Multiuser Detection for Interference-Limited MIMO Systems

Huaiyu Dai, Student Member, IEEE and Andreas F. Molisch, Senior Member, IEEE
AT&T Labs - Research
200 Laurel Avenue
Middletown, NJ 07748
Email: huaiyud@ee.princeton.edu, Andreas.Molisch@ieee.org

Abstract—In this paper, we consider the application of multiuser detection to the downlink of interference-limited multiple-input multiple-output (MIMO) cellular systems. MIMO systems have been shown to yield a tremendous capacity for a single link with white Gaussian noise. In a cellular environment, there will often be co-channel interference from other cells, which becomes the dominating channel impairment. Here space-time layered architectures and turbo processing techniques are combined with multiuser detection techniques for combating intercell interference. Among various multiuser detection methods examined, linear MMSE and successive interference cancellation have been shown to be feasible and effective. Based on these two multiuser detection schemes, one of which may outperform the other for different settings, an adaptive detection scheme is developed.

I. INTRODUCTION

In noises-limited environments, multiple-input multiple-output (MIMO) systems offer much higher capacities than either conventional systems or “smart antenna” systems that have multiple antennas only at one link end [7], [15]. The reason is that each of the $N_t$ transmit antennas can send out a different data stream, while the $N_r$ receive antennas are used to receive and separate those streams. Thus, the capacity goes linearly with the number of antennas. Smart antennas, on the other hand, improve the signal-to-noise ratio, so that the capacity goes only logarithmically with the number of antennas.

The situation is very different in interference-limited environments. Smart antennas can suppress co-channel interference (CCI) from neighboring cells. Essentially, each additional antenna can suppress one interferer. This allows aggressive frequency reuse and thus high cellular capacity, even though the link capacity may only be moderate. MIMO systems, on the other hand, increase the number of interferers (linearly with the number of transmit antennas), so that the receiver cannot suppress them. Recent investigations at AT&T Labs [2] considered the performance of a “standard” MIMO system in interference-limited environments, and showed that smart antenna systems offer almost the same capacity as MIMO systems.

Multiuser detection (MUD) is an increasingly popular technique for the suppression of CCI [16]. In this paper, we investigate whether this technique can be used to significantly decrease the interference in MIMO systems, thus allowing higher capacities. This is especially important for the downlink case, which usually needs higher data rates, especially for web-browsing and multimedia applications.

The paper is organized the following way. In Section II, we outline the system model. Section III discusses various state-of-the-art techniques, i.e., the space-time layered architecture, the turbo processing and multiuser detection, which form the bases for our detection methods. Section IV gives simulation results of various detection schemes. Section V presents the conclusions.

II. SYSTEM MODEL

We consider a TDMA/FDMA multi-cell system, where each base station (BS) and mobile station (MS) has the same number of antennas $N$. We take into account interference from the first tier of the center-excited cell configuration with reuse factor of one. We assume for each cell a frequency-flat, quasi-static Rayleigh fading environment with sufficient physical separation between transmit and receive antennas, with the transmitted signal vector constrained to have overall power $E[|x|^2] \leq P$, and circularly symmetric Gaussian background noise with covariance matrix $\Phi_N = \sigma^2 I$.

In order to make the analysis more tractable, the multicell scenario is usually simplified to a linear array of cells and the interference from the two adjacent cells is characterized by a single attenuation factor [18]. To provide a common framework that is general enough to address multiuser detection across the cell while remaining simple enough for analysis and simulation, we assume such a model that there are four interferers in two groups of two, in which one group is much stronger than the other. This roughly reflects the essential reality as interference from two farthest adjacent cells can typically be ignored, and simulation results verify that the power of the two strongest users usually dominates. The system model used in our study is given by

$$y = H \cdot x + \sum_{i=1}^{2} H_{yi} \cdot x_{y_{i}} + \sum_{i=3}^{4} H_{yi} \cdot x_{y_{i}} + n.$$  \hspace{1cm} (1)

Let $P_{y_{i}} = E[|x_{y_{i}}|^2]$, we have $P_{y_{1}} = aP_{y_{2}}$, $P_{y_{3}} = bP_{y_{4}}$, and $(P_{y_{1}} + P_{y_{2}})/(P_{y_{3}} + P_{y_{4}}) = \gamma > 1$. The $H$ and $\{H_{yi}\}$ matrices are assumed to be independent with independent and identically distributed (i.i.d.) normalized complex Gaussian variables as entries. The signal-to-noise ratio (SNR) is given by $\rho = P/\sigma^2$, and the signal-to-interference ratio (SIR) is given by $\eta = P / \sum_{i} P_{y_{i}}.$
III. ADVANCED TECHNIQUES

A. Space-Time Layered Architecture

The first Bell-labs space-time layered architecture (BLAST) is named D-BLAST [6] as it advocates an elegant diagonally layered space-time coding scheme in which the individual code blocks are distributed along diagonals in the space-time plane. To reduce the implementation complexity, a simplified BLAST architecture, V-BLAST, is proposed as a substitute in [8]. In contrast to the inter-substream coding introduced in D-BLAST, in V-BLAST the vector encoding process is simply a demultiplex operation followed by independent symbol mapping of each substream. As each substream in V-BLAST is tied to a fixed antenna element, it does not make use of the transmit diversity as D-BLAST does. The decoding in the BLAST architecture is essentially a decision feedback multiuser detection, with each substream being taken as a different user. The first detected substream of the V-BLAST will essentially determine the overall system performance, since it has the least receive diversity degree due to interference cancellation. Errors created in the detection of this substream will also influence the performance of the other streams due to error propagation effects. Layer ordering based on the SNR was proposed in the implementation of V-BLAST to remedy this degeneration in receive diversity.

The optimum possible reception scheme is maximum likelihood (ML) detection. We know from multiuser detection theory that joint maximum likelihood multiuser detection will retain the receive diversity for each user; however, it suffers from exponentially increasing complexity as the dimensionality of the problem increases. An alternative sub-optimum scheme is multistage interference cancellation (IC), which often achieves satisfactory results with much lower complexity. The ML solution is one of the fixed points of the multistage IC algorithm, however, the IC approach has the problem of oscillatory behavior and possible lack of convergence [3]. We will see that, with the introduction of a type of turbo processing discussed in the next sub-section, optimum performance will be approached within several iterations between sub-optimum demapping/detection and decoding stages.

B. Turbo Processing

In recent years, iterative processing techniques with soft-in/soft-out (SISO) components have received considerable attention. The basic idea is to break up optimum joint signal processing, which are typically very complex and require large amounts of memory, into separate components, iterating between them with the exchange of probabilities or “soft” information. This approach typically performs almost as well as the much more complex “joint” approach which attempts to achieve the exact ML or MAP optimization. This so-called turbo principle is exemplified through turbo decoding [9], turbo equalization [5] and turbo multiuser detection [12] with application to wireless [17] and wireline [4] communications.

The very same principle can be applied to the BLAST system, resulting in coded V-BLAST and Turbo-BLAST, shown in Fig. 1 and Fig. 2, respectively. For the coded V-BLAST, the information bits are first demultiplexed into \( N \) sub-streams, each of which is independently encoded, interleaved, and symbol-mapped. At the receiver, the MMSE criterion is used to decouple the substreams; then for each substream a soft metric is calculated and fed to the SISO MAP decoder, which produces soft estimates of information and coded bits, used to refine soft metric calculation in the next iteration. After several iterations within a layer, the estimated bits are good enough to be used as output as well as to be fed to the next layer to assist in detection. For Turbo-BLAST, the information bits are coded (not necessarily with Turbo codes) and interleaved as a whole; then the whole coded stream is demultiplexed into \( N \) substreams and symbol-mapped individually. At the receiver, the entire data stream is processed iteratively between a soft metric calculation stage and a decoding stage. Note that in the soft metric calculation stage, either a maximum likelihood joint detection or a MMSE multistage parallel interference cancellation (PIC) scheme can be used. We will show that these two schemes achieve the same performance, owing to the turbo processing.

1. Turbo-BLAST Detection

The turbo decoding procedure of the coded V-BLAST is exactly analogous to that of the Turbo-BLAST to be discussed and therefore is omitted here. The Turbo-BLAST detection algorithm involves two components: decoding and decoding. The MAP algorithm is employed in the decoding stage to take in soft metrics from the demodulation stage and produce soft estimates of information and coded data bits. The demodulation stage with ML is straightforward. Suppose an \( N \times N \) MIMO system is employed by one cell, and each substream adopts \( M \)-QAM. Then for each symbol interval \( B = N \cdot \log_2 M \) bits are jointly detected. The extrinsic information for the \( i\)-th bit, \( 1 \leq i \leq B \), is given by

\[
L_i(i) = \log \frac{\sum_{x \in X_i^1} p(y | x)p(x)}{\sum_{x \in X_i^2} p(y | x)p(x)} - L_0(i),
\]

where \( p(y | x) \) is a multivariate Gaussian distribution;

\[
p(x) = \prod_{i=1}^{B} p(b_i) \text{ and } L_0(i) = \log \left( \frac{P(b_i = 1)}{P(b_i = -1)} \right)
\]

comprise a priori information from the decoding stage; and

\[
X_i^1 = \left\{(x_1, x_2, \ldots, x_N) : b_i = 1\right\}, \quad X_i^2 = \left\{(x_1, x_2, \ldots, x_N) : b_i = -1\right\}.
\]

The demodulation stage with PIC is subter. Suppose the received signal for some substream \( 1 \leq k \leq N \) after interference cancellation is given by

\[
\hat{y}_k = \mathbf{H}(x - \hat{x}_k) + \mathbf{n},
\]

where \( \hat{x}_k = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_k = 0, \ldots, \hat{x}_N) \) is the estimated interference vector. First an MMSE filter is applied to \( \hat{y}_k \) to further suppress the residual interference plus noise, that is,

\[
w_k = E[\hat{y}_k \hat{y}_k^*]^{-1} E[\hat{y}_k] = (\mathbf{h}_k \mathbf{h}_k^H + \mathbf{H}_k \mathbf{Q} \mathbf{H}_k^H + \frac{N}{\rho} I)_k^{-1} \mathbf{h}_k,
\]
where $h_k$ is the k-th column of matrix $H$, $H_k$ is the complement of $h_k$ in $H$, and

$$Q = \text{diag}(1 - \frac{\sigma^2}{P} | x_1 |^2, \ldots, 1 - \frac{\sigma^2}{P} | x_{N-1} |^2, 1 - \frac{\sigma^2}{P} | x_{N+1} |^2, \ldots, 1 - \frac{\sigma^2}{P} | x_{2N} |^2) \tag{5},$$

which approaches 0 when estimates from the decoding stage are accurate enough for constant-modulus signals. As is shown in [13], the output of the MMSE filter $z_k = w_k^H y_k$ can be written as

$$z_k = \mu_k x_k + \eta_k \tag{6},$$

where $\mu_k = \frac{N}{P} E[z_k x_k^*] = w_k^H h_k$ and $\eta_k$ is a Gaussian variable with zero mean and variance

$$v_k^2 = E[|z_k - \mu_k x_k |^2] = E[|z_k |^2] - \frac{\sigma^2}{P} | \mu_k |^2 = \frac{\sigma^2}{N} (\mu_k - | \mu_k |^2).$$

The extrinsic information is given in the same form as (2), but with $y$ replaced by $z_k$ and $x$ with $x_k$, and therefore with much lower complexity.

C. Multiuser Detection

We have already discussed various MUD schemes for detection of different substreams within a MIMO system (intercell interference). Here we will focus on exploiting MUD to combat interference of the same format from adjacent cells (intercell interference).

1. Maximum Likelihood MUD

Maximum likelihood multiuser detection is infeasible for most current applications due to its complexity. Suppose an $N \times N$ MIMO system is employed by one cell, and each substream adopts $M$-QAM. Then the ML-MUD complexity would be in the order of $M^N$. If we want to jointly detect all the information bits for users from desired and K-1 interfering cells, then the complexity would go to the order of $M^{NK}$. Even if we assume the simplest scheme such as $M = 4$, $N = 2$ and $K = 5$ (ignoring the two weakest interfering cells of the first tier), the complexity would be in the order of $2^{20}$, which is unfeasible for all current practical systems.

2. Linear MMSE MUD

We assume knowledge of channel information for the interfering users, which can be obtained either through an initial joint training phase with the coordination of base stations, or through adaptive tracking algorithms from the received signals directly. MMSE MUD, which is generally the most favorable linear MUD, has a detection matrix given by

$$W = \left( HH^H + \sum_{g=1}^{P} \frac{P}{\rho} H_{gP} H_{gP}^H + \frac{N}{\rho} I \right)^{-1} H \tag{7}.$$ 

Thus, the detection process would be to first apply the weight matrix of (7) to the received signal (1) to combat CCI; and then to process the modified signal as in Subsection III-B. As we mentioned, linear MMSE MUD cannot effectively suppress the intercell interference as the receive antenna array does not have enough degrees of freedom. However, the distribution of the residual interference plus noise at the output of a linear MMSE multiuser detector is well approximated by a Gaussian distribution [13]. This nice property will guarantee good performance of the Gaussian-metric-based receivers, which would otherwise deteriorate greatly in a multiuser environment.

3. Linear Channel Shortening MUD

Another linear MUD technique of interest to combat the intercell interference is the so-called channel-shortening multiuser detector, which employs a slightly different optimization criterion than linear MMSE MUD above [11]. For detecting data originating in the desired cell, the idea is to apply some form of array processing to maximize the SINR, where the signal power refers to the power contributions of all the users in the cell to be detected, while interference refers to the power contributions of users in other cells. Note that this criterion is different from linear MMSE MUD (which also maximizes the SINR) in which the signal refers to the very user to be detected while all other users both in cell and out of cell are treated as interferers. In short, the optimal detection matrix for channel-shortening linear MUD is the collection of the first $N$ principal general eigenvectors of the matrix pencil

$$\left( HH^H + \sum_{g=1}^{P} \frac{P}{\rho} H_{gP} H_{gP}^H + \frac{N}{\rho} I \right).$$

This scheme also serves as a linear pre-processing stage, often followed by much more complex processing, such as ML processing, within the desired cell.

4. Group IC MUD

Since ML-MUD is highly complex, while linear MUD is limited in its interference cancellation capability, non-linear MUD often serves as a tradeoff between performance and complexity. In the context of multicell MIMO systems, group detection techniques naturally call for attention, in which information bits for one group (one cell MIMO) are detected at a time. Following a natural extension from BLAST, we can detect one MIMO system at a time, and feed decisions to other group detectors for interference cancellation. Note that generally, the success of interference cancellation relies on the correct detection of interference. In adverse environment where we cannot get good estimates of interference, IC schemes will worsen the performance instead of improving it. The potential benefit of group IC MUD depends highly on the interference structure, which will be further addressed in the next section.

IV. SIMULATION RESULTS

In Section III, various potential advanced techniques have been introduced, the combination of which could yield many detector structures. We now compare them, based on the model (1), to see which one performs best in the interference-limited environment. The performance measure is the block-error rate (BLER) over Rayleigh fading channels. We assume that each cell employs a $4 \times 4$ MIMO system, operating at $\text{SNR} = 30 \text{dB}$. The modulation scheme employed is 4QAM. The coding scheme used is a rate-$1/3$ 64-state convolutional
code with generators \((G_1, G_2, G_3) = (155,117,123)_8\) (proposed for EDGE). It was shown in our simulations that this code achieves better performance than a turbo-code with two identical 16-state recursive encoders with generators \((G_1,G_2) = (23,31)_8\), at a considerably lower complexity. We transmit blocks of 384 information bits, and record the block error probability of this system.

The receiver structure is either V-BLAST or Turbo-BLAST, combined with various MUD schemes. To be specific, the receivers we study are: 1) Coded V-BLAST (V-BLAST); 2) Coded V-BLAST with linear MMSE MUD pre-processing (V-BLAST+MMSE); 3) Turbo-BLAST with a parallel interference cancellation demodulation stage (T-BLAST (PIC)); 4) Turbo-BLAST with a parallel interference cancellation demodulation stage, with linear MMSE MUD pre-processing (T-BLAST (PIC)+MMSE); 5) Turbo-BLAST with a maximum likelihood demodulation stage (T-BLAST (ML)); 6) Turbo-BLAST with a maximum likelihood demodulation stage, with linear channel shortening MUD pre-processing (T-BLAST (ML)+CS); 7) Turbo-BLAST with a parallel interference cancellation demodulation stage, with group IC MUD (T-BLAST (PIC)+IC). We study the performance of these receivers in the framework of (1) in two situations:

\[ \text{(A) } P_{q_1} = P_{q_2} = 4P_{q_3} = 4P_{q_4} \quad \text{and (B) } P_{q_1} = 6P_{q_2} = 6P_{q_3} = 6P_{q_4}. \]

The simulation results for situation (A) are shown in Fig. 3, from which we can see that: 1) Turbo-BLAST offers both diversity and coding gain over V-BLAST; 2) Turbo-BLAST with a PIC demodulation stage performs as well as Turbo-BLAST with ML stage, while it has much lower complexity; 3) Linear MUD pre-processing offers a considerable performance gain in interference-limited environments; and 4) Group IC MUD worsens the performance due to incorrect decision feedback. Note that we attempt to detect all interfering signals in this case.

The failure of group IC MUD is owing to the inability to correctly detect the information bits for interfering cells. There are both theoretical and practical reasons for the errors in the detection of the interfering signals. The practical reason is that the codes that we used in this simulation are comparatively simple, and thus cannot correct all the errors that an “ideal” code could eliminate. However, there is also a theoretical limit: with ideal codes, the codes in neighboring cells would be designed to have rates that achieve capacity in that cell. However, they suffer more attenuation when propagating to the neighboring cell (where they are interferers). The signal-to-noise-ratios of those signals in the neighboring cells are thus worse, so that the data rate is above the capacity of the link to the neighboring cell. Thus, correct decisions for the symbols of interfering signals might not be possible even theoretically.

We would expect that when we have only one dominant interfering signal, group IC MUD will outperform linear MMSE. Therefore, it is worth studying the performance of group IC MUD only for the strongest interfering signal when there is one dominant interferer. The simulation results for situation (B) are shown in Fig. 4. We see that group IC MUD only for the strongest interfering signal achieves the best performance, which is about 4dB and 8dB over Turbo-BLAST and coded V-BLAST, without MUD, respectively, and more than 2dB over Turbo-BLAST with linear MMSE preprocessing, at 1% BLER. (Since T- BLAST (ML) assumes no advantage over T-BLAST (PIC) while having much higher complexity, we do not consider it further.)

We have noticed that group IC MUD performs the best when one interferer dominates. But when two interferers dominate that have the same power, it is no better than the simpler linear MMSE scheme. It is also shown that in the two-dominant-interferer scenario, when the ratio between the two largest interferer power increases, the gap between the performance of group IC MUD and linear MMSE also increases (not presented here due to space limits). In the view of this performance, an idea for adaptive detection arises: namely, in the case of one dominant interferer (3dB or greater) or in the case of two dominant interferers (4dB or greater) with the ratio between the two largest interferer power greater than 3dB, group IC MUD could be adopted, otherwise the simple MMSE scheme could be adopted.

V. CONCLUSIONS

In contrast to the single-cell MIMO system considered in previous studies, where the intercell interference, when accounted for, is added to ambient Gaussian noise, we take the approach of modeling the whole downlink cellular system as a broadcast/interference channel [1], the capacity of which has long been an open question. After discussing the merit of the space-time layered architecture and turbo coding/processing techniques, we have considered multiuser detection for combating intercell interference. Among various multiuser detection techniques examined, linear MMSE and successive interference cancellation have been shown to be feasible and effective. Successive cancellation plays a major role in network information theory from both theoretical and practical points of view. As is known, decoding of the interfering users is not always optimal except in the strong-interference case, nor is treating them as pure ambient noise optimal, except in the very-weak interference case. Based on this phenomenon, we have proposed an adaptive detection idea that offers improved performance.

ACKNOWLEDGMENT

The authors would like to thank Dr. Larry Greenstein, Prof. Peter Driessen, Dr. Jack Winters, Prof. H. Vincent Poor, Dr. Yank Soek Choi, Dr. Martin Clark, and Dr. Moe Win for helpful discussions, and Dr. Severine Catreux for providing us with her results.

REFERENCES

[3] H. Dai and H. V. Poor, “Iterative space-time processing for multiuser detection in multipath CDMA systems,”


