortion and additive noise. The difficulties for joint power control and blind beamforming are to formulate such a cross-layer problem into a joint optimization problem, and develop an algorithm that can be self-trained and adaptively adjust the system parameters. In this paper, we present a novel joint power control and blind beamforming algorithm for a multicell multiantenna system. Based on a reformulated joint problem, our proposed algorithm optimizes the bit error rate (BER) using a quantity directly available from the blind beamforming and estimation, which avoids additional measurements mentioned above. Mobiles' transmit powers are updated in a distributed manner such that the CCI is effectively reduced. Convergence properties of the proposed algorithm are discussed. A Cramer-Rao lower bound (CRB) is derived to show the effect of power control on the symbol estimation performance in the networks. Simulation results illustrate that our algorithm converges to the desired solution and is more robust to channel estimation error compared with traditional joint power control and training-based beamforming algorithm.

The organization of this paper is as follows. In Section 2, we present the system model and the traditional joint power control and beamforming problem. In Section 3, first we choose a blind beamforming algorithm. Then we give the reformulated joint power control and blind beamforming problem. An adaptive algorithm is developed and a system is constructed. In Section 4, the convergence and uniqueness of the solution are analyzed. The CRB is derived to compare the performance. In Section 5, we evaluate our algorithm via numerical studies. In Section 6, we give the summary and conclusion.

2. SYSTEM MODEL, BEAMFORMING AND POWER CONTROL

Consider K distinct cells in wireless networks where cochannel links exist. Each cell consists of one base station and its assigned D mobiles. Antenna arrays with M elements are used only at the base station and $M \geq D$. We assume that coherent detection is possible so that it is sufficient to model this multiuser system by an equivalent baseband model. Each link is affected by the slow Rayleigh fading. The propagation delay is far less than one symbol period. For uplink case, the ith base station antenna array's output vector is given by

$$\mathbf{x}_i(t) = \sum_{k=1}^{K}\sum_{d=1}^{D}\sqrt{G_{ki}^d P_k^d}\,\alpha_{ki}^d \mathbf{a}_{ki}^d(\theta_{ki}^d)\cdot g_k^d(t-\tau_{ki})s_k^d(t-\tau_{ki}) + \mathbf{n}_i(t), \tag{1}$$

where G_{ki}^d is path loss, α_{ki}^d is fading coefficient, P_k^d is transmit power, $\mathbf{a}_{ki}^d(\theta_{ki}^d)$ is the ith base station array response vector to the signal from the dth mobile in the kth cell at direction θ_{ki}^d, $g_k^d(t)$ is shaping function, $s_k^d(t)$ is message symbol, τ_{ki} is the delay, and $\mathbf{n}_i(t)$ is thermal noise vector. We assume the synchronous transmission for all the users within the same cell, that is, $\tau_{ii} = 0$, for all i. The synchronous assumption is reasonable because the symbol timing can be effectively controlled within each cell. We assume that the CCI from other cells is asynchronous for the desired signals within the cell and τ_{ki}, $k \neq i$, is uniformly distributed within the symbol duration. We assume that the channels are flat fading and stable within a frame of hundreds of symbols. Define the impulse response from the dth mobile in the kth cell to the pth element of the ith base station as $h_{ki}^{dp} = \alpha_{ki}^d a_{ki}^{dp}(\theta_{ki}^d) r_{ki}^{dp}$, where r_{ki}^{dp} includes the effect of the transmitter, receiver filter, and shaping function $g_k^d(t-\tau_{ki})$. In the vector form, it is given by $\mathbf{h}_{ki}^d = [h_{ki}^{1d}, \ldots, h_{ki}^{Md}]^T$. The sampled received vector for this DK users and MK antenna outputs multicell system at time n is given by

$$\mathbf{X}(n) = \mathbf{A}\mathbf{S}(n) + \mathbf{n}(n), \tag{2}$$

where $\mathbf{X}(n) = [\mathbf{x}_1^T(n), \mathbf{x}_2^T(n), \ldots, \mathbf{x}_K^T(n)]^T$, $\mathbf{S}(n) = [\mathbf{S}_1^T(n), \mathbf{S}_2^T(n), \ldots, \mathbf{S}_K^T(n)]^T$, $\mathbf{S}_i(n) = [s_i^1(n), \ldots, s_i^D(n)]^T$, $\mathbf{n}(n)$ is the sampled thermal noise vector, and

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{21} & \cdots & \mathbf{A}_{K1} \\ \mathbf{A}_{12} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{K2} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{A}_{1K} & \mathbf{A}_{2K} & \cdots & \mathbf{A}_{KK} \end{bmatrix}_{MK\times DK}, \tag{3}$$

where $\mathbf{A}_{ij} = [\sqrt{P_i^1 G_{ij}^1}\mathbf{h}_{ij}^1 \cdots \sqrt{P_i^D G_{ij}^D}\mathbf{h}_{ij}^D]$.

Let $\mathbf{w}_i^d$ be the beamforming weight vector for the dth mobile in the ith cell. Without loss of generality, we normalize the beamformer weight vector $\|(\mathbf{w}_i^d)^H \mathbf{h}_{ii}^d\|^2 = 1$, which will not change the receivers' SINRs. We assume that the transmitted signals from different sources are uncorrelated and

zero mean, and the additive noise is spatially and temporally white with variance $\mathbf{N}_i = \sigma^2\mathbf{I}_{M\times M}$, where σ^2 is the thermal noise variance. The dth user's SINR at its associated ith base station's beamformer output is

$$\Gamma_i^d = \frac{P_i^d G_{ii}^d}{\sum\sum_{(k,j)\neq(i,d)} P_k^j G_{ki}^j \left\|\left(\mathbf{w}_i^d\right)^H \mathbf{h}_{ki}^j\right\|^2 + \left(\mathbf{w}_i^d\right)^H \mathbf{N}_i \mathbf{w}_i^d}. \quad (4)$$

The issue in question here is how to find the users' beamforming vectors and transmit powers such that each user has the desired link quality and does not introduce unnecessary CCI to other users. In the rest of this section, we will briefly illustrate the traditional joint power control and beamforming.

An adaptive antenna array is designed to receive the signals from the desired directions and attenuate signals' radiations from other directions of no interest. The outputs of the array elements are weighted by a beamformer. In order to suppress the interferences, the beamformer places its nulls in the directions of interference sources and steers to the direction of the target signal. Some most popular beamformers are MMSE and MVDR beamformers [1]. In this paper, we will compare joint power control and MVDR beamforming method with our proposed blind scheme because MVDR beamformer is commonly used in the literature [4].

If the channel responses $\mathbf{h}_{ii}^d$ can be estimated, the beamforming vector can be calculated by the MVDR method, which minimizes the total interferences at the output of a beamformer, while the gain for the desired dth user in the ith cell is kept as a constant. The MVDR problem can be defined as

$$\min_{\mathbf{w}_i^d} \left\|\left(\mathbf{w}_i^d\right)^H \mathbf{x}_i\right\|^2, \quad (5)$$

subject to

$$\left\|\left(\mathbf{w}_i^d\right)^H \mathbf{h}_{ii}^d\right\|^2 = 1, \quad i = 1,\ldots,M. \quad (6)$$

Define correlation matrix as $\boldsymbol{\phi}_i = E[\mathbf{x}_i\mathbf{x}_i^H]$. The optimal weight vector is given by

$$\hat{\mathbf{w}}_i^d = \frac{\boldsymbol{\phi}_i^{-1}\mathbf{h}_{ii}^d}{\left(\mathbf{h}_{ii}^d\right)^H \boldsymbol{\phi}_i^{-1}\mathbf{h}_{ii}^d}. \quad (7)$$

In traditional power control schemes, the overall transmit powers of all links are minimized, while each link's transmit power is selected so that its SINR is equal to or larger than a fixed and predefined targeted SINR threshold γ_i^d required to maintain the link quality. The power control problem can be defined as

$$\min_{P_i^d} \sum_{i=1}^{K}\sum_{d=1}^{D} P_i^d, \quad (8)$$

subject to

$$(\mathbf{I} - \mathbf{BF})\mathbf{P} \geq \mathbf{u}, \quad (9)$$

where $\mathbf{u} = [u_1^1,\ldots,u_1^D,\ldots,u_K^1,\ldots,u_K^D]^T$, $\mathbf{P} = [P_1^1,\ldots,P_1^D,\ldots,P_K^1,\ldots,P_K^D]^T$, $\mathbf{I}$ is the identical matrix, $\mathbf{B} = \mathrm{diag}\{\gamma_1^1,\ldots,\gamma_1^D,\ldots,\gamma_K^1,\ldots,\gamma_K^D\}$, and

$$[\mathbf{F}]_{kj} = \begin{cases} 0 & \text{if } j = k, \\ \dfrac{G_{i'i}^{d'}\left\|\left(\mathbf{w}_i^d\right)^H \mathbf{h}_{i'i}^{d'}\right\|^2}{G_{ii}^d} & \text{if } j \neq k, \end{cases} \quad (10)$$

where $i = \lfloor k/D \rfloor$, $d = \mathrm{mod}(k, D)$, $i' = \lfloor j/D \rfloor$, $d' = \mathrm{mod}(j, D)$, and $k, j = 1\cdots KD$.

If the spectral radius $\rho(\mathbf{BF})$ [17], that is, the maximum eigenvalue of BF, is inside the unit circle, the system has feasible solutions and there exists a positive power allocation vector to achieve the desired targeted SINRs. By Perron-Frobenius theorem [17, 18], the optimum power vector for this problem is $\hat{\mathbf{P}} = (\mathbf{I} - \mathbf{BF})^{-1}\mathbf{u}$. Many adaptive algorithms [3, 4, 19] have been developed to reduce the system complexity by the following distributed iteration:

$$P_i^d(n+1) = \frac{\gamma_i}{G_{ii}^d} I_i^d, \quad (11)$$

where $I_i^d = (\mathbf{w}_i^d)^H \mathbf{N}_i \mathbf{w}_i^d + \sum_{(k,j)\neq(i,d)}^{K,D} \|(\mathbf{w}_i^d)^H \mathbf{h}_{ki}^j\|^2 P_k^j G_{ki}^j$ and I_i^d can be easily estimated at the receivers. The power allocation is balanced at the equilibrium when the power update in (11) has converged.

The level of CCI depends on both channel gain and transmit power. The optimal beamforming vector may vary for different powers. Hence the beamforming and power control should be considered jointly. In [4], a joint power control and beamforming scheme has been proposed. An iterative algorithm is developed to jointly update the transmit powers and beamformer weight vectors. The algorithm converges to the jointly optimal transmit power and beamforming solution. The joint iterative algorithm can be summarized by the following two steps:

(i) beamforming in physical layer: MVDR algorithm,
(ii) power update in MAC layer: $\mathbf{P}^{n+1} = \mathbf{BFP}^n + \mathbf{u}$,

where power update step can be implemented by using only local interference measurement. But the algorithm assumes the knowledge of SINR and directions of the desired signals or the perfect measurements of channel responses, which are very difficult to get in practice.

3. JOINT POWER CONTROL AND BLIND BEAMFORMING

In this section, first we consider how to choose a blind beamforming algorithm that can be used for joint optimization with power control. Then we reformulate the joint power control and blind beamforming problem as a cross layer approach. Finally, an adaptive iterative algorithm is developed.

3.1. Choosing a blind beamforming algorithm

The traditional beamforming needs the measurement of spatial responses of the array. A common practice is the use of training sequences [1]. However, it costs bandwidth which

is very precious and limited in wireless networks. Moreover, the measurement errors can greatly reduce the performance of beamforming. This gives us the motivation to use blind beamforming method to separate and estimate the multiple signals arriving at the antenna array. Since beamforming and power control are two different layer techniques, we need to find the blind beamforming algorithms that allow us to have joint optimization across the layers. In [13, 14], a maximum likelihood approach named iterative least squares projection (ILSP) algorithm is proposed. The algorithm explores the finite alphabet property of digital signals. The channel estimation and symbol detection can be implemented at the same time. In addition, a quantity is available for BER performance and can be used for power control optimization [20]. In this subsection, we will briefly review the ILSP algorithm.

Consider the same channel module in (2). The dth mobile inside the ith cell generates binary data $s_i^d(n)$ with power P_i^d transmitted over a low delay spread Rayleigh fading channel. The channel and antenna array response is $\mathbf{h}_{ii}^d$. The sampled antenna output at the ith base station is given by

$$\mathbf{x}_i(n) = \sum_{d=1}^{D} \mathbf{h}_{ii}^d \sqrt{P_i^d G_{ii}^d} s_i^d(n) + \mathbf{v}_i(n), \tag{12}$$

where $\mathbf{v}_i(n)$ includes the ith base station antenna thermal noise and all the CCIs from the other cells, that is,

$$\mathbf{v}_i(n) = \mathbf{n}_i(n) + \sum_{\substack{k=1 \\ k\neq i}}^{K} \sum_{d=1}^{D} \mathbf{h}_{ki}^d \sqrt{P_k^d G_{ki}^d} s_k^d(n), \tag{13}$$

where $\mathbf{n}_i(n)$ is the $M \times 1$ sampled thermal noise vector.

The ILSP algorithm works with a shifting window on data blocks of size N. Assume that the channel is constant over the N symbol periods. In the ith cell, we obtain the following formulation of the lth data block:

$$\mathbf{X}_i(l) = \mathbf{A}_i \mathbf{S}_i(l) + \mathbf{V}_i(l), \tag{14}$$

where l is block number,

$$\begin{aligned}
\mathbf{X}_i(l) &= \left[\mathbf{x}_i(lN+1) \;\; \mathbf{x}_i(lN+2) \;\; \cdots \;\; \mathbf{x}_i\big((l+1)N\big)\right], \\
\mathbf{V}_i(l) &= \left[\mathbf{v}_i(lN+1) \;\; \mathbf{v}_i(lN+2) \;\; \cdots \;\; \mathbf{v}_i\big((l+1)N\big)\right], \\
\mathbf{S}_i(l) &= \left[\mathbf{s}_i(lN+1) \;\; \mathbf{s}_i(lN+2) \;\; \cdots \;\; \mathbf{s}_i\big((l+1)N\big)\right], \\
\mathbf{s}_i(n) &= \left[s_i^1(n) \;\; \cdots \;\; s_i^D(n)\right]^T, \\
\mathbf{A}_i &= \left[\sqrt{P_i^1 G_{ii}^1}\mathbf{h}_{ii}^1 \;\; \cdots \;\; \sqrt{P_i^D G_{ii}^D}\mathbf{h}_{ii}^D\right].
\end{aligned} \tag{15}$$

We assume that the number of users is known or has been estimated.

The ILSP algorithm uses the finite alphabet property of the input to implement a least squares algorithm that has good convergence properties for the channel with low delay spread. The algorithm is carried out in two steps to alternatively estimate $\mathbf{A}_i$ and $\mathbf{S}_i$ as follows:

$$\min_{\mathbf{A}_i, \mathbf{S}_i} f(\mathbf{A}_i, \mathbf{S}_i; \mathbf{X}_i) = \left|\left|\mathbf{X}_i(l) - \mathbf{A}_i \mathbf{S}_i(l)\right|\right|^2. \tag{16}$$

The first step is a least square minimization problem, where $\mathbf{S}_i$ is unstructured and its amplitude is continuous without considering the discrete nature of modulations, while $\mathbf{A}_i$ is fixed and equal to the estimated $\hat{\mathbf{A}}_i$. In the second step, each element of the solution $\mathbf{S}_i$ is projected to its closest discrete values $\hat{\mathbf{S}}_\mathbf{i}$. Then a better estimate of $\hat{\mathbf{A}}_i$ is obtained by minimizing $f(\mathbf{A}_i, \hat{\mathbf{S}}_i; \mathbf{X}_i)$ with respect to $\mathbf{A}_i$, keeping $\hat{\mathbf{S}}_i$ fixed. We continue this process until estimates of $\hat{\mathbf{A}}_i$ and $\hat{\mathbf{S}}_i$ are converge. The ILSP algorithm is given in Algorithm 1.

3.2. Reformulation of joint power control and beamforming

In traditional joint power control and beamforming, the user's received SINR is larger than or equal to a targeted value to maintain the link quality such as the desired BER. In this paper, we proposed another quantity available from the ILSP algorithm to directly ensure each user's BER. For simplicity, we use BPSK modulation for the analysis and simulation. The other PAM or MQAM modulation methods can be easily extended in a similar way. It has been shown in [14] that the error probability of ILSP algorithm is approximated by

$$P_r(s_i^d) = Q\left(\sqrt{\frac{2}{\mathrm{Var}\left[\hat{s}_i^d(n)\right]}}\right), \tag{17}$$

where each estimated signal $\hat{s}_i^d(n)$ has $E[\hat{s}_i^d(n)] = s_i^d(n)$, that is, ILSP is an unbiased estimator with variance

$$\mathrm{Var}\left[\hat{s}_i^d(n)\right] = 2\sigma_i^2 \left(\mathbf{A}_i^H \mathbf{A}_i\right)_{dd}^{-1}, \tag{18}$$

where, in our case, $\sigma_i^2 = E[\mathbf{v}_i(n)^H \mathbf{v}_i(n)]$ and can be estimated by

$$\sigma_i^2 \approx \frac{1}{N}\left|\left|\mathbf{X}_i - \hat{\mathbf{A}}_i \hat{\mathbf{S}}_i\right|\right|^2 = \frac{1}{N}\left|\left|\mathbf{V}_i\right|\right|^2. \tag{19}$$

In [14], (18) is developed for single cell environment with additive white Gaussian noise. In our case, we need to perform optimization in multicell scenario with CCI. Because there are a large number of CCI sources with similar received powers, by the central limit theorem, we can assume that $\mathbf{v}_i(n)$ approaches a zero-mean Gaussian vector. So (18) still holds in our case. From the simulation results in Section 5, we can show that this assumption is valid.

In our proposed joint power control and blind beamforming scheme, the key issue is the quantity $\mathrm{Var}[\hat{s}_i^d(n)]$ which is directly related to error performance. $\mathrm{Var}[\hat{s}_i^d(n)]$ is a function of σ_i^2 and $\mathbf{A}_i$, so it is also a function of all P_i^d, for all i, d. We want the maximum variance for each user's $\mathrm{Var}[\hat{s}_i^d(n)]$ to be less than or equal to a predefined value var_0 so that each user's BER is less than the desired value. However, if var_0 is too small, each user's transmit power will be too large and cause too much CCI. Under this condition, the system may not be feasible, that is, no matter how large the transmit powers are, the receivers cannot achieve desired BER. So we need a feasibility constraint for var_0. The reformulated joint power control and blind beamforming

(1) Initial $\hat{\mathbf{A}}_{i,0}$, Step $m = 0$.
(2) $m = m + 1$
  (a) $\bar{\mathbf{S}}_{i,m} = \mathbf{A}^{+}_{i,m-1}\mathbf{X}_i$,
    where $\mathbf{A}^{+}_{i,m-1} = (\hat{\mathbf{A}}^{H}_{i,m-1}\hat{\mathbf{A}}_{i,m-1})^{-1}\hat{\mathbf{A}}^{H}_{i,m-1}$,
  (b) projection onto finite alphabet
    $\hat{\mathbf{S}}_{i,m} = \text{proj}[\bar{\mathbf{S}}_{i,m}]$,
  (c) $\hat{\mathbf{A}}_{i,m} = \mathbf{X}_i\hat{\mathbf{S}}^{+}_{i,m}$,
    where $\hat{\mathbf{S}}^{+}_{i,m} = \hat{\mathbf{S}}^{H}_{i,m}(\hat{\mathbf{S}}_{i,m}\hat{\mathbf{S}}^{H}_{i,m})^{-1}$.
(3) Repeat until $(\hat{\mathbf{A}}_{i,m}, \hat{\mathbf{S}}_{i,m}) \approx (\hat{\mathbf{A}}_{i,m-1}, \hat{\mathbf{S}}_{i,m-1})$.

Algorithm 1

problem is given by

$$\min_{P_i^d} \sum_{i=1}^{K}\sum_{d=1}^{D} P_i^d, \tag{20}$$

subject to

$$\text{Var}\left(\hat{s}_i^d(n)\right) \le \text{var}_0, \quad \forall i, d, \tag{21}$$

where var_0 is feasible. In order to solve this problem, we need to develop a distributed algorithm such that each user can adapt its transmit power by using only local information. We need to evaluate the feasible range of var_0 such that the system is feasible, that is, there exists a possible power allocation vector. The convergence and optimality of the adaptive algorithm will be considered in Section 4.

3.3. Adaptive iterative algorithm

In this subsection, we assume that var_0 is feasible for the system. We will discuss the feasibility issue in Section 4.1. In ILSP algorithm, the iteration stops when the estimated channel response matrix and symbol matrix have converged. In the algorithm, we use the final channel response matrix $\hat{\mathbf{A}}_i$ to substitute $\mathbf{A}_i$ in (18). Then the estimation of $\text{Var}(\hat{s}_i^d(n))$ is calculated by

$$\text{var}_i^d = 2\sigma_i^2\left(\hat{\mathbf{A}}_i^H\hat{\mathbf{A}}_i\right)^{-1}_{dd}. \tag{22}$$

In the uplink, the value of var_i^d is obtained in the base station and compared with the desired var_0. If var_i^d is too large, it means that the BER for the dth user is too large and consequently, the dth user's power needs to be increased. If var_i^d is too small, it is unnecessary to have such a high power for the dth user. Consequently, the power needs to be reduced. The power update stops when transmit powers have converged in the consecutive iterations, that is, $\text{var}_i^d \approx \text{var}_0$. Each user's power is updated by the simple feedback of $\lambda = \text{var}_i^d/\text{var}_0$ from the base station. The power update scheme can be easily implemented in a distributed manner. In each iteration, the power is updated by

$$P_i^d(m+1) = \lambda P_i^d(m), \tag{23}$$

where m is the iteration number.

With the above power update equation, we develop the following joint adaptive power control and blind beamforming algorithm. The algorithm is initialized by some feasible power allocation vector $\mathbf{P}(0)$ and some approximate channel estimation $\hat{\mathbf{A}}_{i,0}$ [13]. The user's BER may be larger than the desired value during the initialization. In each iteration, first, ILSP blind estimate algorithm is applied to estimate the antenna array responses and the transmitted signals. Then var_i^d is calculated. The new transmit power is updated by (23). The iteration is stopped by comparing the power vector of the two consecutive iterations. When the algorithm stops, each user's desired BER will be satisfied. The adaptive algorithm is summarized in Algorithm 2.

With the adaptive algorithm, we can construct a joint power control and blind beamforming system as shown in Figure 1. The variance calculator module calculates the estimation var_i^d from the ILSP module. The updating information of transmit powers is computed by the power update module. Then the simple power update information is sent back to mobiles via the feedback channels. When the algorithm converges, the output data from the ILSP module will have the desired BER.

4. ANALYSIS AND CONVERGENCE OF THE ALGORITHM

4.1. Convergence analysis

In this subsection, we analyze the condition for our proposed algorithm to converge, that is, we find the feasible range for var_0. Then we prove that the power update converges to a unique solution when the system is feasible, while the blind beamforming may not converge to a unique solution. So our proposed joint power control and blind beamforming algorithm may have local minima because of the inherited characteristics of the blind estimation. We will propose a method to avoid the local minima. From the simulation results in Section 5, we can show that even with the possible local minima, the proposed algorithm performs comparably well with the traditional joint power control and beamforming algorithm.

Consider the transmission from the dth mobile to its associated ith base station with $\mathbf{h}_{ii}^d$ and G_{ii}^d being the channel response and link gain, respectively, and $\mathbf{A}_i$ being the channel response matrix. We want to find the expression $\text{Var}[\hat{s}_i^d(n)]$ in (18). Then we will analyze the conditions for the convergence of our algorithm. We have

$$\left[\mathbf{A}_i^H\mathbf{A}_i\right]_{jk} = \sqrt{P_i^j P_i^k G_{ii}^j G_{ii}^k}\left(\mathbf{h}_{ii}^j\right)^H\mathbf{h}_{ii}^k. \tag{24}$$

The $\det(\mathbf{A}_i^H\mathbf{A}_i)$ can be expanded by the following alternating sum form:

$$\det\left(\mathbf{A}_i^H\mathbf{A}_i\right) = P_i^1 G_{ii}^1 \cdots P_i^D G_{ii}^D f_1(\mathbf{h}_{ii}), \tag{25}$$

where $\mathbf{h}_{ii} = [\mathbf{h}_{ii}^1, \ldots, \mathbf{h}_{ii}^D]$ and $f_1(\mathbf{h}_{ii})$ is a real function of channel responses $\mathbf{h}_{ii}^d$, for all d. Then it follows from the cofactor method of matrix inverse [17] that

(1) Given $\mathbf{P}(0)$, var_0, $m = 0$, and $\hat{\mathbf{A}}_i = \hat{\mathbf{A}}_{i,0}$.
(2) Received data block at base station i,
  (i) ILSP blind estimation to get $\hat{\mathbf{A}}_i$;
  (ii) For each mobile d inside ith cell,
    $\text{var}_i^d = 2\hat{\sigma}_i^2(\hat{\mathbf{A}}_i^H\hat{\mathbf{A}}_i)_{dd}^{-1}$
    $\lambda = \dfrac{\text{var}_i^d}{\text{var}_0}$
    $P_i^d(m+1) = \lambda P_i^d(m)$;
  (iii) $\hat{\mathbf{A}}_{i,0} = \hat{\mathbf{A}}_i$.
(3) $m = m + 1$. Go to step 2;
  Repeat until $\mathbf{P}_i(m) \approx \mathbf{P}_i(m-1)$, for all i.

Algorithm 2: Joint power control and blind beamforming algorithm.

$$(\mathbf{A}_i^H\mathbf{A}_i)_{dd}^{-1} = \frac{f_2^d(\mathbf{h}_{ii})\prod_{j=1,\,j\neq d}^{j=D} P_i^j G_{ii}^j}{f_1(\mathbf{h}_{ii})\prod_{j=1}^{j=D} P_i^j G_{ii}^j} = \frac{f_3(\mathbf{h}_{ii})}{P_i^d G_{ii}^d}, \tag{26}$$

where $f_2^d(\mathbf{h}_{ii})$ is a real function of channel responses $\mathbf{h}_{ii}^j$, $j \neq d$, and $f_3(\mathbf{h}_{ii}) = f_2^d(\mathbf{h}_{ii})/f_1(\mathbf{h}_{ii})$.

Because the channels are not reused in the adjacent cells in most of the communication system, we assume that the CCI plus thermal noise in (13) is Gaussian noise with the variance:

$$\sigma_i^2 = \sum_{j\neq i}^{K}\sum_{d=1}^{D} \left\|\mathbf{h}_{ji}^d\right\|^2 G_{ji}^d P_j^d + M\sigma^2. \tag{27}$$

Now we can calculate $\text{Var}[\hat{s}_i^d(n)]$ as

$$\text{Var}\left[\hat{s}_i^d(n)\right] = \frac{2\sigma_i^2}{(\mathbf{A}_i^H\mathbf{A}_i)_{dd}} = \frac{2\sigma_i^2}{P_i^d G_{ii}^d} f_3(\mathbf{h}_{ii}). \tag{28}$$

An interesting result is that $\text{Var}(\hat{s}_k^d(n))$ is independent of the transmit powers of the other mobiles in the same cell. So the main concern for power control is intercell CCI. Substitute into (23), the power update equation can be expressed as

$$P_i^d(n+1) = \frac{\sum_{j\neq i}^{K}\sum_{d=1}^{D}\left\|\mathbf{h}_{ji}^d\right\|^2 G_{ji}^d P_j^d + M\sigma^2}{G_{ii}^d\,\text{var}_0} f_3(\mathbf{h}_{ii}). \tag{29}$$

In matrix form, we define a matrix $\mathbf{Q}$ as

$$[\mathbf{Q}]_{kj} = \begin{cases} G_{i'i}^{d'} f_4^{kj}/G_{ii}^d & \text{if } i' \neq i, \\ 0 & \text{otherwise,} \end{cases} \tag{30}$$

where $i = \lfloor k/D \rfloor$, $d = \text{mod}(k, D)$, $i' = \lfloor j/D \rfloor$, $d' = \text{mod}(j, D)$, and $f_4^{kj} = \|\mathbf{h}_{ji}^d\|^2 f_3(\mathbf{h}_{ii})$. The matrix expression of (29) for the whole network can be written as

$$\mathbf{P}(n+1) = \frac{1}{\text{var}_0}\mathbf{Q}\mathbf{P}(n) + \mathbf{u}, \tag{31}$$

where $\mathbf{P} = [P_1^1 \cdots P_1^D, \ldots, P_K^1 \cdots P_K^D]^T$, $\mathbf{u} = [u_1, \ldots, u_{DK}]^T$, and

$$u_j = \frac{f_3(\mathbf{h}_{ii})M\sigma^2}{G_{ii}^d\,\text{var}_0}. \tag{32}$$

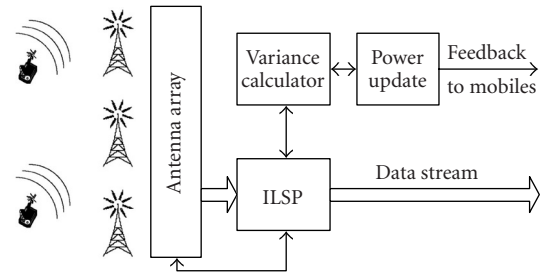


Figure 1: Joint power control and blind beamforming system.

By Perron-Frobenius theorem [17], the power update in (31) has the equilibrium

$$\mathbf{P} = \left(\mathbf{I} - \frac{1}{\text{var}_0}\mathbf{Q}\right)^{-1}\mathbf{u}. \tag{33}$$

If $(\mathbf{I} - (1/\text{var}_0)\mathbf{Q})$ is positive definite, that is, the spectrum radius $|\rho(\mathbf{Q})| < \text{var}_0$, the positive power vector exists and the power update converges. Under this condition, the system is converged when $\text{Var}[\hat{s}_i^d(n)] = \text{var}_0$. From the simulation results in Section 5, we will see that our algorithm converges rapidly to the desired var_0 if $|\rho(\mathbf{Q})| < \text{var}_0$.

When var_0 is too small and less than $\rho(\mathbf{Q})$, the system is not feasible and the adaptive algorithm diverges. In order to prevent the algorithm from diverging, the system will detect the severity of CCI. If the system detect $\rho(\mathbf{Q})$ approaches var_0 or the transmit powers increase very fast, var_0 will be increased so that users will reduce their transmit powers and CCI will be alleviated.

Following the same proof in [19], we can prove that the power update in (29) converges to a unique solution. Suppose that $\hat{\mathbf{P}}$ and $\mathbf{P}^*$ are two different converge power allocation vectors. Without loss of generality, we assume that $\beta = \max_l(\hat{P}_l^d/P_l^{d*}) > 1$ such that $\beta\mathbf{P}^* \geq \hat{\mathbf{P}}$. We can find an index i such that $\beta P_i^{d*} = \hat{P}_i^d$. We have

$$\begin{aligned} \hat{P}_i^d &= \frac{\sum_{j\neq i}^{K}\sum_{d=1}^{D}\left\|\mathbf{h}_{ji}^d\right\|^2 G_{ji}^d \hat{P}_j^d + M\sigma^2}{G_{ii}^d\,\text{var}_0} f_3(\mathbf{h}_{ii}) \\ &\leq \frac{\sum_{j\neq i}^{K}\sum_{d=1}^{D}\left\|\mathbf{h}_{ji}^d\right\|^2 G_{ji}^d \beta P_j^{d*} + M\sigma^2}{G_{ii}^d\,\text{var}_0} f_3(\mathbf{h}_{ii}) \\ &< \beta\frac{\sum_{j\neq i}^{K}\sum_{d=1}^{D}\left\|\mathbf{h}_{ji}^d\right\|^2 G_{ji}^d P_j^{d*} + M\sigma^2}{G_{ii}^d\,\text{var}_0} f_3(\mathbf{h}_{ii}) \\ &= \beta P_i^{d*}. \end{aligned} \tag{34}$$

The above contradiction implies that the power update equation (23) will converge to a unique solution. However, because the solution of blind beamforming may not be unique [14], our proposed joint scheme may fall into local minima. In order to prevent such local minima, we propose the following scheme to avoid the local minima.

When the two users are not well separated in the angle, that is, the array response $\mathbf{A}_i$ is ill-conditioned. The ILSP algorithm can converge to some fixed points that are not the global minima. In this case, instead of projecting

unstructured continuous estimated symbols to the closest discrete values in ILSP algorithm, we enumerate over all Ω^D possible vectors $\mathbf{S}_i^j \in \Omega^D$ and choose the one that minimizes

$$\hat{\mathbf{S}}_i(n) = \arg\min_{\mathbf{S}_i^j \in \Omega^D} ||\mathbf{X}_i(n) - \mathbf{A}_i\mathbf{S}_i^j||^2, \quad \forall j, \tag{35}$$

where Ω is the modulation constellation alphabet. This enumerating method has a better performance but a higher complexity. If the global minimum is still not achieved, it has been shown in [13] that usually one or two reinitializations with random guess are sufficient to yield the global minimum. So we can have two or three parallel structures with different initial values to calculate ILSP algorithm. Then we select the minimal one. The probability of staying in a local minimum will be greatly reduced.

4.2. Cramer-Rao lower bound

In our proposed joint power control and blind beamforming system, the performance of each user's BER is determined by the noise variance, channel conditions, and power allocation. When the additive noise is a zero-mean Gaussian random process, the estimation performance of the unbiased estimator is bounded by the CRB. In this subsection, we derive the covariance matrix for the parameters of the thermal noise variance, the input symbols, and the power allocation vector for the CRB. The results will help us analyze the effects of power control on the users' symbol estimation performances in this multicell system.

For simplicity, we assume that the data are modulated as BPSK, that is, $\mathbf{S}(n) \in \Omega^{KD}$, where $\Omega = \{\pm 1\}$. Similar to the performance analysis of ILSP in [14], we assume that the channel responses are known (the algorithm itself doesn't need such information). The parameters for Fisher information matrix are $\vartheta = [\sigma^2, \mathbf{S}(1), \ldots, \mathbf{S}(N), \mathbf{P}]$. The likelihood function L of the received data $\mathbf{X}(n)$ is given by

$$\begin{aligned} L[\mathbf{X}(1)\cdots\mathbf{X}(N)] &= \frac{1}{(\pi\sigma^2)^{MKN}} \\ &\quad \times \exp\left\{-\frac{1}{\sigma^2}\sum_{n=1}^{N}[\mathbf{X}(n) - \mathbf{A}\mathbf{S}(n)]^H[\mathbf{X}(n) - \mathbf{A}\mathbf{S}(n)]\right\}. \end{aligned} \tag{36}$$

The Fisher information matrix is calculated by

$$\begin{aligned} \mathbf{I}(\theta)_{ij} &= -E\left[\frac{\partial^2 \ln(L)}{\partial\theta_i\partial\theta_j}\right] \\ &= \begin{bmatrix} \frac{MKN}{\sigma^4} & 0 & \cdots & 0 & 0 \\ 0 & \mathbf{Q} & \cdots & 0 & \mathbf{R}(1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{Q} & \mathbf{R}(N) \\ 0 & \mathbf{R}(1) & \cdots & \mathbf{R}(N) & \mathbf{R}_P \end{bmatrix}, \end{aligned} \tag{37}$$

where $\mathbf{Q}$, $\mathbf{R}(n)$, and $\mathbf{R}_P$ are derived in the appendix.

In order to see the effect of the proposed power control on the symbol estimation errors, we define the average mean square error (AMSE) as a performance measure of the symbol estimation:

$$\text{AMSE} = \frac{1}{N}\sum_{n=1}^{N}\frac{||\hat{\mathbf{S}}(n) - \mathbf{S}(n)||^2}{||\mathbf{S}(n)||^2}. \tag{38}$$

Because we use BPSK modulation, $\|\mathbf{S}(n)\|^2 = DK$, for all n, and AMSE is the variance bounded by CRB. The CRB for the symbol estimation can be obtained directly from the inverse of Fisher information matrix, that is,

$$\text{AMSE} \geq \frac{1}{NDK}\sum_{n=1}^{N}\sum_{j=1}^{DK}(\mathbf{I}^{-1}(\theta))_{\mathbf{S}^j(n)\mathbf{S}^j(n)}, \tag{39}$$

where $\mathbf{S}^j(n)$ is the jth element of $\mathbf{S}(n)$. How close AMSE is to the CRB will show the relative efficiency of our proposed algorithm.

5. SIMULATION RESULTS

A network with 50 cells is simulated as shown in Figure 2. Each hexagonal cell's radius is 1000 m. Two adjacent cells do not share the same channel. In each cell, one base station is placed at the center. Two mobiles are placed randomly with uniform distribution. Each mobile transmits BPSK data over Rayleigh fading channels. Each base station employs four-elements antenna array. The noise level is $\sigma = 1$. The transmit frame has $N = 1000$ data symbols. Our shaping function is raised cosine function.

Path loss is due to the decay of the intensity of a propagating radio wave. In our simulations, we use the two slope path loss model [21] to obtain the average received power as a function of distance. According to this model, the average path loss is given by

$$G = \frac{C}{r^a\left(1 + r\lambda_c/(4h_bh_m)\right)^b}, \tag{40}$$

where C is a constant, r is the distance between the mobile and the base station, a is the basic path loss exponent (approximately two), b is the additional path loss component (ranging from two to six), h_b is the base station antenna height, h_m is the mobile antenna height, and λ_c is the wavelength of the carrier frequency. We assume the mobile antenna height is 2 m and the base station antenna height is 50 m. The carrier frequency is 900 MHz.

In Figure 3, we show the analytical and numerical performance of ILSP, compared with MVDR with perfect channel estimation. The numerical results with CCI match the analytical results well especially at high SINR range, which proves our assumption that $\mathbf{V}_i(n)$ can be treated as Gaussian noise when the number of CCIs is large. Our proposed joint power control and blind beamforming has only about 1–2 dB performance loss over traditional power control and MVDR beamforming with perfect channel estimation. However, MVDR beamforming needs additional

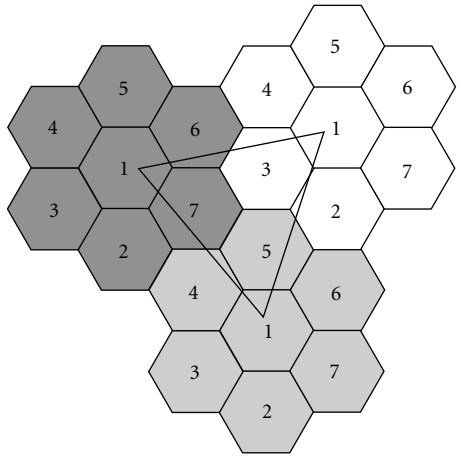


Figure 2: Simulation setup.

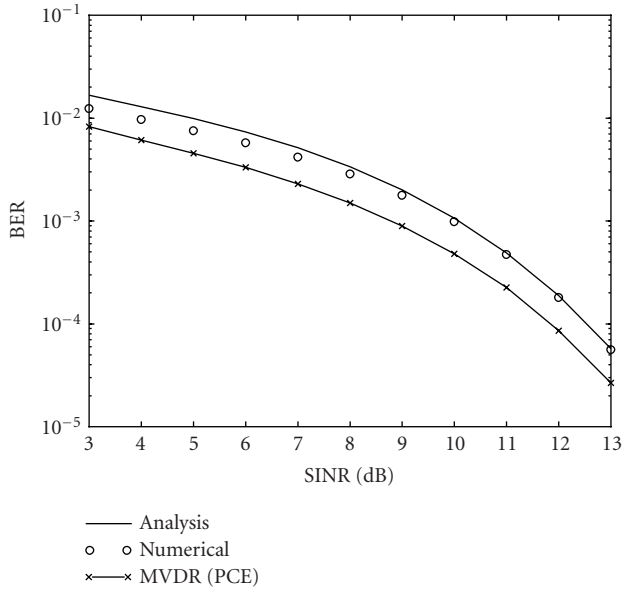
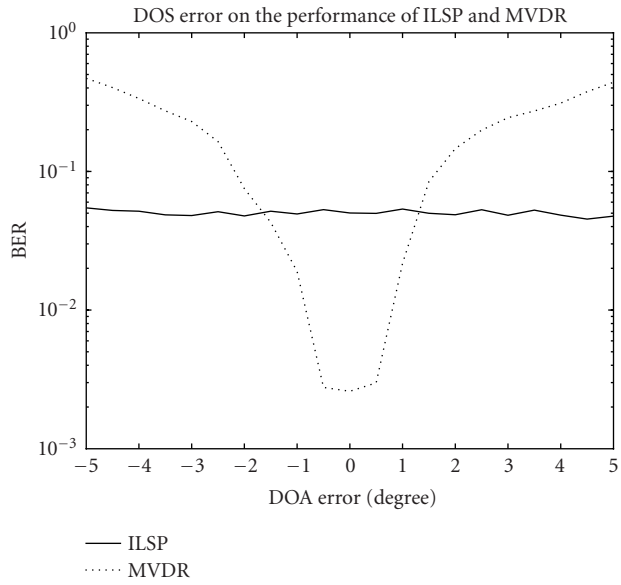
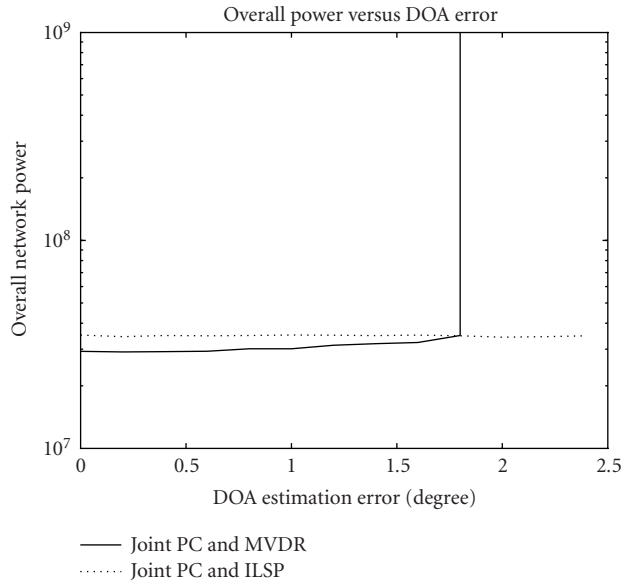


Figure 3: ILSP performance.



(a) BER.



(b) Overall power.

Figure 4: Effects of DOA estimation error.

training sequence to estimate the channel and SINR with prior information that may not be available in practice.

In reality, perfect channel estimation is hard to obtain. In Figure 4, we show the effect of directions of arrivals (DOA) estimation error on the traditional joint power control and MVDR beamforming and our algorithm. In Figure 4a, we compare the BER performance while the transmit power allocation is the same for both algorithms. We can see from the curves that when the channel estimation error for DOA is greater than about 2 degrees, the blind beamforming algorithm outperforms the traditional MVDR. In Figure 4b, we compare the overall transmit power while BER performance is the same for both algorithms. We can see that the blind beamforming algorithm needs a little bit more transmit powers when the DOA estimation error is small. However, the traditional power control with MVDR method will diverge when the DOA estimation error is about 2 degrees. Our proposed joint power control and beamforming algorithm will always converge regardless the DOA variations. When the mobiles are moving, DOA are changing and this will cause the channel estimation errors. The traditional MVDR beamformer may not be aware of the changing and still use the obsolete $\mathbf{h}_{ii}^{d}$ in (7). This will greatly increase BER and transmit powers of the joint power control and MVDR method. The proposed blind scheme will automatically track and adapt to the changes and so it is more robust to channel estimation

errors. Consequently, our algorithm is more robust in applications where usually only the inaccurate channel and SINR estimations are available. It is worthy to mention that the proposed scheme is more sensitive to fast channel varying and the complexity is much higher compared to the traditional training sequence-based algorithm. However, our scheme saves the transmission bandwidth by eliminating the training sequences and is more robust to channel estimation errors.

In Figure 5, we show the numerical results of BER and the overall transmit power versus var_0 for the proposed joint blind beamforming and power control algorithm. When var_0 is decreasing from a large number, BER decreases and the overall power increases slightly. Within a reasonable BER range such as BER $= 10^{-3}$ to BER $= 10^{-5}$, we can calculate the threshold of var_0 for the desired BER. After var_0 decreases to a specific value, the overall transmit power increases and BER decreases quickly. This is because the CCI is too large and $\text{var}_0 \to \rho(\mathbf{Q})$. After var_0 is smaller than some value, the algorithm diverges. Consequently, there is no feasible power control solution, that is, no matter how large the transmit powers are, the receivers cannot ensure the desired BER. This proves that our algorithm behaves exactly the same as the traditional power control algorithm, except that our algorithm directly ensures BER instead of each user's SINR. There is a trade-off between the overall transmit power and BER, while var_0 is the bridge between the two quantities.

In Figure 6, we show the distribution of the number of iterations required for the convergence of our proposed algorithm with different values of var_0. The convergence criteria is that the maximum difference of users' transmit powers between two consecutive iterations is less than 3%. When var_0 is within the range that the system is feasible, we can see that our algorithm converges within a small number of iterations, which demonstrates that our algorithm is robust in the wireless communication systems if the channel gains and topologies have been changed. When var_0 is large, that is, the desired BER is large, the algorithm converges slower. This is because the transmit powers are small when var_0 is large. Consequently, the var_i^d estimation is poor and more iterations are needed for the convergence.

In Figure 7, we compare the AMSE and CRB versus var_0. When var_0 is large and the transmit powers of users are small, the CCI is small. The performance of ILSP is close to CRB. The difference is because discrete alphabets are used for transmitted symbols, while there is no such assumption for CRB. When var_0 is decreasing, the CCI and our algorithm's AMSE are decreasing because of the increasing transmit powers. In this situation, the CRB is much lower than our algorithm performance. This is because we assume that all the channel conditions including $A_{ij}, i \neq j$, are known for CRB, while our algorithm only estimates A_{ii} and treats transmitted signals from other cells as noise. If an algorithm can take consideration of all A_{ij}, for all i, j, its performance will be much better and closer to CRB; however, the complexity will be unacceptably high. When var_0 is smaller than some value, our algorithm diverges. The transmit powers also diverge to arbitrary large values. But the CRB goes extremely low because SINR can be very high, if we know all the channel responses.

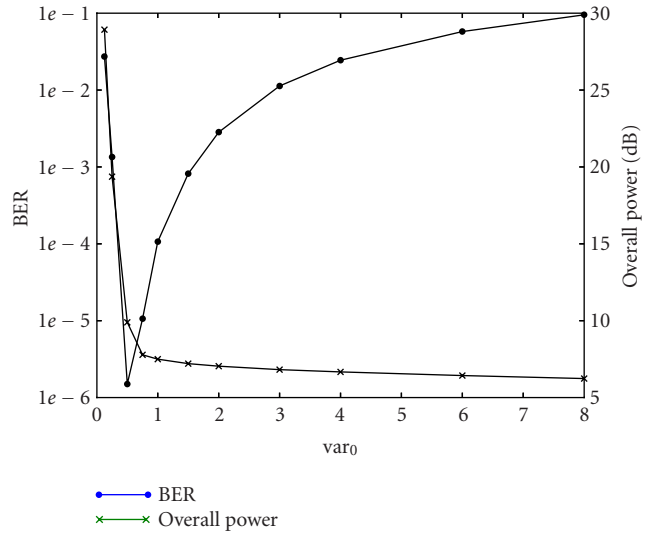


Figure 5: BER and the overall power versus var_0.

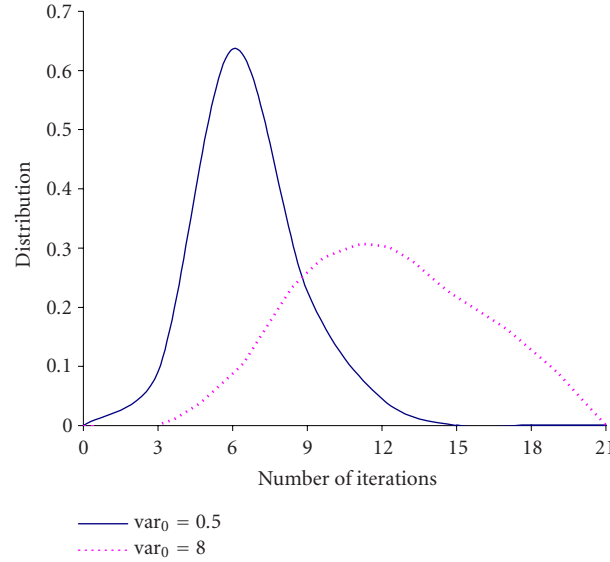


Figure 6: Convergence of the algorithm.

6. CONCLUSION

We have proposed a novel joint power control and blind beamforming algorithm that reformulates the power control problem in terms of a quantity directly related to the error performance of the estimation. First, this approach optimizes BER instead of a theoretically indirect SINR. Secondly, the algorithm does not require additional measurements of

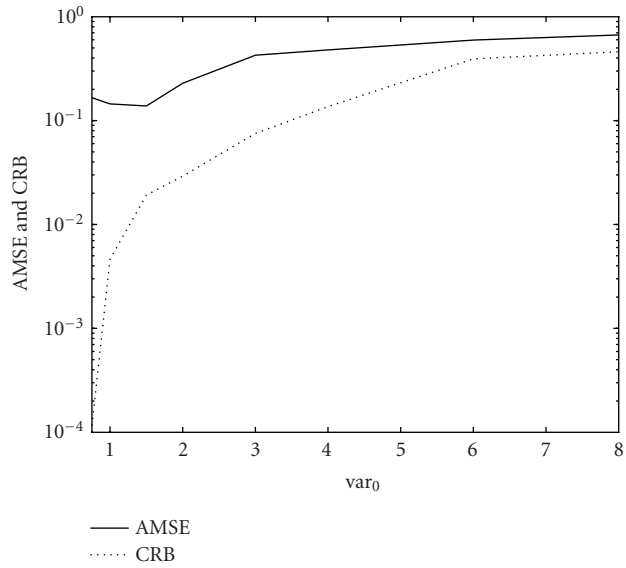


Figure 7: AMSE and CRB versus var_0.

channel or SINR, which saves valuable limited bandwidth. Third, our scheme can be easily implemented in a distributed manner. Fourth, our scheme is more robust to channel estimation error. The proof of convergence of the algorithm is derived and supported by simulation results. Performance results show that our algorithm performs well in the situations where the radio spectrum is limited or the good estimations are hard to obtain.

APPENDIX

From (36), the log likelihood function is

$$\ln(L) = -MKN\ln(\pi) - MKN\ln(\sigma^2) - \frac{1}{\sigma^2}\sum_{n=1}^{N}[\mathbf{X}^H(n) - \mathbf{S}^T(n)\mathbf{A}^H][\mathbf{X}(n) - \mathbf{AS}(n)]. \tag{A.1}$$

We take partial derivatives of (A.1) with respect to σ^2, $\mathbf{S}(n)$, and $\mathbf{P}$:

$$\frac{\partial \ln(L)}{\partial \sigma^2} = -\frac{MKN}{\sigma^2} + \frac{1}{\sigma^4}\sum_{n=1}^{N}\mathbf{e}(n)^H\mathbf{e}(n),$$

$$\frac{\partial \ln(L)}{\partial \mathbf{S}(n)} = \frac{2}{\sigma^2}\,\text{Re}\{\mathbf{A}^H\mathbf{e}(n)\},$$

$$\frac{\partial \ln(L)}{\partial P_i^d} = \frac{2}{\sigma^2}\sum_{n=1}^{N}\text{Re}\left\{\mathbf{S}^T(n)\frac{d\mathbf{A}^H}{dP_i^d}\mathbf{e}(n)\right\},$$

$$\frac{\partial \ln(L)}{\partial \mathbf{P}} = \frac{1}{\sigma^2}\sum_{n=1}^{N}\text{Re}\left\{\text{diag}(\mathbf{S}^T(n))\,\text{diag}\left(\frac{1}{\mathbf{P}}\right)\mathbf{A}^H\mathbf{e}(n)\right\}, \tag{A.2}$$

where $\mathbf{e}(t) = \mathbf{X}(t) - \mathbf{AS}(t)$, and $\text{diag}(1/\mathbf{P}) = \text{diag}(1/P_1^1, \ldots, 1/P_1^D, \ldots, 1/P_K^D)$. Using the several results that are proven in [16, 22], we have

$$E\left[\left(\frac{\partial \ln(L)}{\partial \sigma^2}\right)^2\right] = \frac{MKN}{\sigma^4},$$

$$E\left[\left(\frac{\partial \ln(L)}{\partial \sigma^2}\right)\left(\frac{\partial \ln(L)}{\partial \mathbf{S}(n)}\right)^T\right] = E\left[\left(\frac{\partial \ln(L)}{\partial \sigma^2}\right)\left(\frac{\partial \ln(L)}{\partial \mathbf{P}}\right)^T\right] = 0,$$

$$\mathbf{Q} = E\left[\left(\frac{\partial \ln(L)}{\partial \mathbf{S}(n)}\right)\left(\frac{\partial \ln(L)}{\partial \mathbf{S}(r)}\right)^T\right] = \frac{2}{\sigma^2}\,\text{Re}\{\mathbf{A}^H\mathbf{A}\}\delta_{n,r},$$

$$\mathbf{R}_P = E\left[\left(\frac{\partial \ln(L)}{\partial \mathbf{P}}\right)^2\right] = \frac{1}{2\sigma^2}\sum_{n=1}^{N}\text{Re}\left\{\text{diag}(\mathbf{S}^T(n))\,\text{diag}\left(\frac{1}{\mathbf{P}}\right)\mathbf{A}^H \times \mathbf{A}\,\text{diag}\left(\frac{1}{\mathbf{P}}\right)\text{diag}(\mathbf{S}(n))\right\},$$

$$\mathbf{R}(n) = E\left[\left(\frac{\partial \ln(L)}{\partial \mathbf{S}(n)}\right)\left(\frac{\partial \ln(L)}{\partial \mathbf{P}}\right)^T\right] = \frac{1}{\sigma^2}\,\text{Re}\left\{\mathbf{A}^H\mathbf{A}\,\text{diag}\left(\frac{1}{\mathbf{P}}\right)\text{diag}(\mathbf{S}(n))\right\}. \tag{A.3}$$

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs, NJ, USA, 3rd edition, 1996.

[2] S. A. Grandhi, R. Vijayan, D. J. Goodman, and J. Zander, "Centralized power control in cellular radio systems," *IEEE Trans. Vehicular Technology*, vol. 42, no. 4, pp. 466–468, 1993.

[3] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1341–1347, 1995.

[4] F. Rashid-Farrokhi, L. Tassiulas, and K. J. R. Liu, "Joint optimal power control and beamforming in wireless networks using antenna arrays," *IEEE Trans. Communications*, vol. 46, no. 10, pp. 1313–1324, 1998.

[5] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1437–1450, 1998.

[6] A. Mercado and K. J. R. Liu, "Adaptive QoS for wireless multimedia networks using power control and smart antennas," *IEEE Trans. Vehicular Technology*, vol. 51, no. 5, pp. 1223–1233, 2002.

[7] Y. C. Liang, F. P. S. Chin, and K. J. R. Liu, "Downlink beamforming for DS-CDMA mobile radio with multimedia services," *IEEE Trans. Communications*, vol. 49, no. 7, pp. 1288–1298, 2001.

[8] A. Yener, R. D. Yates, and S. Ulukus, "Interference management for CDMA systems through power control, multiuser detection, and beamforming," *IEEE Trans. Communications*, vol. 49, no. 7, pp. 1227–1239, 2001.

[9] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind channel identification based on second-order statistics: a frequency-domain approach," *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 329–334, 1995.

[10] B. Sampath, K. J. R. Liu, and Y. (G.) Li, "Deterministic blind subspace MIMO equalization," *EURASIP Journal on Applied Signal Processing*, vol. 2002, no. 5, pp. 538–551, 2002.
[11] T. Nguyen and Z. Ding, "Blind CMA beamforming for narrowband signals with multipath arrivals," *International Journal of Adaptive Control and Signal Processing*, vol. 12, pp. 157–172, 1998.
[12] Y. Li and K. J. R. Liu, "Adaptive blind source separation and equalization for multiple-input/multiple-output systems," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2864–2876, 1998.
[13] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array. I. algorithms," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1184–1197, 1996.
[14] S. Talwar and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array. II. performance analysis," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 706–718, 1997.
[15] Z. Ding and Y. (G.) Li, *Blind Equalization and Identification*, vol. 9 of *Signal processing and communications*, Marcel Dekker, New York, NY, USA, 2001.
[16] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communications," *IEEE Transactions on Signal Processing*, vol. 43, no. 11, pp. 2714–2723, 1995.
[17] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, Md, USA, 1996.
[18] A. Mercado and K. J. R. Liu, "NP-hardness of the stable matrix in unit interval family problem in discrete time," *Systems and Control Letters*, vol. 42, no. 4, pp. 261–265, 2001.
[19] G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Transactions on Vehicular Technology*, vol. 42, no. 4, pp. 641–646, 1993.
[20] K. Vissa, F. Rashid-Farrokhi, and K. J. R. Liu, "Combining blind equalization and power control in wireless networks," in *Proc. 32nd Asilomar Conference on Signals Systems, and Computers*, vol. 1, pp. 88–91, Systems, and Computers, Pacific Grove, 1998 (invited paper).
[21] P. Harley, "Short distance attenuation measurements at 900 MHz and 1.8 GHz using low antenna heights for microcells," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 1, pp. 5–11, 1989.
[22] W. M. Steedly and R. L. Moses, "The Cramer-Rao bound for pole and amplitude coefficient estimates of damped exponential signals in noise," *IEEE Transactions on Signal Processing*, vol. 41, no. 3, pp. 1305–1318, 1993.

Zhu Han received the B.S. degree in electronic engineering from Tsinghua University in 1997, and received the M.S. degree in electrical engineering from the University of Maryland, College Park, in 1999. From 1997 to 2000, he was a Graduate Research Assistant. From 2000 to 2002, he was an Engineer in R&D group of ACTERNA. In November 2003, he got his Ph.D. degree from the University of Maryland, College Park, where he is now a Research Associate. His research interests are wireless resource allocation, game theory approach, OFDM, CDMA, multimedia over wireless, blind estimation, and bioimage processing.

Farrokh R. Farrokhi received the B.S. and M.S. degrees (highest honors) in electrical engineering from Sharif University of Technology, Tehran, Iran, in 1988 and 1992, respectively, and the Ph.D. degree in electrical engineering from the University of Maryland at College Park in 1997. From 1998 to 2000, he was a member of the technical staff in the Wireless Research Department, Bell Labs, Lucent Technologies. From 2000 to 2002, he was with the Advanced Research and Development Department, Centillium Communications. Since 2002, he has been with DSP Department, Cognio Inc. His research interests include array and statistical signal processing, wireless communications and networking. Dr. Farrokhi received the 1996–1997 George Harhalakis Outstanding Systems Engineering Graduate Student Award in recognition of outstanding contributions in cross-disciplinary research from the University of Maryland at College Park.

K. J. Ray Liu received the B.S. degree from the National Taiwan University in 1983, and the Ph.D. degree from UCLA in 1990, both in electrical engineering. He is a Professor of Electrical and Computer Engineering Department and Institute for Systems Research of University of Maryland, College Park. His research contributions encompass broad aspects of multimedia communications and signal processing; wireless communications and networking; information security; signal processing algorithms and architectures; and bioinformatics, in which he has published over 300 refereed papers. Dr. Liu is the recipient of numerous honors and awards including the IEEE Signal Processing Society 2004 Distinguished Lecturer, the 1994 National Science Foundation Young Investigator Award, the IEEE Signal Processing Society's 1993 Senior Award (Best Paper Award), IEEE 50th Vehicular Technology Conference Best Paper Award, Amsterdam, 1999. He also received the George Corcoran Award in 1994 for outstanding contributions to electrical engineering education and the Outstanding Systems Engineering Faculty Award in 1996 in recognition of outstanding contributions in interdisciplinary research, both from the University of Maryland. Dr. Liu is a Fellow of IEEE. Dr. Liu is the Editor-in-Chief of IEEE Signal Processing Magazine and was the Founding Editor-in-Chief of EURASIP Journal on Applied Signal Processing. Dr. Liu is a Board of Governor and has served as Chairman of Multimedia Signal Processing Technical Committee of IEEE Signal Processing Society.

EURASIP Journal on Applied Signal Processing 2004:5, 762–771

Approaching the MIMO Capacity with a Low-Rate Feedback Channel in V-BLAST

Seong Taek Chung
STAR Laboratory, Stanford University, Stanford, CA 94305-9515, USA
Email: stchung@dsl.stanford.edu

Angel Lozano
Wireless Research Laboratory, Lucent Technologies, 791 Holmdel-Keyport Road, Holmdel, NJ 07733, USA
Email: aloz@lucent.com

Howard C. Huang
Wireless Research Laboratory, Lucent Technologies, 791 Holmdel-Keyport Road, Holmdel, NJ 07733, USA
Email: hchuang@lucent.com

Arak Sutivong
Information Systems Laboratory, Stanford University, Stanford, CA 94305-9510, USA
Email: arak@stanfordalumni.org

John M. Cioffi
STAR Laboratory, Stanford University, Stanford, CA 94305-9515, USA
Email: cioffi@stanford.edu

Received 8 December 2002; Revised 30 October 2003

This paper presents an extension of the vertical Bell Laboratories Layered Space-Time (V-BLAST) architecture in which the closed-loop multiple-input multiple-output (MIMO) capacity can be approached with conventional scalar coding, optimum successive decoding (OSD), and independent rate assignments for each transmit antenna. This theoretical framework is used as a basis for the proposed algorithms whereby rate and power information for each transmit antenna is acquired via a low-rate feedback channel. We propose the successive quantization with power control (SQPC) and successive rate and power quantization (SRPQ) algorithms. In SQPC, rate quantization is performed with continuous power control. This performs better than simply quantizing the rates without power control. A more practical implementation of SQPC is SRPQ, in which both rate and power levels are quantized. The performance loss due to power quantization is insignificant when 4–5 bits are used per antenna. Both SQPC and SRPQ show an average total rate close to the closed-loop MIMO capacity if a capacity-approaching scalar code is used per antenna.

Keywords and phrases: adaptive antennas, BLAST, interference cancellation, MIMO systems, space-time processing, discrete bit loading.

1. INTRODUCTION

Information theory has shown that the rich-scattering wireless channel can support enormous capacities if the multipath propagation is properly exploited, using multiple transmit and receive antennas [1, 2, 3]. In order to attain the closed-loop multiple-input multiple-output (MIMO) capacity, it is necessary to signal through the channel's eigenmodes with optimal power and rate allocation across those modes [4, 5]. Such an approach requires instantaneous channel information feedback from the receiver to the transmitter, hence a closed-loop implementation. Furthermore, a very specialized transmit structure is required to perform the eigenmode signaling. Therefore, it is challenging to incorporate the closed-loop MIMO capacity-achieving transmit-receive structures into existing systems.

Open-loop schemes that eliminate the need for instantaneous channel information feedback at the transmitter have also been proposed [6, 7, 8, 9, 10, 11]. These schemes can be divided into two categories: multidimensional coding

(e.g., space-time coding) and spatial multiplexing (e.g., vertical Bell Laboratories layered space-time (V-BLAST)). Multidimensional coding [7] requires very specialized coding structures and complicated transceiver structures. Furthermore, its complexity grows very rapidly with the number of transmit antennas. Among spatial multiplexing approaches, V-BLAST [9, 10, 11] uses simple scalar coding and a well-known transceiver structure. This paper focuses on the V-BLAST transmission scheme.

In V-BLAST, every transmit antenna radiates an independently encoded stream of data. This transmission method is much more attractive from an implementation standpoint; the transmitter uses a simple spatial demultiplexer followed by a bank of scalar encoders, one per antenna. The receiver uses a well-known successive detection technique [12]. Furthermore, this scheme is much more flexible in adapting the number of antennas actively used. This flexibility is a strong advantage for the following reasons. First, the channel estimation process requires more time as the number of transmit antennas increases; consequently, the overall spectral efficiency—including training overhead—could actually degrade with an excessive number of transmit antennas in rapidly fading channels. Hence, MIMO systems may need to adapt the number of antennas actively used depending on the environment. Second, it is expected that during initial deployment, not all base stations and terminal units may have the same number of antennas. Therefore, the number of antennas actually being used may need to be adapted, for example, during hand-off processes between different cells.

As previously mentioned, the main weakness of open-loop V-BLAST is that it attains a part of the closed-loop MIMO capacity; as the transmitter cannot adapt itself to the channel environment in an open-loop fashion, V-BLAST simply allocates equal power and rate to every transmit antenna. Consequently, the performance is limited by the antenna with the smallest capacity, as dictated by the channel. Hence, it is natural to consider per-antenna rate adaptation using a low-rate feedback channel.

Using a low-rate feedback channel, [13] introduced rate adaptation at each antenna in V-BLAST to overcome this problem. We extend their approach to both rate and power adaptations at each antenna and theoretically prove that this new scheme, denoted as V-BLAST with per-antenna rate control (PARC), achieves the performance of an open-loop scheme with multidimensional coding. A similar approach was taken at OFDM/SDMA in the downlink of wireless local networks [14]. We show that with per-antenna rate and power control, V-BLAST achieves higher performance than the other open-loop schemes. Moreover, V-BLAST with PARC attains the open-loop MIMO capacity.

In developing the optimal PARC, similarities are noted between the V-BLAST with PARC and the Gaussian multiple-access channel (GMAC) problems. Every transmit antenna within the V-BLAST can be regarded as an individual user in a GMAC. As shown in [15], with optimum successive decoding (OSD), the total sum capacity of the GMAC can be achieved at any corner point of the capacity region. As will be shown, this result translates directly to the V-BLAST context by simply incorporating the notion of PARC.

Next, these theoretical results are applied to practical modulation scenarios. In order to apply the idealized capacity results to a real system, the following points should be considered. First, the idealized results assume an infinite-length codebook to achieve vanishingly small bit error rates (BERs), but in a real system, current coding techniques and practical system requirements allow only for a finite-length coding with nonzero error rates [16]. Second, the idealized results assume a continuous rate set, but in a real system, only rates from a discrete rate set are feasible.

The first issue can be easily solved by adopting the concept of a gap (Γ) [17]:

$$b = \log_2\left(1 + \frac{\mathrm{SINR}}{\Gamma}\right). \tag{1}$$

The number of bits transmitted at a specific SINR and specific coding and BER can be expressed as (1), where b is the number of bits transmitted per symbol, SINR is the signal-to-interference-and-noise ratio, and Γ is a positive number larger than 1, which is a function of the BER and specific coding method. Note that this is a capacity expression, except that the SINR is scaled by a penalty Γ, which is a function of the target BER and coding method. Γ can take various values; for uncoded M-QAM with the target BER 10^{-3}, Γ is 3.333 (5.23 dB). For a very powerful code (e.g., Turbo code), Γ is close to 1 (0 dB). When Γ equals 0 dB, the gap expression (1) equals the actual capacity [17]. Works in [13] also utilize the gap expression in considering the rate adaptation per antenna.

The second issue is investigated using ad hoc methods since the optimal solution for discrete rates is difficult to obtain analytically. Successive quantization with power control (SQPC) is first proposed. Here, the rate is quantized efficiently with continuous power control. However, a continuously variable transmit power level can be impractical since the feedback channel data rate is limited. Therefore, SQPC is extended to successive rate and power quantization (SRPQ) by considering power level quantization as well.

The organization of this paper is as follows. The system model is introduced in Section 2. V-BLAST is specifically described in Section 3, with optimal PARC, when the transmit antenna powers are given. The antenna power allocation that maximizes the capacity is derived in Section 4. Section 5 shows that the open-loop capacity can be approached using V-BLAST with equal power allocation; additional power control only leads to a slight increase in capacity. Section 6 first suggests a simple discrete bit loading algorithm based on rounding off the rate from a continuous set with equal power allocation. Then, a new discrete bit loading is presented along with continuous power control, SQPC, in Section 7. In Section 8, a discrete bit loading with quantized power levels, SRPQ is suggested. Results are shown in Section 9. Conclusions follow in Section 10.

2. SYSTEM MODEL

We assume a general architecture with M transmit and N receive antennas and perfect channel estimation at the receiver. Rate and/or power information can be fed back to the transmitter. The $M \times 1$ transmit signal vector is $\mathbf{x}$; the $N \times 1$ received signal vector is $\mathbf{y}$. The $N \times M$ channel matrix $\mathbf{H}$ can take any value; however, for a rich scattering environment, we assume that $\mathbf{H}$ is composed of independent zero-mean complex Gaussian random variables. The zero-mean additive white Gaussian noise (AWGN) vector at the receiver, denoted by $\mathbf{n}$, has a covariance matrix equal to the identity matrix scaled by σ^2. For simplicity, we assume $\sigma^2 = 1$ and scale the channel appropriately. The average power of each component of the $\mathbf{H}$ matrix is indicated by g, while the total power available to the transmitter is denoted by P_T. An average SNR ρ is defined as $P_T g$.

This model can be expressed mathematically as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{2}$$

where $E[\mathbf{n}\mathbf{n}^H] = \mathbf{I}_N$ and $E[H(n_1, m_1)^* H(n_2, m_2)] = g\delta(n_1 - n_2, m_1 - m_2)$ for all n_1, n_2, m_1, and m_2. $\mathbf{I}_N$ denotes the identity matrix of size $N \times N$, $\delta(m, n)$ denotes the 2-dimensional Kronecker delta function, and $H(n, m)$ indicates the nth row and mth column element of the $\mathbf{H}$ matrix. Consistent with the open-loop V-BLAST concept, the signals radiated from different antennas are independent. Hence, the covariance matrix of $\mathbf{x}$ can be expressed as follows when the power allocated to antenna m is equal to P_m:

$$E[\mathbf{x}\mathbf{x}^H] = \begin{bmatrix} P_1 & 0 & \cdots & 0 & 0 \\ 0 & P_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & P_{M-1} & 0 \\ 0 & 0 & \cdots & 0 & P_M \end{bmatrix}, \tag{3}$$

where $\sum_{m=1}^{M} P_m = P_T$. When we simply allocate equal power to all the transmit branches, we assign $P_m = P_T/M$. We use $(\cdot)^T$ and $(\cdot)^H$ to denote transposition and Hermitian transposition, respectively. For scalars, $(\cdot)^*$ denotes complex conjugate.

3. V-BLAST WITH PARC

With respect to minimum mean square error (MMSE) V-BLAST, the natural extension is PARC, which is explained in detail below.

The capacity of the mth transmit antenna C_m can be expressed in terms of the channel matrix and the transmit power of each antenna. We define $\mathbf{h}_m$ as the mth column of $\mathbf{H}$ and $\mathbf{H}(m)$ $(m = 1, \ldots, M)$ as the $N \times (M - m + 1)$ matrix $[\mathbf{h}_m \mathbf{h}_{m+1} \cdots \mathbf{h}_{M-1} \mathbf{h}_M]$. We also define $\mathbf{P}(m)$ as an $(M - m + 1) \times (M - m + 1)$ diagonal matrix with $(P_m, P_{m+1}, \ldots, P_{M-1}, P_M)$ along the diagonal.

According to the OSD procedure described in [15], the signals radiating from the M transmit antennas are decoded in any agreed-upon arbitrary order. In the remainder, it is assumed, without loss of generality, that they are decoded according to their index order. It is interesting to note that, unlike the open-loop V-BLAST, the ordering has no impact on the capacity attained by the sum of all M antennas.[1] It does, however, impact the fraction of that capacity that is allocated through rate adaptation to each individual antenna. It also affects the total rate when both rate and power are quantized.

The process is parameterized by a set of projection vectors $\mathbf{F}_m$ $(m = 1, \ldots, M)$ and cancellation vectors $\mathbf{B}_{m1}$, $\mathbf{B}_{m2}$, $\ldots$, $\mathbf{B}_{mm}$ $(m = 1, \ldots, M - 1)$, all with a dimension of $N \times 1$. In decoding the mth transmit antenna signal, interference from the $(m - 1)$ already decoded signals is subtracted from $\mathbf{y}$ by applying the proper cancellation vectors to reencoded versions of their decoded symbols. An inner product of that cancellation process result and the projection vector corresponding to the mth antenna is fed into the mth antenna decoder.

The first antenna, in particular, is decoded based on Z_1, which is obtained as the inner product of $\mathbf{F}_1$ and the receive vector $\mathbf{Y}_1 = \mathbf{y}$ expressed as $Z_1 = \langle \mathbf{F}_1, \mathbf{Y}_1 \rangle = {\mathbf{F}_1}^H \mathbf{Y}_1$. The decoded bits are reencoded to produce $\hat{x_1}$. The second antenna is similarly decoded based on Z_2, where Z_2 is now the inner product of $\mathbf{F}_2$ and a vector $\mathbf{Y}_2$ obtained by subtracting the vector $\mathbf{B}_{11}\hat{x_1}$ from $\mathbf{y}$. Therefore, $\mathbf{Y}_2 = \mathbf{y} - \mathbf{B}_{11}\hat{x_1}$ and $Z_2 = \langle \mathbf{F}_2, \mathbf{Y}_2 \rangle$. In general, the mth antenna is decoded based on $Z_m = \langle \mathbf{F}_m, \mathbf{Y}_m \rangle = \mathbf{F}_m^H (\mathbf{y} - \sum_{j=1}^{m-1} \mathbf{B}_{(m-1)j} \hat{x_j})$. Here, it is assumed that all decoded bits are error-free, which is legitimate in the analysis of capacity [16].

The optimal cancellation vectors are given by $\mathbf{B}_{(m-1)j} = \mathbf{h}_j$, and the optimal projection vectors are $\mathbf{F}_m = (\mathbf{H}(m + 1)\mathbf{P}(m + 1)\mathbf{H}(m + 1)^H + \mathbf{I}_N)^{-1}\mathbf{h}_m$ [15].

Furthermore, the capacity of the mth antenna can be expressed as

$$C_m = \log_2 \left(1 + P_m {\mathbf{h}_m}^H (\mathbf{H}(m+1)\mathbf{P}(m+1) \times \mathbf{H}(m+1)^H + \mathbf{I}_N)^{-1} \mathbf{h}_m\right) \quad (m = 1, \ldots, M), \tag{4}$$

and it was proved in [15] that

$$\sum_{m=1}^{M} C_m = \log_2 \det (\mathbf{I}_N + \mathbf{H} E[\mathbf{x}x^H] \mathbf{H}^H), \tag{5}$$

which, with equal power per antenna, is precisely the open-loop MIMO capacity attainable with multidimensional coding [1]. Hence, the same capacity can be achieved using scalar coding, but at the expense of rate adaptation using a low-rate feedback channel. For a practical coding scheme with a nonzero BER, the rate R_m is expressed as follows, using (1)

[1] It should be emphasized that this is true only in a capacity sense. In practice, due to error propagation, error rate performances can differ depending on the ordering.

and (4):

$$R_m = \log_2\left(1+\frac{P_m \mathbf{h}_m{}^H(\mathbf{H}(m+1)\mathbf{P}(m+1)\mathbf{H}(m+1)^H+\mathbf{I}_N)^{-1}\mathbf{h}_m}{\Gamma}\right) \quad (m=1,\ldots,M) \tag{6}$$

It is interesting to note that as the number of antennas grows large, the capacities C_m become increasingly predictable from the statistics of the channel, and hence the feedback need for each transmit antenna actually vanishes progressively [18].

4. POWER CONTROL IN V-BLAST WITH PARC

In this section, the power P_m $(m = 1,\ldots,M)$ allocation methods are considered under the total power constraint. For any set of powers $P_m(m = 1,\ldots,M)$, the optimal capacity and rate are those given by (4) and (6). The optimal power allocation scheme here is different from the waterfilling solution in [4].

4.1. *Optimal scheme for* $N = 1$ *or* $N = 2$

The optimal power control was found only when the number of receive antennas is 1 or 2. The optimal power allocation for more extensive cases was independently derived in [19].

When N=1, the open-loop MIMO capacity can be expressed as

$$C = \log_2\left(\sum_{m=1}^{M} P_m |\mathbf{h}_m|^2 + 1\right), \tag{7}$$

where $\mathbf{h}_m$ is a scalar. Under the total power constraint, the optimal power allocation corresponds to assigning the entire power budget to the transmit antenna with the largest $|\mathbf{h}_m|$.

When N=2, following (5), the open-loop MIMO capacity can be expressed as

$$\begin{aligned} C = \log_2\Bigg[&\left(\sum_{m=1}^{M} P_m |H(1,m)|^2 + 1\right)\\ &\left(\sum_{m=1}^{M} P_m |H(2,m)|^2 + 1\right)\\ &-\left(\sum_{m=1}^{M} P_m H(1,m)^* H(2,m) + 1\right)\\ &\times\left(\sum_{m=1}^{M} P_m H(2,m)^* H(1,m) + 1\right)\Bigg]. \end{aligned} \tag{8}$$

Under the total power constraint, the optimal power allocation can be found using a Lagrangian method:

$$\begin{aligned} &J(P_1,\ldots,P_M)\\ &= \left(\sum_{m=1}^{M} P_m |H(1,m)|^2 + 1\right)\left(\sum_{m=1}^{M} P_m |H(2,m)|^2 + 1\right)\\ &\quad -\left(\sum_{m=1}^{M} P_m H(1,m)^* H(2,m) + 1\right)\\ &\quad\times\left(\sum_{m=1}^{M} P_m H(2,m)^* H(1,m) + 1\right) + \lambda\left(\sum_{m=1}^{M} P_m - P_T\right), \end{aligned} \tag{9}$$

where $J(P_1,\ldots,P_M)$ is convex with respect to P_m. The optimal power allocation should satisfy the Karush-Kuhn-Tucker condition [20]; if the optimal power allocation P_m is positive for all $m = 1,\ldots,M$, then the optimal power assignment policy is found from $\partial J/\partial P_l = 0$ $(l = 1,\ldots,M)$ and the total power constraint. $\partial J/\partial P_l = 0$ becomes

$$\begin{aligned} &\sum_{m=1}^{M} P_m |H(1,l)H(2,m) - H(1,m)H(2,l)|^2\\ &\quad = \lambda - |H(1,l)|^2 - |H(2,l)|^2 \quad (l = 1,\ldots,M). \end{aligned} \tag{10}$$

If some P_m's are zero in the optimal power allocation, then $\partial J/\partial P_l$ should be zero only for the nonzero P_l's and the total power constraint should be satisfied. By checking this condition numerically, the optimal power allocation can be found. Simulation results are shown in Section 5.

4.2. *Suboptimal scheme for* $N > 2$

We were not able to find the optimal power and rate allocations when the number of receive antennas is more than 2. By solving the nth-order linear equations, we can get the optimal power solution, but obtaining a closed form, even for $N = 3$, is extremely complicated. However, from the optimal solution for $N = 1$ and $N = 2$, we observe the following:

(i) the optimal power allocation scheme usually corresponds to selecting 1 or 2 antennas while switching off the remaining ones completely;

(ii) with suboptimal power allocations (e.g., equal-power allocation), the capacity loss is small.

Based on these observations, we suggest a suboptimal power allocation algorithm that works for any combination of M and N. First, divide the total power P_T by M and consider P_T/M as a power unit. There are M such power units. Then, consider every possible power unit distribution over antennas, calculate the sum capacity (5) of each distribution, and select the one that yields the largest sum capacity of all the distributions.

5. CAPACITY RESULTS

Numerical values for the capacity are shown in this section. Equation (1) is equivalent to the capacity formula for two dimensions when the gap (Γ) is 0 dB. The average (ergodic)

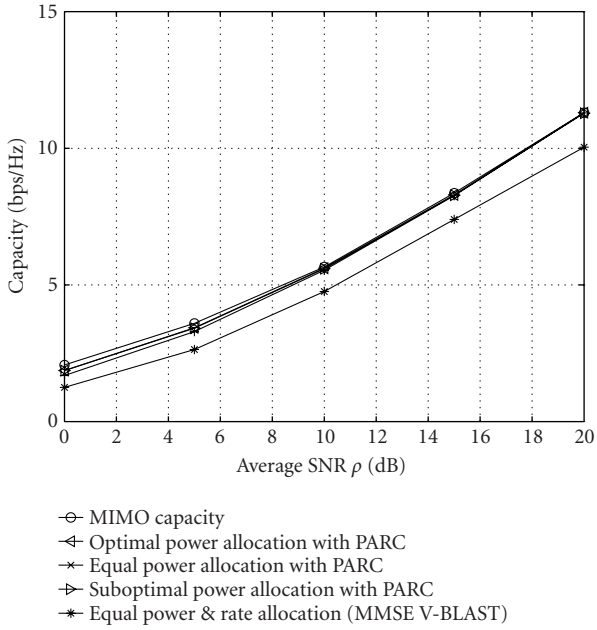


Figure 1: Average capacity when $M = 2$ and $N = 2$.

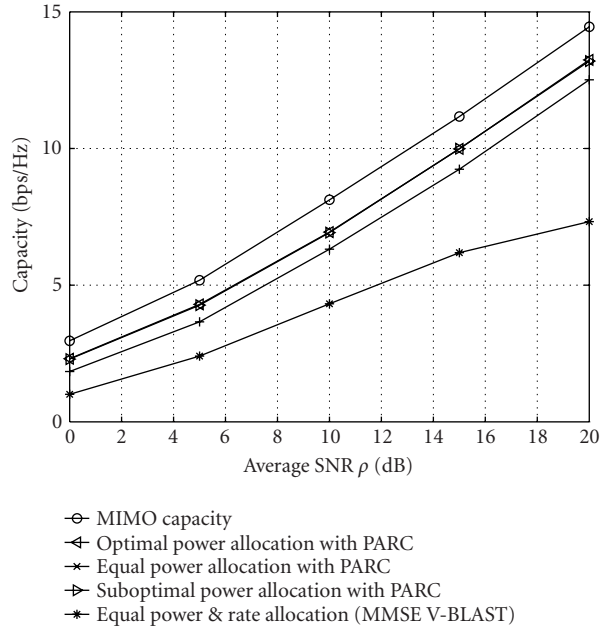


Figure 2: Average capacity when $M = 4$ and $N = 2$.

capacity is used as a performance measure. We have also tested the outage capacity at small levels of outage, which shows a performance trend similar to that of the average capacity. Hence, the outage capacity results are not shown here.[2] Figures 1, 2, and 3 show such average capacity for various combinations of M and N. For each combination, the following cases are depicted: MIMO capacity, optimal power allocation with PARC, equal power allocation with PARC, suboptimal power allocation with PARC, and equal power and equal rate allocation. The MIMO capacity is the maximum rate achievable by transmitting over the channel eigenmodes when both the transmitter and the receiver know the channel matrix [4]. In other words, the MIMO capacity here is the closed-loop MIMO capacity. Furthermore, the spectral efficiency of equal power allocation with PARC is equal to the open-loop MIMO capacity.

In a moderate to high SNR regime, equal power allocation across antennas works almost as well as the optimal (or suboptimal) power allocation as long as the rate is controlled under OSD. Hence, power adaptation becomes largely irrelevant with PARC in a moderate to high SNR region. However, in a low SNR region, it is observed that power allocation improves the capacity. This is in line with conclusions drawn in other research literatures in similar cases. In a single user time-varying channel, a close-to-optimal performance is achieved by transmitting a constant power when the channel path gain is larger than a certain threshold value [21].

[2]In general, unless all the schemes produce the same probability density function of achievable capacity, the outage capacity does not follow the same trend as the average capacity.

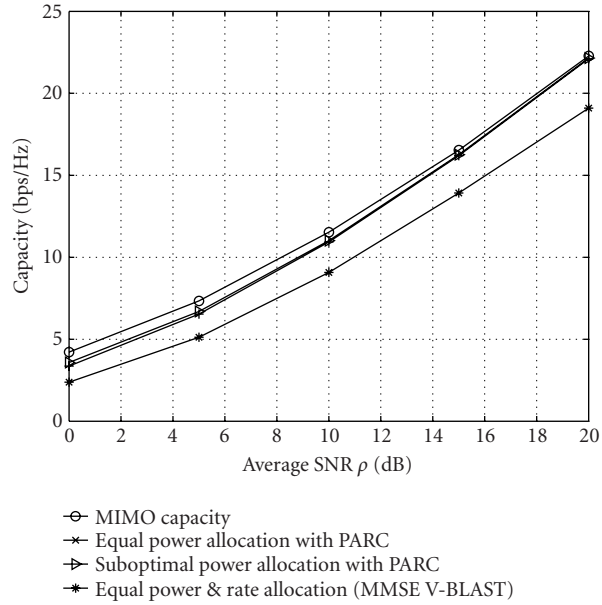


Figure 3: Average capacity when $M = 4$ and $N = 4$.

Results also show that the capacity loss relative to the closed-loop MIMO capacity is not significant (except in Figure 2, where the gap between MIMO capacity and equal-power capacity is not reduced even though we increase the average SNR). Therefore, equal power allocation combined with PARC under OSD is a practical and efficient method to approach the MIMO capacity. All the schemes proposed in

this paper perform better than the equal power and rate allocation (MMSE) V-BLAST.[3]

6. SIMPLE ROUNDING-OFF

Here, a simple, discrete bit loading algorithm is proposed. Given that PARC under equal power allocation achieves the open-loop MIMO capacity as seen in (5), a natural practical extension is to simply round off each rate per antenna with equal power allocation. Here, it is assumed that all the decisions are correct during OSD process.

Given the rate R_m as described in (6), round off R_m and assign the rounded-off rate $\overline{[R_m]}$, where $\overline{[x]}$ is the largest integer which is smaller than or equal to x. The rate set can be reduced further by considering only every qth integer. In this case, the rounded-off rate is $q\overline{[R_m/q]}$. This quantization method does not limit the maximum rate used, but simulation results in Section 9 show that the maximum rate per antenna calculated with this algorithm is less than or equal to 16 QAM when an average SNR is 10 dB. Hence, clipping in quantization is not considered.

As there is no power control, this is simpler than the following two schemes. However, unlike in the continuous rate case, results in Section 9 show that the spectral efficiency loss is significant when power is not adapted.

7. SUCCESSIVE QUANTIZATION WITH POWER CONTROL

A more efficient discrete bit loading algorithm is proposed by also adapting the power levels at each transmit antenna. Obviously, the performance is maximized by using optimal power control under the assumption that discrete rates are available at each transmit antenna. However, a closed-form solution for the optimal discrete rate and continuous power control cannot be found analytically; furthermore, an exhaustive search over the set of rate and power levels is too complicated to be conducted in real time. Hence, instead of the optimal rate and power control scheme, an ad-hoc discrete bit loading method, successive quantization with power control (SQPC) (Figure 4), is suggested in the following. Here also all the decodings are assumed perfect in OSD.

The transmit antennas are labeled according to the order in which they are decoded at the receiver. The SINR of the kth transmit antenna contains interference from all the antennas decoded after it (i.e., $k+1,\ldots,M$). The available rates are assumed to be $0, q, 2q, 3q$, and so on. Therefore, q is the interval between rate quantization levels. Again, there is no clipping; from numerical calculations, the maximum rate per antenna is less than or equal to 16 QAM when an average SNR is 10 dB.

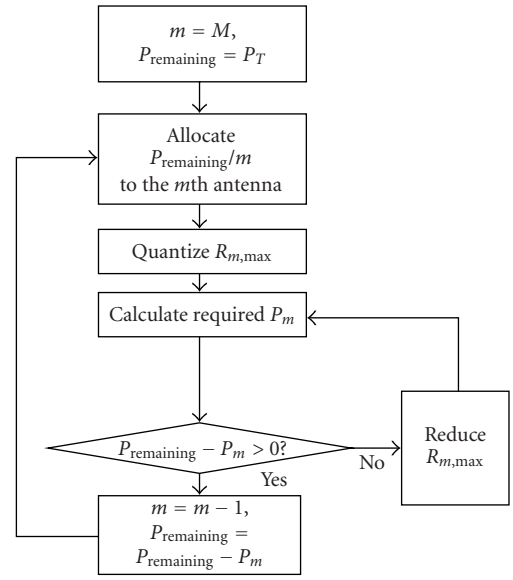


Figure 4: SQPC algorithm.

First, the power and rate for the Mth antenna are allocated. The rate of this Mth antenna is independent of the power of all other antennas. P_T is divided by M and then assigned as the transmit power of the Mth antenna. Then, we calculate the maximum rate $R_{M,\max}$ possible for $P_M = P_T/M$ from (6). Next, round $R_{M,\max}$ and recalculate how much P_M is needed to support rounded $R_{M,\max}$ from (6). Here "round x" means $q\{x/q\}$, where $\{x\}$ means the integer closest to x. If that power exceeds P_T, then subtract q from $R_{M,\max}$. Then, recalculate how much power is necessary to support the reduced $R_{M,\max}$ from (6).

Second, the power and rate for the $(M-1)$th antenna are allocated. Given the interference due to the Mth antenna from the previous stage, calculate the maximum rate for the $(M-1)$th antenna, assuming $(P_T - P_M)/(M-1)$ is allocated as the transmit power of the $(M-1)$th antenna. Round $R_{M-1,\max}$ and recalculate how much P_{M-1} is needed to support this rounded $R_{M-1,\max}$. If $(P_M + P_{M-1})$ exceeds P_T, then subtract q from $R_{M-1,\max}$ and recalculate P_{M-1} which can support the reduced $R_{M-1,\max}$.

Iteratively, at step j $(j < M-1)$, the power and rate for the $(M-j)$th antenna are determined. The exact amount of interference from $M, M-1,\ldots,(M-j+1)$th antennas is known at this stage. Calculate the maximum rate for the $(M-j)$th antenna, $R_{M-j,\max}$, assuming $(P_T - (P_M + P_{M-1} \cdots + P_{M-j+1}))/(M-j)$ is allocated as the transmit power of the $(M-j)$th antenna. Round $R_{M-j,\max}$ and calculate the new P_{M-j} which can support rounded $R_{M-j,\max}$. If $(P_M + P_{M-1} + \cdots + P_{M-j})$ exceeds P_T, then reduce $R_{M-j,\max}$ by q and find the new P_{M-j} which can support the reduced $R_{M-j,\max}$.

At step $M-1$, where the power and rate for the first antenna are determined, $R_{1,\max}$ is calculated, assuming $(P_T - (P_M + P_{M-1} \cdots + P_2))$ is allocated as the transmit power of

[3] Equal power and rate allocation should be interpreted carefully. This is achieved when a codebook designer knows the channel and then allocates equal power and rate across the antennas. However, in practice, MMSE V-BLAST is designed without any prior knowledge regarding the channel. Therefore, one MMSE V-BLAST can achieve one point on the curve not the entire curve.

the 1st antenna. Round off $R_{1,\max}$ and recalculate a new P_1 which can support rounded-off $R_{1,\max}$. Here, rounding up is not an option since it would violate the power budget.

SQPC will inherently leave some part of the total power P_T unused. This residual power is not sufficient to increase the rate of any antenna to the next higher quantized level.

8. SUCCESSIVE RATE AND POWER QUANTIZATION

SQPC in Section 7 can become infeasible, especially when frequent rate and power level updates are necessary. As power levels still assume infinite precision, frequent power level updates cannot be supported due to a limited data rate on the feedback channel. Here, we look into the case in which both rate and power are adapted, while limiting the number of available rate and power levels. Here also, a closed-form solution for the optimal discrete rate and discrete power control does not exist; again, an exhaustive search over the set of rates and powers is too complicated to be conducted in real time. Hence, an ad hoc suboptimal discrete bit loading, successive rate and power quantization (SRPQ) (Figure 5), is also suggested as follows. Here also, all the decoding stages are assumed perfect during OSD.

We use the same notation for the antenna labeling and the achievable rates as in Section 7. Furthermore, the available transmit power levels are 0, $P_T/(N_P-1)$, $2P_T/(N_P-1),\ldots,$ and P_T, where N_P is the number of available transmit power levels. In SQPC, only rate per antenna was quantized while the power levels could take any continuous values.

First, the power and rate for the Mth antenna branch are allocated. P_T is divided by M and then assigned to the Mth branch. Then, the maximum rate $R_{M,\max}$ possible is calculated for $P_M = P_T/M$ from (6). Next, round $R_{M,\max}$ and recalculate how much P_M is needed to support rounded $R_{M,\max}$ from (6). Then round up P_M considering the number of power levels available. In other words, P_M is updated as $q_p[P_M/q_p]$, where $q_p = P_T/(N_P-1)$ and $[x]$ means the integer closest to and larger than x. Round-off is not an option since it would ruin the reliability according to (1). If that power exceeds P_T, then subtract q from $R_{M,\max}$. Recalculate how much power is required to support the reduced $R_{M,\max}$ from (6). Then round up P_M so that P_M can take one of N_P transmit power levels as before. If this P_M still violates the power budget, subtract q from $R_{M,\max}$ again and repeat the process until the power budget is satisfied.

Second, the power and rate for the $(M-1)$th antenna are allocated. Given the interference due to the Mth antenna from the previous stage, calculate the maximum rate for the $(M-1)$th antenna while assuming that $(P_T-P_M)/(M-1)$ is allocated as the transmit power of the $(M-1)$th antenna. Round $R_{M-1,\max}$ and recalculate how much P_{M-1} we need to support this rounded $R_{M-1,\max}$. Then round up P_{M-1} so that P_{M-1} can take one of N_P transmit power levels. If (P_M+P_{M-1}) exceeds P_T, then subtract q from $R_{M-1,\max}$ and recalculate the smallest P_{M-1} which is among the available N_P power levels and can support reduced $R_{M-1,\max}$. If the power budget cannot be satisfied, keep reducing $R_{M-1,\max}$ by q until the power budget is satisfied.

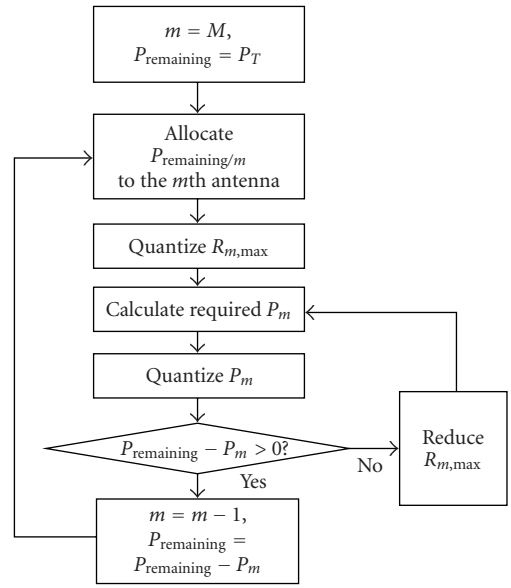


Figure 5: SRPQ algorithm.

Iteratively, at step j $(j < M-1)$, the power and rate for the $(M-j)$th antenna branch are allocated. The exact amount of interference from $M, M-1,\ldots,(M-j+1)$th antenna branches is known. Calculate the maximum rate for the $(M-j)$th antenna branch, $R_{M-j,\max}$, assuming $(P_T-(P_M+P_{M-1}\cdots+P_{M-j+1}))/(M-j)$ is allocated as the transmit power of the $(M-j)$th branch. Round $R_{M-j,\max}$ and calculate new P_{M-j} which is one of the available N_P power levels and can support rounded $R_{M-j,\max}$. If $(P_M+P_{M-1}+\cdots+P_{M-j})$ exceeds P_T, then reduce $R_{M-j,\max}$ by q and find a new P_{M-j} which is one of the available N_P power levels and can support reduced $R_{M-j,\max}$. If the power budget is not satisfied, keep reducing $R_{M-j,\max}$ and calculate appropriate P_{M-j}.

At step $M-1$, where the power and rate for the first antenna are decided, the maximum rate $R_{1,\max}$ is calculated assuming that $(P_T-(P_M+P_{M-1}\cdots+P_2))$ is allocated as the transmit power of first branch. Round off $R_{1,\max}$ and recalculate a new P_1, which is one of the available N_P power levels and can support rounded $R_{1,\max}$. If the power budget is not satisfied, keep reducing $R_{1,\max}$ and calculate appropriate P_1. Here, rounding up is not an option since it would definitely violate the power budget.

Several variations are shown in the following subsections. The first one is a variation in which residual power is used efficiently to reduce error propagation, while the second one is a variation in which an efficient decoding order is found.

8.1. SRPQ1: efficient use of residual power

SRPQ inherently leaves some part of the total power P_T unused. This residual power is not sufficient to increase the rate of any antenna to the next higher quantized level. However, this residual power can be used efficiently to reduce the error rate. Therefore, by pouring residual power into the first antenna, which is decoded first, its BER performance can be improved. This reduction in BER, in turn, helps improve the

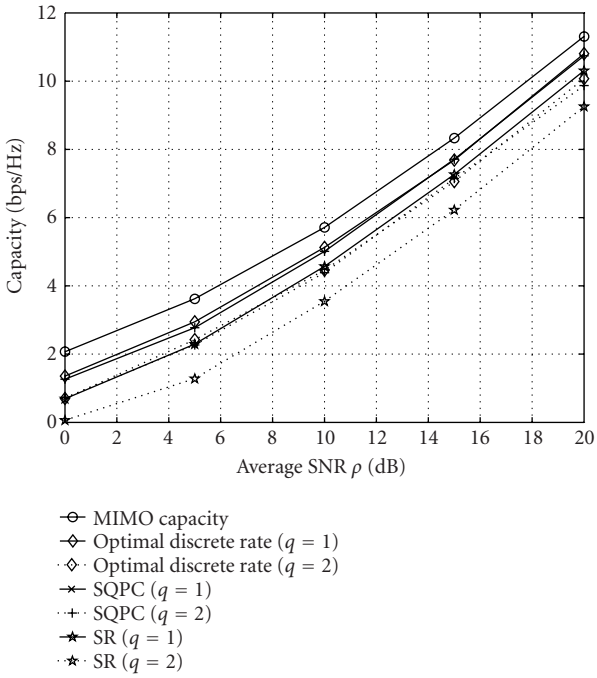


Figure 6: Effect of rate quantization when $M = N = 2$.

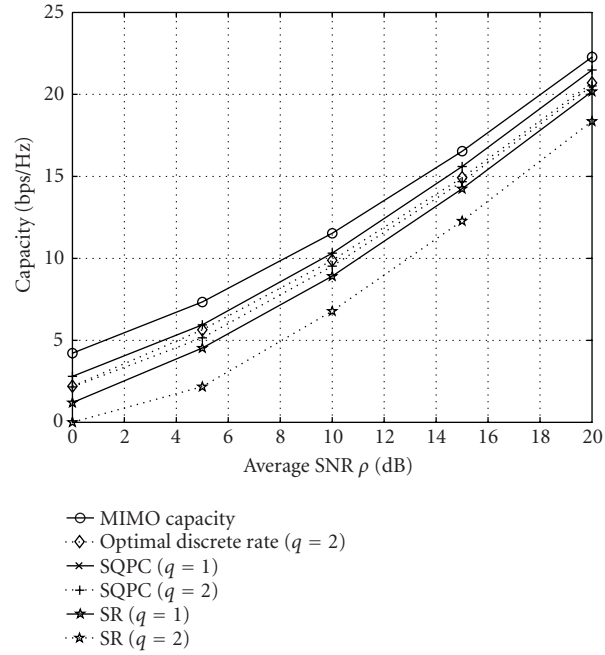


Figure 7: Effect of rate quantization when $M = N = 4$.

decoding reliability at later stages. Pouring all the residual power into the first antenna does not increase the feedback channel rate, even though P_1 is not within the N_P possible power levels since P_1 equals $(P_T - \sum_{m=2}^{M} P_m)$, which can be calculated at the transmitter once P_m $(2 \le m \le M)$ are fed back. This variation of the SRPQ scheme is called SRPQ1.

8.2. SRPQ2: efficient decoding order

So far, the decoding order has been chosen arbitrarily. In a capacity sense, it was proved that the same total rate is achieved regardless of the decoding order. However, for the quantized rate power case, it is unclear whether the optimization of decoding order is helpful or not. Here, a decoding order is optimized by doing a full search over all possible decoding orders. This variation of SRPQ scheme is called SRPQ2.

9. RESULTS

The following schemes are considered: MIMO Capacity, SR, SQPC, SRPQ1, and SRPQ2. The MIMO capacity is the closed-loop MIMO capacity as in Section 5. For each average SNR ρ, H is generated 1000 times and the average capacity is calculated assuming that a scalar capacity-achieving code is used: $\Gamma = 1$ at (1). First, the effect of rate quantization is investigated; later, power quantization is also considered.

9.1. Effect of rate quantization levels

When q is equal to 1, both square and cross QAM (0 bits/symbol, 1 bit/symbol, 2 bits/symbol, and so on) are allowed as a signal constellation. On the other hand, when q is equal to 2, only square QAM (0 bits/symbol, 2 bits/symbol, 4 bits/symbol, and so on) is allowed. For each q, optimal discrete rate is the case in which the spectral efficiency is maximized under a total power constraint when only discrete rates $(0, q, 2q, \ldots)$ are available per antenna. In Figures 6 and 7, the average capacity is displayed as function of the quantization levels. When the power on each transmit antenna is not adapted at all (SR case), using a smaller number of discrete rate levels ($q = 2$) results in poor performance compared with using a larger number of discrete rate levels ($q = 1$). However, in other schemes (SQPC, optimal discrete rate), the performance difference is not significant between $q = 1$ and $q = 2$. The trade-off between feedback information and performance is observed; power levels at each antenna in SR do not need to be fed back. However, more rate levels (smaller q) need to be fed back for SR than for SQPC in order to achieve the same performance level. Hence, it is concluded that $q = 2$ is a reasonable quantization level choice, where power control is also available.

9.2. Effect of power quantization levels

In this section, q is assumed to be 2 and the capacities of the various schemes are compared, depending on the power quantization levels. In Figures 8 and 9, SQPC always performs better than SR for the same M, N, and q. Furthermore, the performance gap increases with M and N. Moreover, for low SNR, the capacity of SQPC falls short of the MIMO capacity by 4 dB in SNR when $q = 2$. Due to space limitations, the result for $q = 1$ cannot be presented, but in this case, the performance of SQPC is less than the MIMO capacity by 3 dB in SNR.

For a low average SNR ρ, a small number of power levels does not degrade the performance significantly from a large number of power levels. The reason is that, for a low SNR, usually only a single antenna is activated. However, for a high

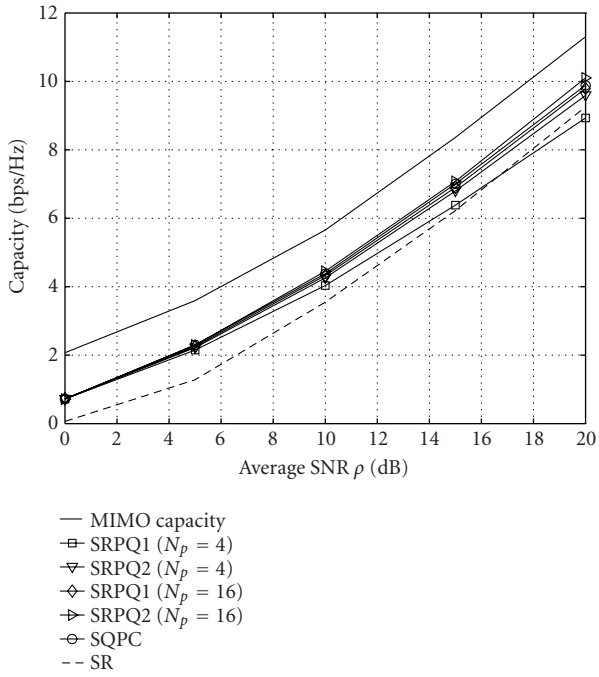


Figure 8: Average capacity when $M = 2$ and $N = 2$ for $q = 2$.

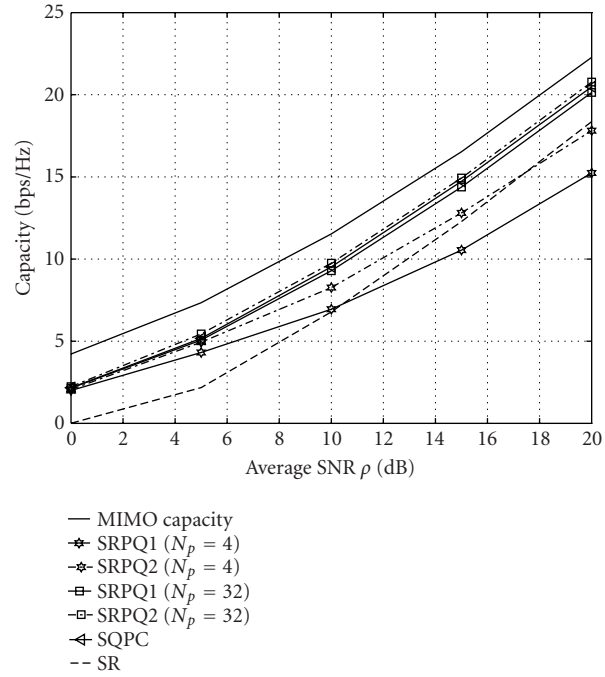


Figure 9: Average capacity when $M = 4$ and $N = 4$ for $q = 2$.

average SNR ρ, the performance loss is considerable as the number of power levels is decreased. Indeed, when $N_p \leq 4$, the degradation caused by power control quantization becomes so great that it is better not to do power allocation at all, since SR scheme outperforms both SRPQ1 and SRPQ2.

Our results suggest that $N_p = 16$ and $N_p = 32$ for $M = N = 2$ and $M = N = 4$, respectively, result in minimal degradation compared to the scheme in which continuous power is allowed. Moreover, this choice of N_p leads to only 2 dB away from the MIMO capacity if a capacity achieving scalar coding is used. Finally, as can be seen, SRPQ2 outperforms SRPQ1 in terms of spectral efficiency. This shows that the decoding order indeed matters when continuous rate and power cannot be used.

10. CONCLUSIONS

This paper proposes an extension of V-BLAST in which the MIMO capacity is approached closely with rate and/or power control using scalar coding with successive interference cancellation. Two practical discrete bit loading algorithms are proposed: SQPC and SRPQ. Simulation results show that power control is necessary, especially in a low SNR regime. Furthermore, it is shown that 4 or 5 bits are sufficient for power quantization levels in order to sustain a similar spectral efficiency to that achieved by continuous power levels.

ACKNOWLEDGMENT

This paper was presented in part at the IEEE Vehicular Technology Conference (VTC) Fall 2001 and the 2002 IEEE Wireless Communications and Networking Conference (WCNC).

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[2] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1273–1289, 2001.

[3] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Transactions on Communications*, vol. 51, no. 3, pp. 389–399, 2003.

[4] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," Tech. Rep., AT & T Bell Laboratories Internal Technical Memorandum, 1995.

[5] M. A. Khalighi, J. Brossier, G. Jourdain, and K. Raoof, "Water filling capacity of Rayleigh MIMO channels," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '01)*, vol. 1, pp. 155–158, San Diego, Calif, USA, September 2001.

[6] G. J. Foschini, "Layered space-time architecture for wireless communication in fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[7] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.

[8] F. R. Farrokhi, G. J. Foschini, A. Lozano, and R. A. Valenzuela, "Link-optimal BLAST processing with multiple-access interference," in *IEEE 52nd Vehicular Technology Conference (VTC '00)*, vol. 1, pp. 87–91, Boston, Mass, USA, September 2000.

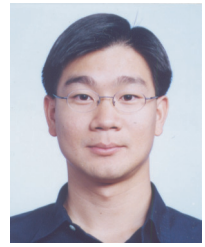
[9] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. of URSI International Symposium on Signals, Systems, and Electronics (ISSSE '98)*, pp. 295–300, Pisa, Italy, September 1998.

[10] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 11, pp. 1841–1852, 1999.
[11] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electronics Letters*, vol. 35, no. 1, pp. 14–15, 1999.
[12] S. Verdu, *Multiuser Detection*, Cambridge University Press, New York, NY, USA, 1998.
[13] S. Catreux, P. F. Driessen, and L. J. Greenstein, "Data throughputs using multiple-input multiple-output (MIMO) techniques in a noise-limited cellular environment," *IEEE Transactions on Wireless Communications*, vol. 1, no. 2, pp. 226–235, 2002.
[14] S. Thoen, L. Van der Perre, B. Gyselinckx, M. Engels, and H. De Man, "Adaptive loading in the downlink of OFDM/SDMA-based wireless local networks," in *IEEE 51st Vehicular Technology Conference Proceedings (VTC '00)*, vol. 1, pp. 235–239, Tokyo, Japan, May 2000.
[15] 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. Asilomar Conf. on Signals, Systems and Computers*, vol. 1, pp. 1405–1409, Monterey, Calif, USA, November 1997.
[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1991.
[17] T. Starr, J. M. Cioffi, and P. J. Silverman, *Understanding Digital Subscriber Line Technology*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 1999.
[18] A. Lozano, "Capacity-approaching rate function for layered multiantenna architectures," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 616–620, 2003.
[19] K. J. Hwang and K. B. Lee, "Transmit power allocation with small feedback overhead for multiple antenna system," in *IEEE 56th Vehicular Technology Conference (VTC '02)*, vol. 4, pp. 2158–2162, Vancouver, Calif, USA, September 2002.
[20] S. G. Nash and A. Sofer, *Linear and Nonlinear Programming*, McGraw-Hill, New York, NY, USA, 1996.
[21] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1561–1571, 2001.

Seong Taek Chung received the B.S. degree in electrical engineering from Seoul National University, Korea, in 1998. He received the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Calif, in 2000 and 2004, respectively. During the summer of 2000, he was an intern with Bell Laboratories, Lucent Technologies, where he worked on multiple antenna systems. He is currently a Senior Engineer at Qualcomm Inc., San Diego, Calif. His research interest includes communication theory and signal processing.

Angel Lozano was born in Manresa, Spain, in 1968. He received the Engineer degree in telecommunications (with honors) from the Polytechnical University of Catalonia, Barcelona, Spain, in 1992 and the Master of Science and Ph.D. degrees in electrical engineering from Stanford University, Stanford, Calif, in 1994 and 1998, respectively. Between 1996 and 1998 he worked for Pacific Communication Sciences Inc. and for January 1999 he was with Bell Laboratories (Lucent Technologies) in Holmdel, NJ. Since October 1999, he has served as an Associate Editor for IEEE Transactions on Communications. Dr. Lozano holds 6 patents.

Howard C. Huang was born in Texas in 1969. He received the B.S.E.E. degree from Rice University, Houston, TX, in 1991, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 1995. He is currently a distinguished member of the technical staff in the Wireless Communications Research Department at Bell Laboratories, Lucent Technologies in Holmdel, NJ. His interests include communication theory and multiple antenna networks.



Arak Sutivong received the B.S. and M.S. degrees in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, Pa, in 1995 and 1996, respectively. He received the Ph.D. degree in electrical engineering from Stanford University, Stanford, Calif, in 2003. From 1997 to 1998, he was a Systems Engineer at Qualcomm Inc., San Diego, Calif, developing a satellite-based CDMA system, while at Stanford University, he has served as a Technical Consultant to numerous companies. He returned to Qualcomm Inc. in October 2002, where he is currently a Staff Engineer. His research interests are in information theory and its applications, wireless communications, and signal processing.

John M. Cioffi received his B.S.E.E. degree in 1978 from the University of Illinois and his Ph.D. degree in electrical engineering from the University of Stanford in 1984. He was with Bell Laboratories from 1978 to 1984 and worked at IBM Research from 1984 to 1986. In 1986, he became a Professor of electrical engineering at the University of Stanford. Cioffi founded the Amati Communications Corporation in 1991 (purchased by Texas Instruments (TI) in 1997) and was the Officer/Director from 1991 till 1997. He is currently on the board of directors of Marvell, Teknovus, Ikanos, Clariphy, and Tranetics. He is on the advisory boards of Charter Ventures, Halisos Networks, and Portview Ventures. Cioffi's specific interest is in the area of high-performance digital transmission. Dr. Cioffi was granted the Hitachi America Professorship in electrical engineering at Stanford in 2002 and was a member of the National Academy of Engineering in 2001. He received the IEEE Kobayashi Medal in 2001 and the IEEE Millennium Medal in 2000. Moreover, he was the IEEE Fellow in 1996 and received the IEE JJ Tomson Medal in 2000. He is the 1999 University of Illinois Outstanding Alumnus and received 1991 IEEE Communication Magazine Best Paper Award and 1995 ANSI T1 Outstanding Achievement Award. He was the National Science Foundation (NSF) Presidential Investigator from 1987 till 1992. Cioffi has published over 200 papers and holds over 40 patents.

EURASIP Journal on Applied Signal Processing 2004:5, 772–782

Multiple ARQ Processes for MIMO Systems

Haitao Zheng
Wireless Research Laboratory, Lucent Technologies, 791 Holmdel-Keyport Road, Holmdel, NJ 07733, USA
Email: haitaoz@lucent.com

Angel Lozano
Wireless Research Laboratory, Lucent Technologies, 791 Holmdel-Keyport Road, Holmdel, NJ 07733, USA
Email: aloz@lucent.com

Mohamed Haleem
Wireless Research Laboratory, Lucent Technologies, 791 Holmdel-Keyport Road, Holmdel, NJ 07733, USA
Email: haleem@lucent.com

Received 4 December 2002; Revised 19 August 2003

We propose a new automatic repeat request (ARQ) scheme for MIMO systems with multiple transmit and receive antennas. The substreams emitted from various transmit antennas encounter distinct propagation channels and thus have different error statistics. When per-antenna encoders are used, separating ARQ processes among the substreams results in a throughput improvement. Moreover, it facilitates the interference cancellation in certain MIMO techniques. Quantitative results from UMTS simulations demonstrate that the proposed multiple ARQ structure yields more than 30% gain in link throughput.

Keywords and phrases: MIMO systems, automatic repeat request, throughput, wireless communication, UMTS.

1. INTRODUCTION

Third-generation cellular systems are being designed to support high-speed packet data services. In the downlink, which has more stringent requirements in many of such services, high-speed packet access is provided through a shared channel where time-division multiplexing is used. Time slots are assigned to users at specific data rates through a scheduling algorithm based on the user data backlog and on channel quality indication (CQI) received via a feedback channel.[1] Such a transmission scheme allows multiple users to share the system resources efficiently by adapting to traffic and channel variations and it also avoids possible resource limitations that might occur if each user were allocated a dedicated code-multiplexed channel. Therefore, it has the potential to improve the capacity for delay-tolerant bursty services. Examples where this scheme will be implemented include the CDMA 1x EV-DO and 1x EV-DV and the UMTS high-speed downlink packet access (HSDPA) [1, 2]. Several advanced technologies are employed in high-speed downlink transmission to improve link throughput or reduce packet delay by adapting to the time-varying channel conditions, traffic statistics, and quality-of-service requirements. Some of these adaptive techniques, relevant to this paper, are summarized below.

Multiple transmit and receive antennas. The use of multiple antennas at each base station sector is already part of every third-generation standard. In the downlink, specifically, these antennas can be used to provide transmit diversity and/or to direct a beam towards the intended terminal. The deployment of multiple receive antennas at data terminals is also being considered. The combination of multiple transmit and receive antennas will enable the implementation of a number of multiple-input multiple-output (MIMO) techniques that promise spectacular increases in throughput without the need for additional power or bandwidth [3, 4, 5].

Dynamic link adaptation through adaptive modulation and coding. Typically, each transmission in the downlink shared channel is at the maximum available power, with no power control. Therefore, link adaptation [6, 7], which adjusts the modulation and coding schemes (MCS), provides an efficient way of maximizing the instantaneous usage of the wireless channel. Specifically, it enables the use of aggressive MCSs when channel conditions are favorable while it reverts to MCSs that are more robust but with lower transmission rates when channel conditions degrade. The base station

[1] Each terminal measures its channel condition and translates it into a metric to be fed back to the serving base station.

selects the appropriate MCS based on the CQI for the user served at each time slot. We hereby refer to the MCS selection process as the mapping design.

Automatic repeat request (ARQ) or hybrid ARQ (HARQ). The performance of MCS-based link adaptation largely depends on the accuracy of the CQI, which is difficult to maintain as velocity increases. The delay tolerance of many data services enables the use of retransmission schemes to recover erroneous packets. Recently, HARQ techniques have been adopted by several wireless standardization bodies, for example, 3GPP and 3GPP2. HARQ [8, 9, 10] can improve throughput performance, compensate for link adaptation errors, and provide a finer granularity in the rates effectively pushed through the channel. Upon detecting a transmission failure, mostly by cyclic redundancy check (CRC), the terminal sends a request to the base station for retransmission. The delay due to packet acknowledgement can be significantly reduced by placing the HARQ functionality in the base station (Node B in UMTS) rather than in the radio network controller (RNC in UMTS). The packet decoder at the mobile combines the soft information of the original transmission with those of the subsequent retransmissions. The combined signal has higher probability of successful decoding. In general, there are two ways of soft combining. With *chase combining*, the base station repeatedly sends the same packet and the receiver aggregates the energy from the (re)transmissions to improve the signal-to-noise ratio (SNR) [11, 12]. A more sophisticated HARQ mechanism, named *incremental redundancy* (IR), transmits additional redundant information in each retransmission and gradually reduces the coding rate until successful decoding occurs [13, 14, 15]. Compared with chase combining, IR requires larger receiver buffers but it can achieve better performance [16]. It also provides finer granularity in the encoded rates and allows for better adaptation to channel variations.

Scheduler. In a multiuser system where user channel conditions change over time, a scheduler can exploit those channel variations by giving certain priority to the users with transitorily better channels. The scheduler critically impacts the system performance. Several scheduling algorithms have been proposed in the literature to maximize the packet data throughput, subject to various fairness conditions [17].

The above technologies are tightly coupled. However, since some of them reside in different layers, that is, HARQ in the medium access control (MAC) layer and MIMO in the physical layer, they are usually discussed and treated separately. The evaluation of each technology fails to take into account the performance improvement or degradation brought about by the other one. In particular, the link layer performance of any MIMO algorithm is usually selected according to the raw data rate at some operating point, for example, 10% packet error rate. However, when some level of channel uncertainty exists and the system supports HARQ, it may be beneficial to transmit aggressively at higher packet error rates and recover channel errors through retransmissions [18]. The throughput depends heavily on the transmission strategy. An overly aggressive transmission could produce too many unsuccessful packet transmissions that diminish the overall throughput, while an overly conservative one fails to fully utilize the channel. In this case, the overall throughput depends on the algorithms at both layers and only cross-layer design can enable the most efficient use of the channel.

In this paper, we address some of the key design issues associated with the choice of the HARQ structure to be used for MIMO physical layer transmission. We propose a new HARQ structure that matches the layered structure of the most popular MIMO architectures [19]. Simulation results show that the performance sensitivity to the choice of HARQ depends on the aggressiveness of the transmissions and on the type of CQI.

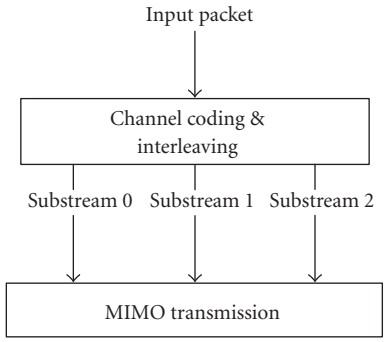
The paper is organized as follows. In Section 2, we describe the layered architectures with per-antenna encoding. Modifications to the conventional HARQ structure to fit these layered architectures are discussed in Section 3. We compare the performance of different HARQ structures in Section 4. Conclusions are drawn in Section 5.

2. LAYERED ARCHITECTURES WITH PER-ANTENNA ENCODING

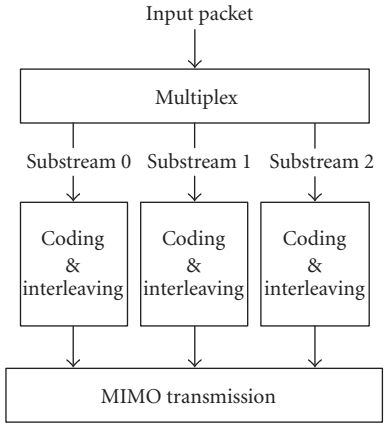
In order to approach the MIMO channel capacity in rich multipath environments, the substreams radiated from the various transmit antennas should be uncorrelated [20, 21]. Nonetheless, it may in practice be advantageous to jointly encode them (Figure 1a). This has motivated a blossoming interest in the design of space-time (vector) codes [22]. Clearly, when the substreams are jointly encoded, they should share a single CRC.

The complexity of joint detection, however, explodes as the number of transmit antennas grows large. As a result, there has also been strong interest in devising alternative approaches. One such approach is that of layered architectures, which incorporate multiple scalar encoders, one per transmit antenna. In these architectures, input data is demultiplexed into multiple substreams, which are then separately encoded and radiated from the various transmit antennas (Figure 1b). At the receiver, the substreams are successively detected and cancelled [4, 5]. Specifically, the information extracted from each substream is reencoded, interleaved, and modulated to construct a replica of the transmitted substream. This replica, properly combined with the channel response, is then subtracted from the overall received signal so that—if there are no errors—the interference contribution of this substream is removed. The complexity of these architectures increases more gracefully with the number of antennas. Furthermore, they can capitalize on existing scalar coding formats.

A layered architecture can approach the MIMO channel capacity if the data rates of the different transmit antennas are appropriately adjusted [23, 24]. This adjustment requires separate CQI, one per transmit antenna, and thus the amount of feedback required increases linearly with the number of transmit antennas. We hereby refer to it as per-antenna rate and CQI. Alternatively, a common CQI—and thus the same data rate—can be used for all transmit



(a) MIMO with joint coding.



(b) MIMO with per-antenna coding.

Figure 1: MIMO transmitter architecture with different coding structures.

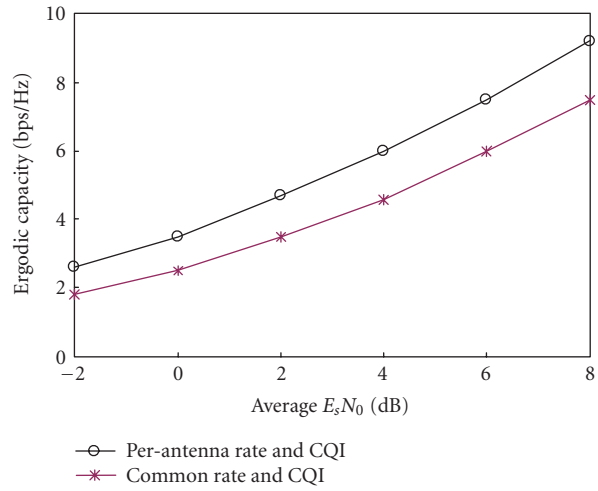


Figure 2: Ergodic Shannon capacity with 4 transmit and 4 receive antennas obtained via Monte Carlo simulation on a Rayleigh-faded channel with no antenna correlation.

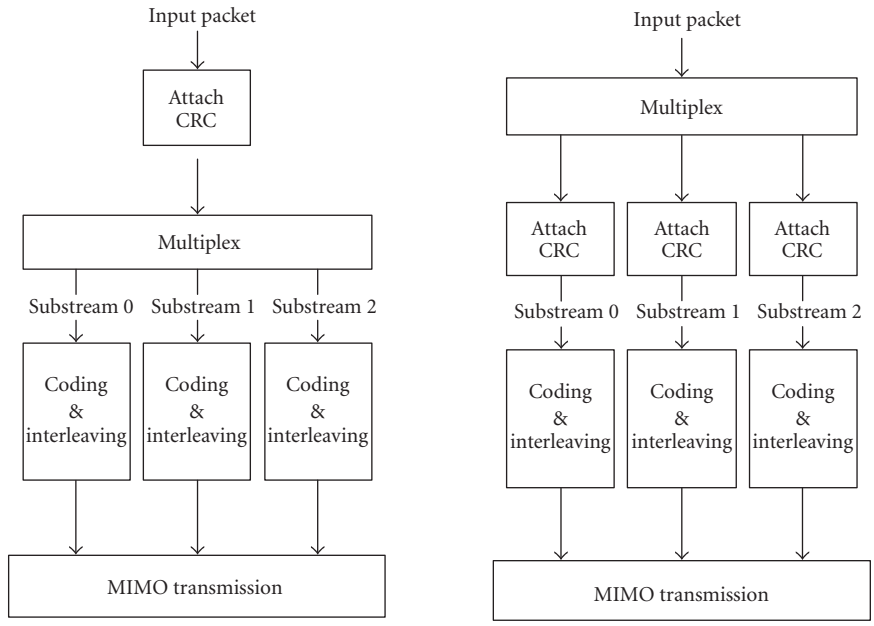
antennas at the expense of some loss in capacity [23]. To illustrate this point, Figure 2 depicts the difference between the capacity with and without the constraint that the data rate at each of the transmit antennas be equal, for the specific case of 4 transmit and 4 receive uncorrelated antennas with Rayleigh fading. For the purpose of this paper, in any event, the most relevant feature of a layered architecture is that it does not constraint the transmit antennas to be jointly encoded and share a unique CRC.

3. HARQ MECHANISMS FOR MIMO SYSTEMS

If the MAC layer is unaware of the presence of MIMO at the physical layer, HARQ simply attaches a single CRC to the packet with such CRC encompassing the data radiated from the various transmit antennas. We refer to this scheme, depicted in Figure 3a, as MIMO single ARQ (MSARQ). Since substreams transmitted from different antennas encounter distinct propagation channels, they have different error statistics. Using a typical channel propagation model with 4 transmit and 4 receive uncorrelated antennas [21], we observe that in more than 70% of error events,[2] only the substreams from 1 or 2 transmit antennas are corrupted and thus require a retransmission (Figure 4). However, upon an error event, an MSARQ receiver has to request a retransmission of the entire packet because it relies on the single CRC over the whole packet. Retransmitting substreams that have already been correctly received wastes throughput. When multiple per-antenna encoders are used, it becomes possible to remove the constraint that the substreams radiated from multiple transmit antennas share a single ARQ process.

For per-antenna MIMO encoding architectures, we herein propose to employ multiple ARQ processes, 1 for each substream radiated from 1 transmit antenna or group of antennas. This scheme is independent of the receiver-processing algorithm and only requires that the receiver decodes substreams independently. We refer to this scheme as MIMO multiple ARQ (MMARQ). As shown in Figure 3b, a CRC symbol is appended to each substream. At the receiver, each such substream is decoded and the associated CRC is used to validate the content. Multiple acknowledgment (NACK/ACK) indications are then sent back to the transmitter. After receiving these acknowledgements, the transmitter sends fresh packets from the transmit antennas that have been successfully acknowledged and retransmits the substreams that have been negatively acknowledged through their associated transmit antennas. Hence, the HARQ operations at different transmit antennas are independent of each other. We focus on high-speed downlink data transmission so that the overhead due to multiple CRC symbols is negligible. However, we need to consider the uplink signaling overhead due to multiple acknowledgements. For each ARQ process, NACK/ACK requires an overhead of 1 bit plus error protection redundancy. Therefore, the amount of ARQ feed-

[2] An error event occurs when any of the substreams contains an error.



(a) MSARQ transmitter.

(b) MMARQ transmitter.

Figure 3: Transmitter structures of MSARQ and MMARQ.

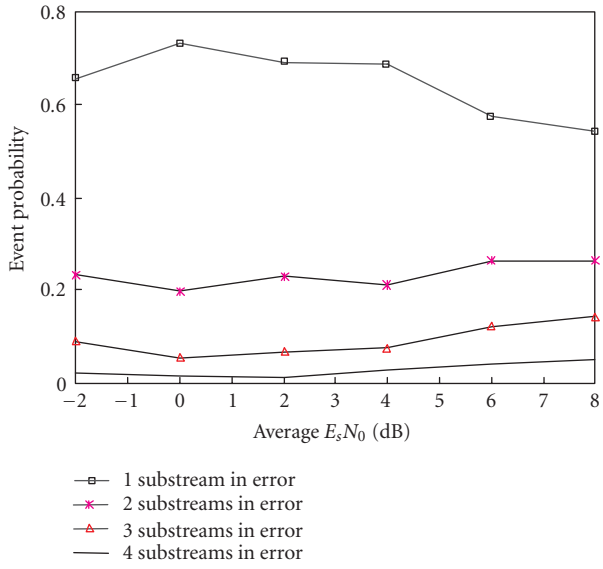


Figure 4: Probability distribution of the number of corrupted substreams in an error event with 4 transmit and 4 receive uncorrelated antennas and frequency-flat fading.

back overhead scales with the number of transmit antennas. When that number is large, grouping the transmit antennas and assigning a single ARQ process to each group can reduce the signaling overhead.

Next, using per-antenna encoders with successive decoding and cancellation at the receiver as an example, we describe the receiving procedures for both MMARQ and MSARQ. The receiver decodes the transmitted substreams sequentially following a certain order, which can be optimized to achieve the best throughput performance. The first substream is decoded from the overall aggregate received signal $\mathbf{Y}(t)$. The information data $S_0(t)$, extracted from substream 0, is then reencoded, interleaved, and modulated to construct a replica of the transmitted substream. This replica, combined with the channel response, that is, $F(S_0(t), \mathbf{H}(t))$, is then subtracted from $\mathbf{Y}(t)$ so that the interference contribution of this substream to the others is removed. This procedure is the so-called interference cancellation. The same process is then applied to the remaining substreams, which are thus successively extracted.

For MMARQ, the interference cancellation and HARQ packet combining procedures can be blended advantageously. In that case, the receiver would decode a substream and use its associated CRC to validate the content. If this substream carries a retransmission packet and contains uncorrectable errors, the soft symbols of the packet would be combined with those of the previous transmission(s) to extract the information data. The receiver would then perform interference cancellation to remove the interference due to this substream. Interference cancellation is performed regardless of the results of the CRC validation; therefore, all the subsequent substreams can be decoded without waiting for the retransmission of the current substream. However, the reliability of the decoded data is much higher after HARQ packet combining and, thus, using such data to reconstruct the signal replicas for interference cancellation reduces error propagation. The detailed receiver procedure is shown in Figure 5.

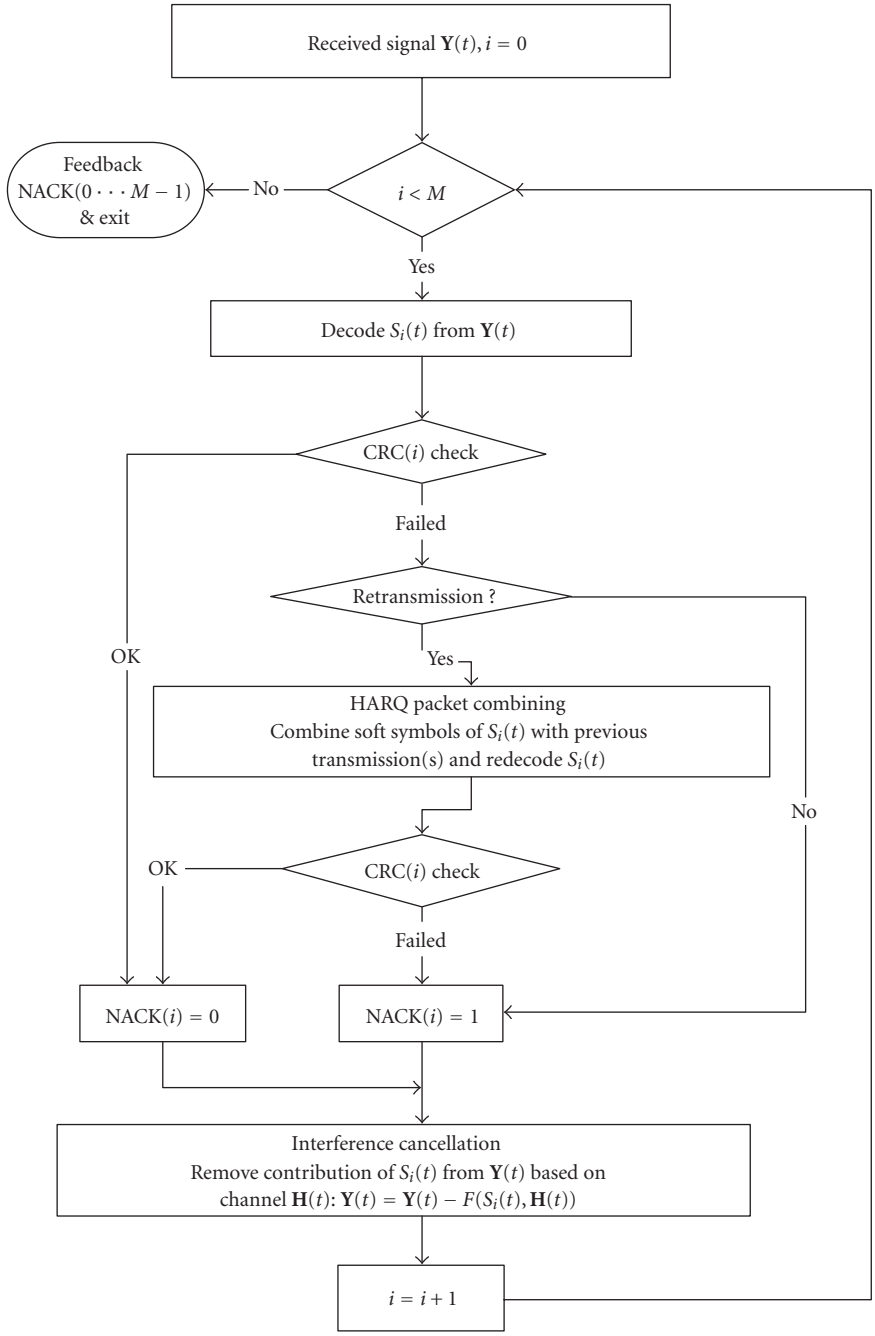


Figure 5: MMARQ receiver flow chart.

In contrast, it is not so easy to combine HARQ with interference cancellation when MSARQ is employed. As illustrated in Figure 6, MSARQ separates HARQ packet combining from interference cancellation. The receiver performs packet decoding and interference cancellation to extract the substreams and then combines those substreams into a compound packet. In this case, decoding errors at each substream could propagate to the substreams that are decoded afterwards. Such error propagation could severely degrade the performance. Another alternative would be to recancel interference on the HARQ combined signal upon a CRC failure. This procedure is shown in Figure 7, wherein interference cancellation is conducted twice. We refer to it as MSARQ IC. The resulting hardware design, however, could be problematic, as the receiver would need to quickly feedback the NACK indicator to the transmitter.

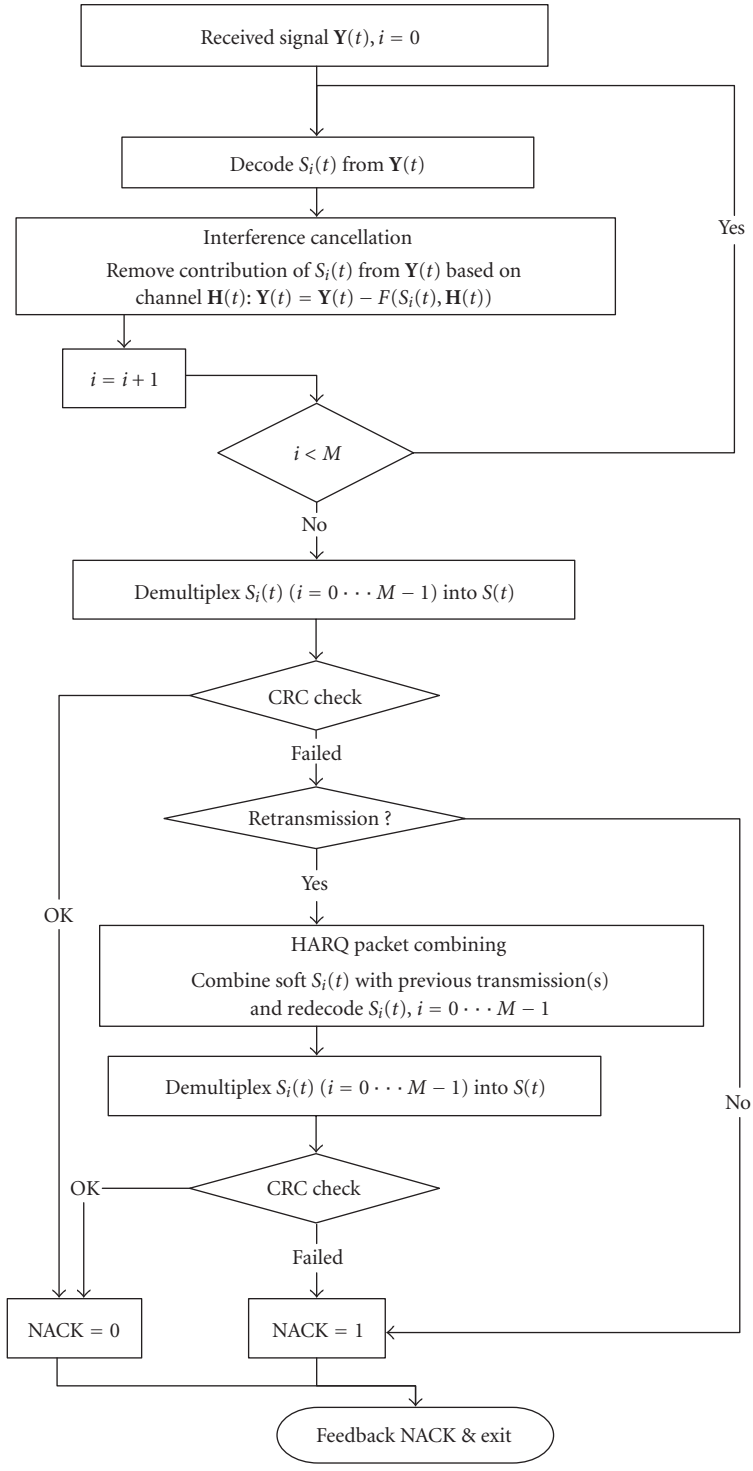


Figure 6: MSARQ receiver flow chart type I.

4. COMPARISON OF MSARQ AND MMARQ

In this section, we compare the performance of MSARQ and MMARQ in the context of UMTS HSDPA [25]. The most prominent features of HSDPA, which is specifically geared towards delay-tolerant data, are as follows.

(1) A fraction of the power and code space available at the base station is allocated to HSDPA while the rest is assigned to pilots, overhead channels, and voice traffic.
(2) HSDPA users are time-multiplexed in short frames. A scheduler at the MAC layer determines the user to be served at each frame. Each scheduling interval or

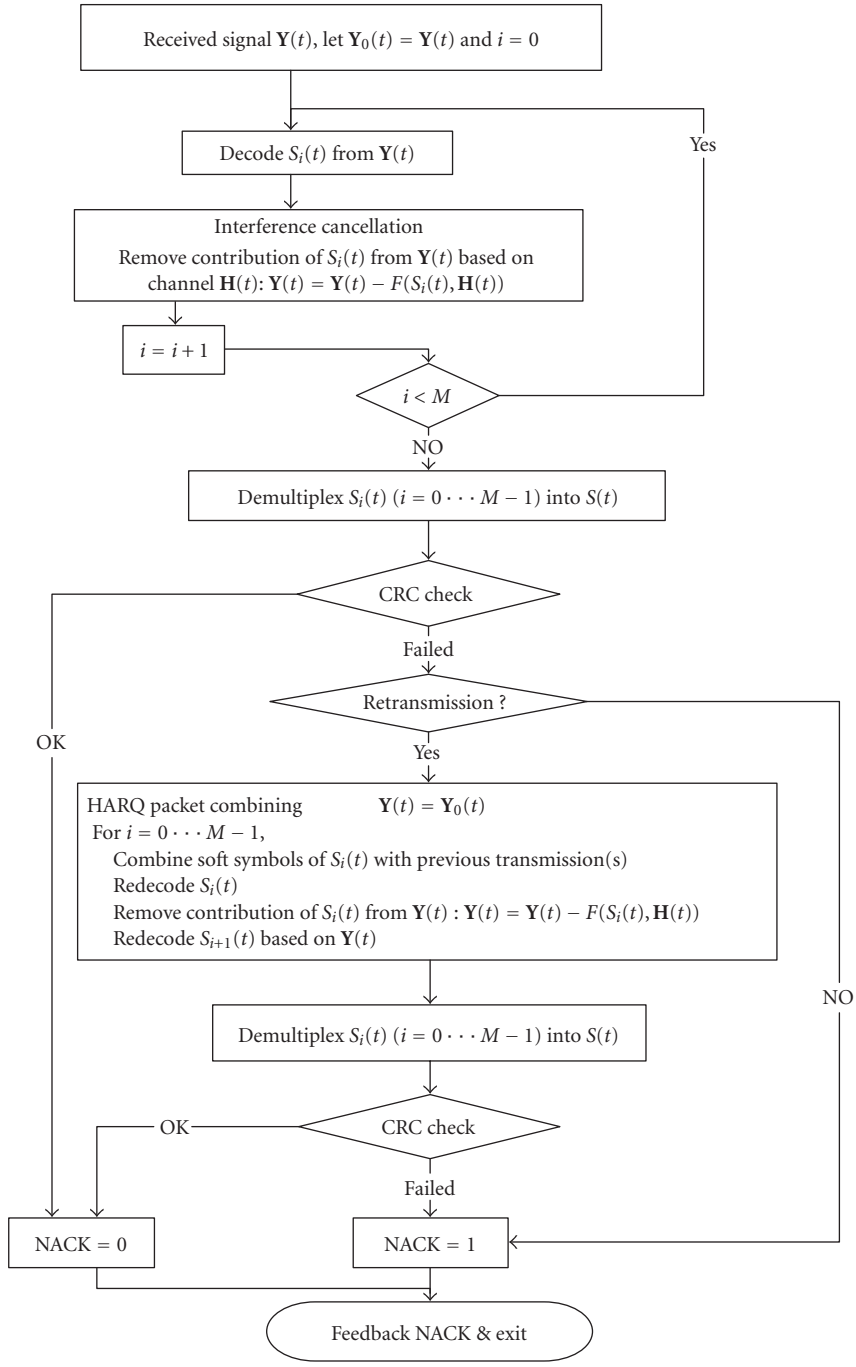


Figure 7: MSARQ IC receiver flow chart.

frame lasts 2 milliseconds. We assume that the entire HSDPA code space (10 codes in this paper) and transmit power are assigned to the scheduled user. That is, the base station transmits to only one user in each frame using 10 codes and full power. The transmit signal consists of a superposition of such 10 orthogonal codes.

(3) The Node B (or base station) MAC determines the transmission rate for the user being served, based on the CQI.

(4) The HARQ functionality resides between the Node B and the mobile terminal to permit soft combining and fast NACK/ACK feedback.

We have developed a simulation tool that captures the dynamic processes in a radio network. The simulated radio network consists of a base station (Node B) and multiple user

terminals. The Node B possesses the following functionalities.

(a) $\text{MAC}_{\text{HSDPA}}$. It performs scheduling, MCS selection, and HARQ, based on the CQI feedback and the NACK/ACK signaling from each terminal.

(i) *Scheduler*. System performance depends heavily on the scheduling algorithm. For the purpose of this work, we limit ourselves to a round-robin scheduler, which exhibits maximum fairness across users. Additionally, with such scheduler, it is easy to quantify the system-level performance from the single-user performance.

(ii) *MCS selection*. The MCS at each transmit antenna is separately controlled through CQI feedback from the receiver [23, 24].

(iii) *HARQ*. The downlink HARQ operates asynchronously, that is, the retransmissions can take place anytime after the Node B receives a NACK/ACK. The scheduler determines the exact time. To compensate for the NACK/ACK feedback delay of 2 frames, each HARQ entity operates in terms of three stop-and-wait (SAW) processes. This allows HARQ to operate continuously without waiting for a NACK/ACK signal. For MSARQ, all transmit antennas use a single HARQ entity with 3 processes while, for MMARQ, each transmit antenna uses one HARQ entity with 3 processes. Chase combining is used to combine the initial transmission with the retransmissions. The maximum number of retransmissions is 30. If a corrupted packet cannot be recovered after exhausting the maximum number of retransmissions, the packet is discarded and the associated loss should be recovered by higher layer error control mechanism.

(b) *PHY*. The physical layer simulation consists of a sequence of events such as transmission and reception of signals, signal-to-interference-and-noise ratio (SINR) evaluation, and channel estimation. It employs a bandwidth of 5 MHz with 3.33-milliseconds frames. We assume that the uplink channel operates at a rate of 64 kbps. At the terminal, the substreams radiated by the various transmit antennas are decoded according to a fixed order. The MCS of each such substream is selected based on its detected SINR at the receiver and it is then fed back as a CQI message. Some additional premises are summarized below:

(i) fading is Rayleigh-distributed and frequency-flat and the channel is either perfectly known at the receiver or modeled by adding simulated estimation noise onto the actual channel;

(ii) pedestrian speed (3 Km/hr);

(iii) 70% of transmit power dedicated to HSDPA;

(iv) 10 out of 16 orthogonal codes dedicated to HSDPA;

(v) 4 uncorrelated transmit and 4 uncorrelated receive antennas;

(vi) 7 MCSs employing turbo codes with varying rates and symbol repetition [4]: QPSK rate 1/4 repeated 4 times, QPSK rate 1/4 repeated 2 times, QPSK rate 1/4, QPSK rate 1/2, QPSK rate 3/4, 16-QAM rate 1/2, and 16-QAM rate 3/4.

The probability of each substream being detected erroneously is given by a frame error rate (FER) versus instantaneous SINR curve for each MCS. For the above MCS schemes, these curves are displayed in Figure 8.

The ultimate performance measure is the single-user throughput, defined as the ratio between the number of information bits correctly received by a user and the time that the channel is allocated to that user:

$$\text{throughput} = \frac{\text{total good bits}}{(\text{total frames with transmissions}) \cdot \text{frame duration}}. \quad (1)$$

Notice that the throughput represents the peak net throughput that can be delivered to a user.

It should be pointed out that the throughput depends on the mapping between the detected SINR and the selected MCS per antenna. Such mapping is adjusted in order to maximize the throughput while maintaining some target FER measured prior to HARQ operation. When this target FER is small (less than 5%), the probability of retransmission is low and there is no large gain with any kind of ARQ. As the target FER increases, the probability of retransmissions grows and there is a considerable gain with MMARQ. Hence, we optimize the FER to maximize the throughput.

4.1. Performance with perfect channel estimation and feedback

Our initial simulations assume perfect channel estimation and error-free uplink feedback. We first examine the advantage of combining HARQ with interference cancellation by comparing the compound packet error performance of MSARQ and MSARQ IC. Separating interference cancellation from HARQ combining fails to eliminate the interference from any corrupted substream even if the substream is later fully recovered through HARQ packet combining. Such inefficiency results in a higher compound packet error rate (Figure 9). To quantify the advantage of per-antenna HARQ in MMARQ, the throughput performances of MMARQ, MSARQ, and MSARQ IC are compared in Figure 10. We observe that MMARQ achieves 10%–20% improvement over MSARQ IC and 26%–40% over MSARQ. Thus, the contributions of combined operation and multiple ARQ structures are roughly equal. The ergodic Shannon capacities for open-loop single-transmit single-receive and 4-transmit 4-receive configurations are also shown in the same figure as references.

In the above example, MSARQ, MSARQ IC, and MMARQ use the same MCS/SINR settings, which maximize the throughput for MMARQ but not necessarily for MSARQ and MSARQ IC. Through additional simulations, we find that the optimal MCS/SINR settings for MSARQ and MSARQ IC yield a compound FER of 8%–10%, while the optimal FER for MMARQ is around 15%–20%. The

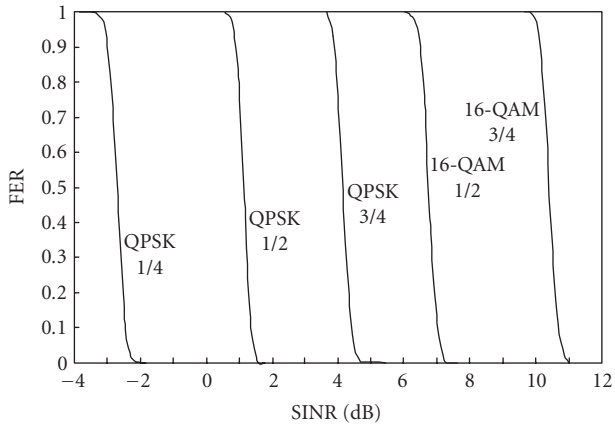


Figure 8: Frame error rate (FER) versus SINR for a single transmit and a single receive antenna.

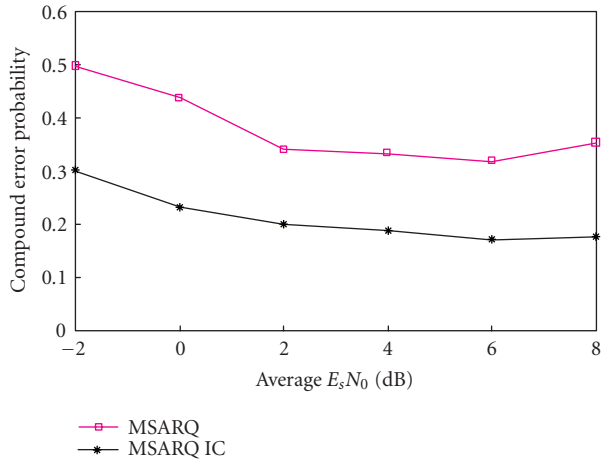


Figure 9: Compound packet error rate of MSARQ and MSARQ IC.

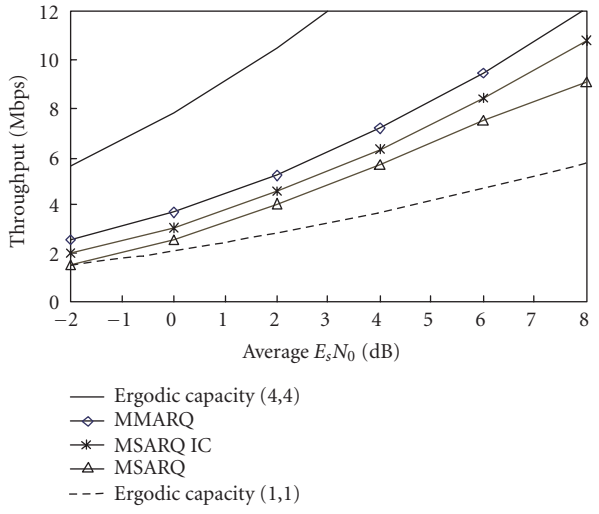


Figure 10: Throughputs of MMARQ, MSARQ, and MSARQ IC with interference cancellation in ideal conditions.

corresponding individual substream error rates are 2%–5% and 8%–18%, respectively. In practice, it is quite difficult to guarantee a substream error rate of 5% or less. Therefore, the optimal throughput of MMARQ would be easier to achieve in a realistic environment. Nevertheless, the optimized throughputs are shown in Figure 11, where the improvement of MMARQ drops to around 10% with respect to MSARQ IC and 20% with respect to MSARQ. By operating at a low packet error rate, channel coding and packet combining can eliminate most channel errors. As such, the throughput gap between MSARQ and MSARQ IC also diminishes.

4.2. Performance with imperfect channel estimation and feedback

Next, we examine the performance of MSARQ, MSARQ IC, and MMARQ in more realistic conditions, with imperfect channel estimation and imperfect uplink feedback. The main sources of imperfection are limited pilot power, finite channel coherence time, and feedback delay. We model these nonidealities by adding noise to the SINR, that is,

$$\hat{\gamma}_{\mathrm{dB}} = \gamma_{\mathrm{dB}} + N(0, \sigma_a^2), \tag{2}$$

where $\hat{\gamma}_{\mathrm{dB}}$ represents the SINR in dB as estimated by the receiver, γ_{dB} represents the actual SINR in dB, and $N(0, \sigma_a^2)$ represents Gaussian noise with variance σ_a^2. The estimation error not only impacts the MIMO signal detection and decoding process, but also impacts the MCS selected for each transmit antenna. In addition, the uplink feedback channel also encounters a uniformly distributed binary error rate of 6%, which could corrupt the CQI and the NACK/ACK indication(s). Figure 12 illustrates the throughput performance of MMARQ, MSARQ, and MSARQ IC for $\sigma_a^2 = 1.5$ dB. The performance degradations range from 10% to 18% for MMARQ, 17% to 32% for MSARQ, and 16% to 24% for MSARQ IC. Relatively, MMARQ is less sensitive to channel estimation noise and feedback errors. As the level of uncertainty increases, it becomes more difficult to guarantee a successful transmission without sacrificing packet throughput. In this case, it is beneficial to transmit aggressively and use HARQ to recover from channel errors. Overall, MMARQ achieves 30%–45% throughput improvement over MSARQ, while per-antenna ARQ contributes to a 15%–25% throughput improvement.

5. CONCLUSION

We have proposed a new ARQ scheme suitable for any MIMO scheme in which substreams radiated from different antennas are encoded separately. Conventionally, a single ARQ process is applied to each data packet. Upon an error event, all constituent substreams—including those that have already been correctly received—are retransmitted. In contrast, our proposed scheme separates the ARQ processes for the substreams. We have quantified the gains of the new scheme within the context of UMTS high-speed downlink data access. We first considered ideal conditions with perfect

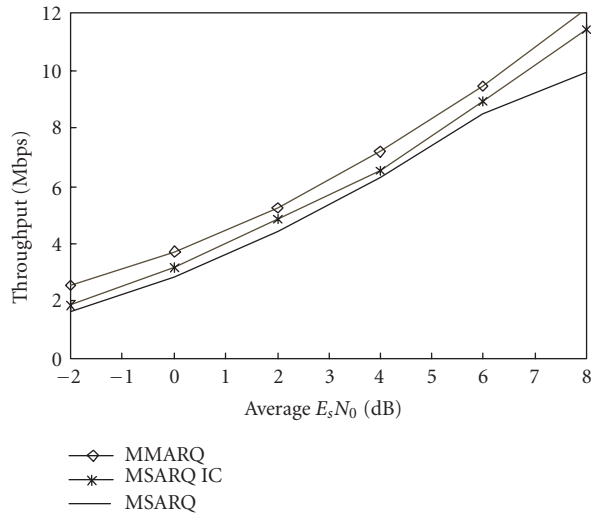


Figure 11: Throughputs of MMARQ, MSARQ, and MSARQ IC with interference cancellation in ideal conditions using the optimized MCS/SINR settings.

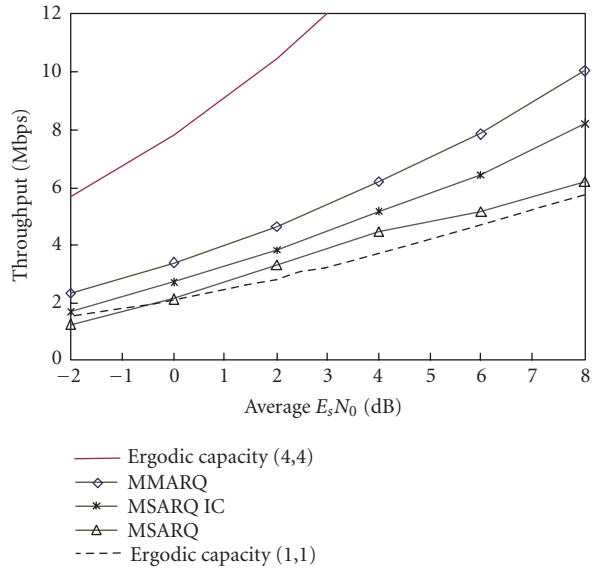


Figure 12: Throughput of MMARQ, MSARQ, and MSARQ IC with interference cancellation in realistic conditions (with imperfect channel estimation and imperfect uplink feedback).

channel estimation and error-free uplink feedback, where MMARQ improves the throughput by 25%–40%. We then performed the simulations in more realistic conditions, with imperfect channel estimation and possibly erroneous uplink feedback. Such uncertainty leads to a higher loss rate, and HARQ becomes a major technique for efficient error control and recovery. Hence, MMARQ is even more favorable with the performance improvement increasing to 30%–45% compared with MSARQ. It should be pointed out that the results presented here are based on the premise of frequency-flat fading and uncorrelated antennas. Frequency-selective fading may modify this conclusion, and this problem is currently under investigation.

Traditionally, the physical layer had been considered the performance bottleneck in wireless systems due to the unpredictable nature of the radio channel. Higher layer issues, such as scheduling, link adaptation, retransmissions, and mobile routing, used to be discussed and treated separately from major physical layer issues. With the convergence of mobile communications and data services, however, there is a growing need for a cross-layer design that facilitates the interaction of multiple protocol layers. In particular, one can couple the design of link layer (i.e., MAC and RLP) with that of the physical layer. The superior performance of MMARQ confirms the benefits of such joint layer design.

REFERENCES

[1] Third Generation Partnership Project, "Physical layer aspects of UTRA high speed downlink packet access (Release 2000)," 3GPP Tech. Rep. 25.848, Technical Specification Group Radio Access Network, 2001.

[2] 3GPP2 C.S00024, "CDMA2000 high rate packet data air-interface specification," 2000.

[3] A. Lozano, F. R. Farrokhi, and R. A. Valenzuela, "Lifting the limits on high speed wireless data access using antenna arrays," *IEEE Communications Magazine*, vol. 39, no. 9, pp. 156–162, 2001.

[4] G. J. Foschini, "Layered space-time architecture for wireless communication in fading environments when using multiple antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[5] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 11, pp. 1841–1852, 1999.

[6] K. Balachandran, S. R. Kadaba, and S. Nanda, "Channel quality estimation and rate adaptation for cellular mobile radio," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 7, pp. 1244–1256, 1999.

[7] B. Vucetic, "An adaptive coding scheme for time-varying channels," *IEEE Trans. Communications*, vol. 39, no. 5, pp. 653–663, 1991.

[8] S. Lin, D. Costello, and M. Miller, "Automatic-repeat-request error-control schemes," *IEEE Communications Magazine*, vol. 22, no. 12, pp. 5–17, 1984.

[9] D. J. Costello Jr., J. Hagenauer, H. Imai, and S. B. Wicker, "Applications of error-control coding," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2531–2560, 1998.

[10] Third Generation Partnership Project, "Hybrid ARQ methods for FDD in Release 2000," TSG-R WG1 document, TSGRl#13(00)0637, May 2000.

[11] G. Benelli, "An ARQ scheme with memory and soft error detectors," *IEEE Trans. Communications*, vol. 33, no. 3, pp. 285–288, 1985.

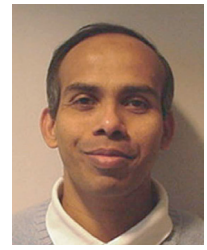
[12] D. Chase, "Code combining–a maximum-likelihood decoding approach for combining an arbitrary number of noisy packets," *IEEE Trans. Communications*, vol. 33, no. 5, pp. 385–393, 1985.

[13] D. N. Rowitch and L. B. Milstein, "On the performance of hybrid FEC/ARQ systems using rate compatible punctured turbo (RCPT) codes," *IEEE Trans. Communications*, vol. 48, no. 6, pp. 948–959, 2000.

[14] R. Love, B. Classon, A. Ghosh, and M. Cudak, "Incremental redundancy for evolutions of 3G CDMA systems," in *Proc. IEEE 55th Vehicular Technology Conference, 2002. VTC Spring 2002*, vol. 1, pp. 454–458, Birmingham, Ala, USA, May 2002.
[15] A. Das, F. Khan, A. Sampath, and H. J. Su, "Performance of hybrid ARQ for high speed downlink packet access in UMTS," in *Proc. IEEE VTS 54th Vehicular Technology Conference, 2001. VTC 2001 Fall*, vol. 4, pp. 2133–2137, Atlantic City, NJ, USA, 2001.
[16] P. Frenger, S. Parkvall, and E. Dahlman, "Performance comparison of HARQ with chase combining and incremental redundancy for HSDPA," in *Proc. IEEE VTS 54th Vehicular Technology Conference, 2001. VTC 2001 Fall*, vol. 3, pp. 1829–1833, Atlantic City, NJ, USA, 2001.
[17] M. Andrews, K. Kumaran, A. Ramanan, K. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, vol. 39, no. 2, pp. 150–154, 2001.
[18] H. Zheng and H. Viswanathan, "Optimizing the ARQ performance in downlink packet data systems with scheduling," in *Proc. IEEE VTS Vehicular Technology Conference, 2003. VTC 2003 Fall*, Orlando, Fla, USA, October 2003.
[19] H. Zheng, A. Lozano, and M. Haleem, "Multiple ARQ processes for MIMO systems," in *Proc. of IEEE 13th Symposium on Personal, Indoor and Mobile Radio Communication (PIMRC '02)*, Lisbon, Portugal, September 2002.
[20] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.
[21] D. Chizhik, F. R. Farrokhi, J. Ling, and A. Lozano, "Effect of antenna separation on the capacity of BLAST in correlated channels," *IEEE Communications Letters*, vol. 4, no. 11, pp. 337–339, 2000.
[22] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.
[23] A. Lozano, "Capacity-approaching rate function for layered multiantenna architectures," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 616–620, 2003.
[24] S. T. Chung, A. Lozano, and H. C. Huang, "Approaching eigenmode BLAST channel capacity using V-BLAST with rate and power feedback," in *Proc. IEEE VTS 54th Vehicular Technology Conference, 2001. VTC 2001 Fall*, vol. 2, pp. 915–919, Atlantic City, NJ, USA, October 2001.
[25] Third Generation Partnership Project, "UTRA high-speed downlink packet access (Release 4)," 3GPP Tech. Rep. 25.950, Technical Specification Group Radio Access Network, March 2001.

Haitao Zheng received her B.S. degree with highest honor in electrical engineering from Xian Jiaotong University, China, in 1995, the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, MD, in 1998 and 1999, respectively. From 1995 to 1998, she was an Institute for System Research Fellow at University of Maryland. She received the 1998-1999 George Harhalakis Outstanding Systems Engineering Graduate Student Award in recognition of outstanding contributions in cross-disciplinary research from the University of Maryland. Since August 1999, she is with Wireless Research Laboratory, Bell Labs, Lucent Technologies, Holmdel, NJ. Her research interests include wireless communications and networking, multimedia communications, and signal processing. Recently, she received the Bell Laboratories 2002 Presidents Gold Award in recognition of outstanding level of innovation, technical excellence, and business impact. She currently serves as the TPC Member of IEEE Multimedia Signal Processing Technical Committee, TPC of ICME 2003, TPC of Globecom 2003, Guest Editor of the IEEE JSAC Special Issue on Advanced Mobility Management and QoS Protocols for Wireless Internet, and Guest Editor of the EURASIP Journal on Applied Signal Processing Special Issue on Cross Layer Design for Communications and Signal Processing. She has served as the TPC Member of ICME 2002, ICC 2003, ICASSP 2002, and so forth.

Angel Lozano was born in Manresa, Spain, in 1968. He received his Engineer degree in telecommunications (with honors) from the Polytechnical University of Catalonia, Barcelona, Spain, in 1992 and his M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, Calif, in 1994 and 1998, respectively. Between 1996 and 1998, he worked for Pacific Communication Sciences Inc. and for Conexant Systems in San Diego, Calif. Since January 1999, he has been with Bell Laboratories, Lucent Technologies, Holmdel, NJ. Since October 1999, Dr. Lozano has served as Associate Editor for IEEE Transactions on Communications. He holds 6 patents.



Mohamed Haleem received his B.S. Eng. degree from the Department of Electrical and Electronic Engineering, University of Peradeniya, Sri Lanka in 1990, the M.Phil. degree from the Department of Electrical & Electronic Engineering, Hong Kong University of Science & Technology, in 1995, and is currently working toward his Ph.D. degree in electrical engineering at Stevens Institute of Technology, Hoboken, NJ. He was with the academic staff of the Department of Electrical & Electronic Engineering, University of Peradeniya, from 1990 to 1993, and has been with the Wireless Communications Research Department, Bell Laboratories, Lucent Technologies at Crawford Hill, Holmdel, NJ, from 1996 to 2002. His research interests include dynamic resource assignment, channel adaptive transmission techniques for wireless communication systems, multiple antenna communication systems, and application of stochastic dynamic programming and optimization techniques to wireless communication systems.