Adaptive CSI Prediction in Linear Multi-User MIMO Systems

Bartosz Mielczarek
Feedback Limited
Edmonton, AB, Canada T5M 0N8
Email: bmielcza@feedbacklimited.com
Witold A. Krzymień
University of Alberta/TRLabs
Edmonton, AB, Canada T6G 2V4
Email: wak@ece.ualberta.ca

Abstract—In this paper, we discuss channel adaptation techniques in linear multi-user multiple-input multiple-output (MU-MIMO) systems with imperfect Channel State Information (CSI) caused by limited feedback bandwidth and feedback channel delay. We assume that such systems will operate in environments, where channel characteristics constantly change due to the movement of mobile terminals.

We consider an approach where the impact of delayed CSI is partially mitigated by channel prediction performed at the base station side. In this paper, we use a dictionary based predictor that operates on phrases consisting of indices of quantized channel state information. To address variations in channel characteristics, we propose an algorithm that adapts the prediction dictionary to the current channel state.

We demonstrate the performance of our algorithms in broadcast MU-MIMO systems with and without CSI prediction operating on a channel with changing Doppler rates. The simulation results show that significant sum-rate gains are achievable for a large range of feedback delays. The additional advantage of our proposed adaptive prediction algorithms is that they can be used with any type of CSI quantization implementations and contribute very little extra complexity to the base station.

Index Terms—MU-MIMO systems, channel state information, channel prediction, channel adaptation, feedback channel.

I. INTRODUCTION

One of the most promising solutions for increased spectral efficiency in high capacity multi-user wireless systems is the use of multiple antennas on fading channels. The fundamental issue in such systems is availability of the channel state information (CSI) at the hub transmitters and mobile user receivers. In general, if the mobile receivers and base station transmitter have access to CSI, the system throughput can be significantly increased. However, while it is usually assumed that perfect CSI is available at the mobile receivers, the hub transmitter may only have partial CSI due to the feedback channel delay and errors, channel estimation errors and, primarily, limited reverse channel bandwidth, which forces CSI to be quantized at the mobile to minimize feedback rate [1]–[8].

In [9], [10], we proposed to alleviate those feedback channel imperfections by using channel prediction algorithms that work with the quantized CSI information. The common part of these algorithms is transmitter-side prediction (extrapolation) of the correct channel information codebook index based on the previously received indices. The predictors can be implemented as either look-up tables or as channel dictionaries but regardless of the method used, they work directly with integer CSI indices. Our approach has three significant advantages over the conventional direct channel coefficient prediction such as in [11]:

1) It works with any channel model,
2) It can be used with any CSI quantization method,
3) Integer operations are much less complex than fixed or floating point operations required for direct channel coefficient prediction.

In general, however, a simple static predictor may not suffice when the channel properties change between transmission epochs. For example, a relative speed of mobile terminals will change all the time, causing fluctuations of the Doppler shift of the received signal. Moreover, in urban areas, a mobile user may frequently move in and out of the direct line of sight of the base station which alters the characteristics the received signal. In such cases, the CSI predictors should be able to dynamically adjust their underlying channel model to allow for variations of speed, correlation etc.

This paper extends the work presented by us in [10] by introducing an adaptation mechanism to update the underlying dictionaries of the CSI index predictor. We consider a system where channel changes its statistics with time and the channel state information at the hub transmitter (CSIT) is delayed.

We propose three algorithms that are able to cope with such situations and analyze their results in terms of prediction success rate and sum-rate improvements.

In Section II, we introduce the system model and assumptions. In Section III, we discuss the basic idea behind the compensation of feedback delay using channel state index prediction. In the following Section IV, we address the prediction issues arising from changing channel characteristics and introduce three methods of alleviating those issues. In Section V, we show the method for adaptive creation channel state dictionaries at the base station. In Section VI, we present the results of the simulations and, finally, in Section VII, we summarize the paper and present directions for future research.
II. SYSTEM MODEL

We assume that the communication system consists of a transmitter equipped with $n_T$ antennas and $K \geq n_T$ mobile receivers with $n_R^k$ antennas, where $k = 1, 2, \ldots K$. The mobile user channels in the $l$th epoch are modeled by a set of complex channel matrices $\mathbf{H}^k(l)$ of dimension $n_R^k \times n_T$. The received signal of the $k$th user is then given by $n_R^k$-element vector $y^k(l)$ defined as

$$y^k(l) = \mathbf{H}^k(l)x^k(l) + n^k(l)$$

where $x^k(l)$ is the $n_T$-element vector of the transmitted signal and $n^k(l)$ is the $n_R^k$-element vector containing independent circularly symmetric complex Gaussian entries with zero means and unit variances. Matrices $\mathbf{H}^k(l)$ are time-correlated between time epochs $l$ and their elements may or may not be spatially correlated as well. Note that the proposed algorithms work with any statistical channel models as long as $\mathbf{H}^k(l)$ are temporally correlated (fundamental requirement for predictive algorithms). In this paper, we assume that the elements of matrices $\mathbf{H}^k(l)$ are correlated in time as specified by the Clark’s model [12]. The frame duration is given as $T_{\text{frame}}$ and we normalize the Doppler frequency as $f_D^k(l)T_{\text{frame}}$, where $f_D^k(l)$ is the maximum Doppler shift of the $k$th user in the $l$th epoch. The system works as follows

1) Before every transmission epoch $l$, each mobile receiver $k$ estimates its channel matrix $\mathbf{H}^k(l)$.
2) The receiver CSI is quantized by selecting best-matching channel information codeword using any method of choice.
3) The indices of the selected CSI codewords from all mobile receivers are fed back to the hub transmitter. It is assumed that reverse channel introduces delay of $\Delta$ frames.
4) The base station uses received delayed CSI indices in a prediction algorithm to extrapolate actual CSI indices at mobile receivers.
5) The base station uses the predicted indices to choose the best broadcast combination of mobile receivers.
6) The signal is multiplexed and transmitted to the selected subset of mobile receivers.

This paper addresses problems arising in the third and fourth steps of the above algorithm, when the statistical characteristics of $\mathbf{H}^k(l)$ change and the predictor has to adjust its underlying channel dictionary continually.

III. BASIC PREDICTION ALGORITHM

The typical feedback channel imperfections are transmission errors, erasures and delays. All those problems may cause the base station to schedule the transmission in non-optimum way, impacting the throughput in the served cell. In [9], we showed how to use the channel index prediction techniques to mitigate all those problems. In this paper, we concentrate on the most severe and unavoidable problem of reverse channel delay when feeding back the indices from a mobile receiver to the hub transmitter.

In order to alleviate feedback delay that results in base station using outdated channel state indices, channel predictor processes the received channel state indices and attempts to extrapolate the most probable actual index. As shown in [9], [10], such an extrapolation can significantly improve system’s performance at the minimum increase of hardware complexity as the extrapolation of integer indices is orders of magnitude less complex than direct channel coefficient prediction.

The fundamental premise for CSI index prediction is that, for a given channel coherence time and transmission epoch duration, the CSI index trajectories are constrained with high probability to a limited set of allowed paths within a CSI quantization space. If these paths can be statistically characterized, the basic channel prediction problem is to select the most probable future value of the vector quantizer index $I[\ell + \Delta]$, which is separated $\Delta$ transmission epochs from the current one $I[\ell]$. To model the predictor, we use a function

$$f(I[\ell], I[\ell - 1], I[\ell - 2], \ldots, I[\ell - L + 1])$$

which, based on up to $L$ past indices, returns pairs $(I[\ell + \Delta], f)$, which contain possible future CSI indices at the $I[\ell + \Delta]$ epoch together with their relative frequencies $f$.

There are two basic approaches to the CSI index prediction [9], [10]. The first one is based on a set of matrices $\mathbf{P}$ containing the statistical information about the channel behaviour together with the given CSI quantizer. The second technique of constructing the prediction tables (used in this work) is based on relative sparsity of the prediction matrices. This fact can be used to construct sequences of valid indices, called henceforth phrases, that are used by the prediction algorithm to identify valid CSI transitions and select values of the index $I[\ell + \Delta]$. Based on the example in Fig. 1, two phrases in such a dictionary would be $D_1 = 317$ and $D_2 = 315$ appearing with relative frequencies 0.8 and 0.2, respectively. Such a predictor will contain a dictionary of phrases, corresponding to all allowed index sequences for a given quantizer structure and channel characteristics.
Fig. 2. Dictionary size for different normalized Doppler shifts $f_D T_{frame}$ and number of past samples $L$ used in the predictor. $\Delta = 1$

IV. ALLEVIATION OF THE VARYING CHANNEL CHARACTERISTICS

The channel index prediction algorithms depend on the quality of the underlying channel dictionary to provide index extrapolations with high success rate. As a result, when channel characteristics change between the epochs $l$, the channel dictionary must also change to include phrases reflecting the current channel characteristics best.

1) The base station has no a-priori channel prediction data and builds a channel dictionary online for each mobile terminal.

2) The base station has a set of a-priori channel dictionaries, created for different types of channels, and uses the one that matches the instantaneous channel characteristics best.

3) A hybrid approach, where the base station starts with the known a-priori channel dictionary and updates it adaptively.

The advantage of the first approach is that the base station can work with any type of channel, as long as it is temporally correlated. Every realization can be tracked and the channel state dictionary updated accordingly. The disadvantage of this method is that sometimes it may take a considerable amount of time to adaptively create a good quality dictionary (fast fading, long feedback delays), the time that may not be available to the base station if the channel characteristics change often.

As an example, Fig. 2 shows dictionary sizes for different Doppler shifts and number of past CSI indices $L$. For fast changing channels, full size dictionaries can reach hundreds of thousands of phrases. In realistic situations, a channel would change long before the dictionary would be fully adapted.

The advantage of the second approach is that in certain instances, when the pre-computed dictionary matches the current state characteristics very well, the predictor success rate will be very high. However, the system will not have a flexibility to adapt to actual channel variations and the dictionary storage requirements may be very high as seen in Fig. 2.

The third, hybrid approach, if designed in a proper way, may alleviate problems with both the above approaches. The predictor will be able to either use the pre-defined dictionary or adapt it to the current channel characteristics, hopefully improving the overall system sum-rate.

V. PREDICTOR ADAPTATION ALGORITHM

In order to allow a CSI predictor to adaptively adjust its channel model to increase the success rate of its extrapolation, we implemented predictor adaptation algorithm based on sliding window approach applied to the incoming CSI indices. The algorithm works as follows:

1) Initialize system operation without predictor.

2) Perform regular transmission operation collecting received CSI indices $I$.

3) Parallel with step 2, update predictor channel as the CSI indices are received by the base station.

4) When $M$ indices have been received, switch the predictor on.

5) At any transmission epoch $l$, use predictor to extrapolate next index $I[l + \Delta]$ and use it in regular transmission scheduling.

6) After the actual $I[l + \Delta]$ index is received, remove the first entry in the CSI index list and store it in a temporary buffer. Reduce the histogram value of the corresponding predictor channel model entry by 1.

7) Add the $I[l + \Delta]$ index to the end of the CSI index list.

8) Update the predictor channel model as follows:

   a) If the predictor could not be used to correctly extrapolate the new index $I[l + \Delta]$, update the predictor channel model with a new entry.

   b) If the predictor could correctly predict the new index $I[l + \Delta]$ without ambiguity, go to 5.

   c) If the new index $I[l + \Delta]$ could be predicted but with ambiguity, check if temporary buffer entries could be used to expand the current predictor’s channel model to eliminate ambiguous extrapolation.

   d) Update the channel model histogram values;

9) Increase $l$ with 1.

10) Go to 5.

The basic premise of the predictor operation is to eliminate, as time progresses, the oldest entries in the channel model and exchange them with the new ones. The crucial parameter of this process is the length of the sliding windows $M$. With too short window, the dictionary will never be fully trained. When $M$ is too large, the dictionary may contain old phrases that are no longer relevant. In future papers, we will show another layer of adaptation that changes the length of $M$ based on the measured success of prediction.
VI. SIMULATIONS

We have implemented the presented adaptive prediction method in a system using a base station with $n_T = 2$ and two mobile users $K = 2$ with identical statistical properties and $n_R(k) = n_T = 2$. In order to demonstrate the sum-rate improvement of the adaptive CSI prediction, we used a linear block diagonalization approach [13] with the CSI codebooks representing the dominant eigenmodes of the channel matrices [14]. Note, that any type modulation method and CSI codebook designed could be easily used as our predictors work solely with integer indices received from mobile terminals.

Each setup has been simulated using 100,000 channel trace realizations with complex Gaussian entries in $H_k$. To model changes in Doppler shift we simulated the system, in which two users were moving with a constant speed of 60 km/h on a street 1 km from the base station. The users would enter the base station range at a distance of 0.5 km from the perpendicular bisector of the street drawn from the base station and travel a distance of 1 km, see Fig. 3. At frame duration of 2 ms, corresponding to HSPDA [15], the number of frames transmitted to the mobile terminal will be thus equal to 30,000. Correspondingly, at the carrier frequency of 2 GHz, the resulting normalized Doppler shift varies from $f_{LD}(k)T_1 = 0.1$ at the range border to $f_{LD}(k)T_{15000} = 0$ when the mobile terminal reaches the bisector line and again $f_{LD}(k)T_{30000} = 0.1$.

We assumed that the base station was able to track the mean received power of all the users. This assumption was validated by the fact that the rate of change of large-scale fading processes was orders of magnitude lower than the rate of change of small-scale processes. Moreover, we implemented the dictionary with maximum number of past samples used to predict the future indices equal to $L = 3$.

In Fig. 4, we present the aggregate sum-rates for the considered scenario with and without the basic adaptation algorithm. For both considered reverse channel delays $\Delta = 1$ and $\Delta = 3$, one can see sum-rate improvements in range of 1-2 dB for the shorter delay and 2-4 dB for longer delay.

Figs. 5 and 6 show the interaction of reverse channel delay $\Delta$ and the length of the predictor sliding window $M$ in terms of contribution to the sum-rate of the system. As one can see in Fig. 5, the adaptive system’s sum-rates increase with increasing sliding window length $M$ for system with long reverse channel delay and decrease when channel delay is shorter. This is a result of longer training required to create a dictionary for a system with long reverse channel delay and the corresponding number of past samples $L$ as shown in Fig. 2.

It is also interesting to compare the time traces of the sum-rates for such situations as shown in Fig. 6. As one can see, for small reverse channel delay and short sliding window, the adaptive system maintains almost constant gain, while for the
other situation one can clearly see that there is a loss of performance when old channel entries are still used by the system even when channel characteristics have changed.

Finally, in Fig. 7 one can see an implementation of a hybrid approach, where the system uses 20 pre-computed dictionaries corresponding to the range of normalized Doppler between 0.1 and 0.01. The system selects the best dictionary for current channel conditions and uses it together with the above adaptive algorithm when the channel changes. This solution adds further 1-2 dB to the gains in Fig. 4 for $\Delta = 3$.

VII. CONCLUSIONS AND FUTURE WORK

We have considered multi-user MIMO system with limited feedback operating on a changing flat fading channel with reverse channel delay. We extended our earlier proposed channel state index prediction approach to include channel dictionary adaptation. The proposed solution is easily implementable in any system using quantized CSI and does not introduce a significant processing burden on either the base station or the mobile terminal. We have shown that system sum-rate gains can be significant for a wide range of channel Doppler spreads and reverse channel delays.

Our future work will be focused on more detailed analysis of adaptive prediction algorithms, especially on the problem of selection between multiple prediction possibilities.

ACKNOWLEDGMENT

Funding for this work has been provided by TRLabs, the Rohit Sharma Professorship and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

REFERENCES