Design of Differential Quantization for Low Bitrate Channel State Information Feedback in MIMO-OFDM Systems

Woon Hau Chin and Chau Yuen
Modulation and Coding Department
Institute for Infocomm Research
Singapore 119613
Email: w.h.chin@ieee.org, cyuen@i2r.a-star.edu.sg

Abstract—The feedback of channel state information have always posed a problem for the design of a communications system where the information is required at the transmitter. In most cases, changing the quantization levels or size of the codebook is used to control the amount of feedback bits. However, the number of quantization levels must be maintained at a sufficiently high level such that the feedback information can still be of use to the transmitter. This paper discusses the use of differential quantization methods to reduce the number of feedback bits required.

I. INTRODUCTION

Channel state information (CSI) is sometimes required at the transmitter for designing schemes which can improve the overall performance of the system. These include adaptive modulation, power control, and transmit beamforming. However, due to the high bandwidth required for full CSI feedback, usually only limited feedback, such as signal to noise levels, are implemented. Limited feedback are insufficient for use in many precoding schemes.

Orthogonal frequency division multiplexing (OFDM) is a transmission scheme which divides the channel bandwidth into multiple narrowband channels. The transmitter and receiver will effectively see multiple flat fading channels in place of a single frequency selective fading channel. However, due to the multiple subchannels/subcarriers introduced by OFDM, CSI feedback in OFDM systems are usually not feasible due to the large amount of parameters involved. For multiple antenna systems, or multiple-input/multiple-output (MIMO) systems, this is even more difficult as there are now multiple channel state information per-subcarrier to feedback.

There are several techniques currently used for reducing CSI feedback. Many of them revolve around quantizing the CSI. The IEEE P802.11n draft standard [1] also includes CSI feedback. The method was to use Givens Rotation to compress the MIMO CSI for each subcarrier. However, the scheme would still require a large amount of feedback due to the large number of subcarriers present.

A commonly used method for representing correlated sequences is to quantize the difference between two consecutive elements instead of quantizing the elements themselves. This was first used in speech coding [2], commonly known as differential pulse code modulation (DPCM) or delta modulation (DM) for the single bit version. Such schemes, their simplest form, have recently been applied to communication problems, where delta modulation was to encode the feedback for time varying channel information [3], and for encoding feedback for OFDM [4]. [5] also looked at using adaptive DM for tracking Givens parameters in time varying channels.

In this paper, we proposed the use of differential quantization schemes for encoding feedback information in MIMO-OFDM systems. Here, we make use of differential quantization to exploit the frequency correlation of the subcarriers, as compared to other works which deals mainly with time correlation. Additionally, we propose the use of multilevel quantization as opposed to single quantization which was used in other literature. We will show that this is critical, especially in certain channel conditions, for a respectably accurate feedback. We also discuss the design of the differential quantizer and the Givens parameters such that the performance of the feedback scheme is optimized.

The paper is organized as follows. In the following section, we describe the differential quantization process and the choice of the dynamic range. This is followed by a description of using Givens Rotation for the MIMO-OFDM system and the design of the Givens parameters. We then briefly compare the feedback rate required for the proposed scheme and normal quantization. In Section V, we present some simulation results showing the performance of the proposed scheme.

II. DIFFERENTIAL QUANTIZATION

Differential encoding [2] is a technique frequently used in speech coding to reduce the number of bits used to represent the speech signal. The encoding exploits the strong time correlation of speech signals. The most commonly used schemes used for differential encoding is the differential pulse code modulation.

The motivation behind differentially encoding of consecutive samples is to exploit the correlation which exist between the samples so as to reduce the amount of storage required for
their representation. This is only feasible if
\[ E[(x_n - x_{n-1})^2] < E[x_n^2], \]
i.e. the variance of the sample is greater than the variance of the difference of the samples. As a result, the storage required to represent the difference (up to a certain degree of accuracy) is less than that required to represent the sample itself [2].

In OFDM systems, adjacent subcarriers are usually correlated. The strength of the correlation between the adjacent subcarriers is inversely proportional to the maximum delay of the fading channel. Therefore, in situations where the time delay of the channel is small, the variance of the fading coefficient difference between adjacent subcarriers is small, and this increases with the time delay of the fading channel.

The scheme works by encoding the difference between the adjacent samples. While traditionally, DPCM also incorporates a predictor, we assumed that only the immediate past sample is used. This is to reduce the overhead of the scheme and to reduce the complexity of the encoder and decoder.

Suppose that we are to differentially quantize a sequence \( x_n, n \in \{1, \ldots, N\} \). The encoder starts with the first sample, \( x_1 \), and proceeds sequentially. We assume that the first sample, \( x_1 \), is quantized separately using \( A \) bits. At the \( n \)th sample, the difference between the current sample and quantized version of the previous sample was acquired and quantized,
\[ \beta_n = Q(x_n - \hat{x}_{n-1}), \]
where \( Q(\cdot) \) is the quantization function whose output is dependent on the number of bits assigned to the quantization, \( B(\ll A) \), and the dynamic range, \( \Delta \). This is followed by constructing the quantized output for the current sample,
\[ \hat{x}_n = \hat{x}_{n-1} + \beta_n. \]

The output of the encoding algorithm is the first sample(\( x_1 \)) quantized at \( A \) bits followed by a string of \( N - 1 \) differences (\( \beta_n \)) quantized at \( B \) bits.

A special case of DPCM is where the number of bits used for quantization is 1. This method is commonly known as delta modulation.

The decoding process is the inverse of the encoding process. The algorithm starts with the available first sample, \( \hat{x}_1 \) and sequentially reconstructs the following sample by
\[ \hat{x}_n = \hat{x}_{n-1} + \beta_n. \]

### A. Dynamic Range

The dynamic range for the quantization is an important factor in the performance of the whole quantization process. If the chosen dynamic range is too small, the encoding would not be able to represent parts of the sequence where there are large fluctuations, resulting in what is commonly known as slope overload distortion. On the other hand, if the chosen dynamic range is too large, encoding parts of sequences where the differences are small would result in what is known as granular noise. As the encoding process is sequential, there is no close form solution for the optimal dynamic range.

However, a ‘rule of thumb’ guide for choosing the dynamic range for speech signals was given as [2],
\[ \Delta = \sqrt{\frac{1}{N - 1} \sum_{n=1}^{N} (x_n - x_{n-1})^2 \log_2 2}, \]
where \( N \) is the length of the sequence.

The characteristics of communication channel parameters, however, is in general vastly different from that of speech signals. In order to adequately describe the channel parameters which can fluctuate abruptly especially in long delay scenarios, we use the following criteria for choosing the dynamic range,
\[ \Delta = \alpha \max\{|x_n - x_{n-1}|\}_{n=2}^{N}, \]
where \( \alpha \) is a scalar multiplier. The \( \alpha \) parameter represents the tradeoff between having slope overload distortion and granular noise. We will show later through simulations that the optimal range of \( \alpha \) for a small number of quantization bits is 0.7-0.8.

### III. MIMO-OFDM and Givens Rotation

For a MIMO-OFDM system, the \( n \)th carrier is usually represented by the system equation
\[ Y_n = H_n X_n + N_n, \]
where \( Y_n, H_n, X_n, \) and \( N_n \) are the received signal, channel matrix, transmitted signal and the observation noise respectively.

By applying singular value decomposition (SVD) on \( H_n \),
\[ Y_n = U_n \Sigma_n V_n^H X_n + N_n, \]
we can observe that the optimal beamforming matrix required at the transmitter is \( V_n \). Therefore, only the feedback of the singular matrix, \( V_n \), is required. This is also the feedback requirement adopted in the IEEE 802.11n draft standard [1].

#### A. Givens Rotation

To feedback \( V_n \), Givens rotation can be used to compress the unitary matrix into two parameters, \( \phi_{11} \) and \( \psi_{21} \) [1].

\( V_n \), and any unitary matrix of dimension \( N_T \times K \), can be represented as [6]
\[ V = \prod_{i=1}^{\min(N_T-1,K)} D_i \left( \begin{array}{cccc} 1 & -1 & \cdots & -e^{j\phi_{1,i-1}} \\ & \ddots & \vdots & \vdots \\ & \cdots & \ddots & \cdots \\ & \cdots & \cdots & 1 \\ \end{array} \right) \times \prod_{i=1}^{N_T} G_{li}^T(\psi_i) \times I_{N_T \times K}, \]
where \( D_i \) is a diagonal matrix and \( G_{li}(\psi_i) \) is defined as
\[
G_{li}(\psi_i) =
\begin{bmatrix}
I_{i-1} & 0 & 0 & 0 \\
0 & \cos(\psi_{li}) & 0 & \sin(\psi_{li}) \\
0 & 0 & I_{i-1} & 0 \\
0 & -\sin(\psi_{li}) & 0 & \cos(\psi_{li}) \\
0 & 0 & 0 & I_{N_T-1}
\end{bmatrix}.
\]
For example, a $3 \times 2$ unitary matrix can be described as

$$
V = \begin{bmatrix}
  e^{j\phi_{11}} & 0 & 0 \\
  0 & e^{j\phi_{21}} & 0 \\
  0 & 0 & 1 \\
\end{bmatrix} \times G_{21}^T(\psi_{21})G_{31}^T(\psi_{31}) \\
\times \begin{bmatrix}
  1 & 0 & 0 \\
  0 & e^{j\phi_{22}} & 0 \\
  0 & 0 & 1 \\
\end{bmatrix} \times G_{32}^T(\psi_{32}) \times \begin{bmatrix}
  1 & 0 \\
  0 & 1 \\
  0 & 0 \\
\end{bmatrix}
$$

(10)

Hence, the $3 \times 2$ unitary matrices $V$ can be fully described by just six parameters: $\phi_{11}, \phi_{21}, \psi_{21}, \psi_{31}, \phi_{22}, \psi_{32}$. Table I lists down the number of parameters for Givens rotation of various dimensions. We shall not go into details on how to solve for the Givens parameters, as the reader can refer to books such as [6] for details.

We would however like to point out that the Givens parameters are angles and hence would satisfy the property

$$
\theta = \theta + k\pi, \quad k \in \mathbb{Z}^+,
$$

(11)

where $\mathbb{Z}^+$ denotes the set of positive integers. Keeping this in mind, and denoting the set of all equivalent angels by $\Theta$, when solving for the Givens parameters, we choose the parameter such that

$$
\theta(n) = \arg \min_{\theta(n) \in \Theta} |\theta(n) - \theta(n - 1)|^2,
$$

(12)

where $n$ is the subcarrier index.

For example, in Figure 1, we show the Givens parameters for a realization of Channel Type ‘E’ [7] using normal Givens rotation and applying (12). The red boxes show the spikes in using normal Givens, which could result in a non-optimal set of differentially quantized values.

By selecting the Givens parameters based on (12), we can apply the differential quantization described in the preceding section, by treating the Givens parameters, $\phi(n)$ and $\psi(n)$, $n = 1 \ldots N$, as correlated sequences.

IV. FEEDBACK RATE

The output of the encoding algorithm is the first sample quantized at $A$ bits followed by a string of $N - 1$ differences quantized at $B$ bits. If we assume that the dynamic range of the quantization and the parameter $\alpha$ is known a priori, then the total feedback requirement for the scheme is $A + (N - 1)B$ bits, as opposed to $AN$ bits for quantizing all the channel information at $A$ bits. If $A \gg B$ and $N \gg 1$, which is normally the case, it could be seen that differential encoding offers an enormous advantage over normal quantization schemes.

V. SIMULATION RESULTS

In the simulations, we use the IEEE 802.11n channel models described in [7]. We use two of the channel types, namely ‘B’ and ‘E’. The channel types have different maximum delays, with Type ‘B’ model having short delay (highly correlated in frequency) and Type ‘E’ model having extremely long delay (less correlated in frequency). Without the loss of generalization, we assume a two transmit, two receive MIMO system. Extensions to larger MIMO systems are straightforward. We also assume that the system bandwidth is 40MHz, and the transmit and received filters used are root raised cosine filters with rolloff factor of 0.25. The number of subcarriers, or Fast Fourier Transform (FFT) size is 128.

To measure the accuracy of the quantized CSI, we use the mean squared error (MSE), normalized to the power of the parameter, defined as

$$
MSE = \frac{1}{N} \sum_{i=1}^{N} \frac{||\phi_i - \hat{\phi_i}||^2}{E(||\phi||^2)}.
$$

(13)

Each simulation point for the average MSEs is obtained from at least 10,000 channel realizations.

In Figure 2, we show a realization of Channel Type ‘B’. Differential quantization of the channel parameter, $\phi_{11}$, using 1 and 2 bits were also shown. As Channel Type ‘B’ is highly correlated in frequency, 1 bit differential quantization can already represent the channel parameter quite well. With 2 bits, we can observe that the quantization error is extremely small.

We present a realization of Channel Type ‘E’ in Figure 3. As pointed out earlier, Channel Type ‘E’ is much less correlated than Type ‘B’. Using 1 bit differential quantization results in a worse approximation as compared to Type ‘B’ (Please note the different scales of the two diagram), suggesting that a larger number quantization bits are required to sufficiently represent the channel parameters.

Next, we study the impact of varying the $\alpha$ parameter. In Figure 4, we have the average MSEs for the parameter $\phi_{11}$ ($\psi_{21}$ results in similar diagrams, but we omit these due to space constraints). For both Channel Type ‘B’ and ‘E’, we can observe that the optimal value of $\alpha$ increase with the number of quantization bits, $b_{Q}$, used. As we recall from Section II-A, the $\alpha$ parameter represents the tradeoff between slope overload distortion and granular noise, with $\alpha = 1$ having no slope.
overload distortion but large granular noise. As $b_Q$ increases, the granular noise possible in the representation is lowered, and the limiting factor in the performance is the slope overload distortion. Therefore, the optimum $\alpha$ value increases with $b_Q$, up to $b_Q \geq 4$, where it is always optimum to use $\alpha = 1$. We can also observe from the diagrams that for smaller values of $b_Q$ (1 or 2 bits), the optimal range of $\alpha$ value to use would be 0.7-0.8.

Finally, we look at the impact of $b_Q$. In Figure 5, we have the averages MSEs for the parameter $\psi_{21}$. We vary $b_Q$ while fixing $\alpha$. We can observe that for each $\alpha \neq 1$, there is an error floor as $b_Q$ is increased. This confirms our earlier conclusion that it is prudent to use $\alpha = 1$ when $b_Q$ is large. The error floor is caused by the slope overload distortion which does not decrease with the increase in $b_Q$ due to $\alpha$.

VI. CONCLUSION

In this paper, we proposed using differential multilevel quantization to efficiently feedback channel parameters in MIMO-OFDM systems. By exploiting the frequency correlation, the amount of feedback can be significantly reduced.

We have also shown methods to design the differential quantizer and the Givens rotation so as to reduce quantization errors in the system. Simulations results have verified that the methods can produce very accurate CSI feedback while using a largely reduced number of bits.

ACKNOWLEDGEMENTS

The authors would like to thank Y. Wu for the helpful discussion on the IEEE 802.11n channel models.

REFERENCES


Fig. 4. Average MSE with different $\alpha$ and quantization bits, $b_Q$, for $\phi_{11}$ in (a) Channel Type ‘B’ (b) Channel Type ‘E’.

Fig. 5. Average MSE with different $\alpha$ and quantization bits, $b_Q$, for $\psi_{21}$ in (a) Channel Type ‘B’ (b) Channel Type ‘E’.