

# Peak to Average Power Ratio Reduction for Multicarrier Systems Using Dirty Paper Coding

Pin-Hsun Lin, Shih-Chun Lin, Hsuan-Tien Liu, Hsuan-Jung Su

Department of Electrical Engineering, and Graduate Institute of Communication Engineering,  
National Taiwan University, Taiwan, R. O. C.

Email: {f89921145, f89921044, r92942098}@ntu.edu.tw, hjsu@cc.ee.ntu.edu.tw

**Abstract**—In this paper, we improve the peak to average power ratio reduction scheme for multicarrier systems proposed by Collings and Clarkson by applying dirty paper coding with peak power constraint. We compare the bit error rate performance among conventional orthogonal frequency division multiplexing, Collings and Clarkson's method, and our proposed scheme with bit loading. From simulation we find that when channel coding is considered, Collings and Clarkson's method is the worst. The proposed method performs the best when the number of bits is large and hence is suitable for high speed transmission.

**Index Terms**—peak to average power ratio, orthogonal frequency division multiplexing, discrete multi-tone, dirty paper coding, Tomlinson Harashima precoding, trellis precoding

## I. INTRODUCTION

Due to the ability to cope with the frequency selective fading of wideband communication with reasonable complexity, the sinusoidal based multicarrier modulation (MCM) has been widely used. The two most well known variants of MCM are the orthogonal frequency division multiplexing (OFDM) for the wireless local area network and the discrete multitone for the digital subscriber loop. However, the high time domain peak to average power ratio (PAPR) of MCM has been a major obstacle in the implementation of power efficient transmitters. Many techniques have been proposed to alleviate the power efficiency problem. The most straightforward method is to clip the peaks which exceed some predefined thresholds in the time-domain [1]. The clipped signals must then be filtered to suppress the out-of-band signal. This filtering, however, tends to re-generate peaks in the time-domain. Other PAPR reduction methods are mostly redundancy based. These include selected mapping [2], partial transmit sequence [3], tone reservation [4], and tone rejection [5]. Selected mapping requires an optimization which may not be computationally feasible if the size of OFDM blocks is large. The approach with partial transmit sequence needs to transmit side information to

let the receiver know the rotating vector. Moreover, its complexity grows exponentially with the number of sub-blocks and the log size of the alphabets of rotating vectors. The tone reservation method needs to reserve tones to achieve PAPR reduction. So the transmission is at a lower rate. On the other hand, its complexity is quite low and is linearly proportional to the size of the OFDM symbol. The tone injection approach has a larger transmission power because it uses a larger constellation sizes to represent the same information. In addition, the complexity for optimization is prohibitive for implementation in practical systems.

Collings and Clarkson [6] recently proposed a non-sinusoidal based low PAPR transmission scheme. This idea is implemented by applying the QR decomposition to the channel matrix, which is different from the sinusoidal based MCM, where the eigenvalue decomposition is used. In addition, they employed a block Tomlinson Harashima precoding (THP) [7] [8] to remove the interference caused by other subchannels. The modulo size of the THP was set to be equal to the linear range of the PA. Therefore clipping of the transmitted signal by the PA is avoided. This method is, however, only suitable for high signal-to-noise ratio (SNR) where the additive noise does not alter the outcome of the receiver modulo operation much. Moreover, channel coding was not considered in their work. The most important performance index in modern communication systems is the coded bit error rate (BER). Thus channel coding should be taken into account to assess the performance degradation caused by the clipping effect.

In this paper, we use the dirty paper coding (DPC) [9] with peak power constraint to replace the THP used in [6]. The information-theoretic results revealed in [9] claim that with average power constraint, if we know the interference sequence (side information) noncausally at the transmitter, the interference can be removed completely even if the receiver has no information of it. The DPC has several advantages over the THP, such as avoiding the shaping loss, power loss, and modulo loss [10] which will be explained in Section III. The practical code design issues of DPC can be found in [11]–[13]. To gain these advantages, we first show that the vector transmission problem with PA constraint can be recast as several parallel scalar transmission problems with the same PA constraint and noncausal side information at the

This paper is based on "Peak to Average Power Ratio Reduction for Multicarrier Systems Using Dirty Paper Coding," by P. H. Lin, S. C. Lin, H. T. Liou, and H. J. Su, which appeared in the Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toulouse, France, May 2006. © 2006 IEEE.

This work was supported in part by the Excellent Research Projects of National Taiwan University, 95R0062-AE00-04, and the National Science Council, Taiwan, R.O.C., under grant NSC 94-2219-E-002-005.

transmitter. Although the original DPC problem focuses on average power constraint, we can still use the same concept with additional peak power constraint. The coded BER performances of conventional OFDM, Collings and Clarkson's method, and the proposed one with bit loading are performed by simulation. When channel coding is considered, Collings and Clarkson's method is the worst. The proposed method performs the best when the number of bits is large and hence is suitable for high speed transmission.

This paper is organized as follows. In Section II we introduce the system model and explain how to obtain the parallel scalar transmission problems. In Section III we review the development of several precoding schemes including DPC. After that, the general code design concept is presented in Section IV. Section V provides a practical code design. Simulations and discussions are given in Section VI. Finally, Section VII concludes this paper.

## II. SYSTEM MODEL AND PROBLEM FORMULATIONS

The following notations will be used throughout the paper. Vectors and matrices are denoted by lowercase boldface and uppercase boldface alphabets respectively. The identity matrix is denoted by  $\mathbf{I}$ . Subscripts indicate the dimension of the vector or matrix.  $(\cdot)^H$  means transpose and complex conjugate. Random variables are denoted by uppercase alphabet. As in the common transmission model used in OFDM systems, we assume that the guard interval is appended to each transmitted block and removed at the receiver. The guard interval is assumed to be long enough to avoid the inter-block interference. Without loss of generality, we assume the linear range of PA is within  $[-1/2, 1/2]$ . To avoid the clipping effect of PA, we consider the following vector transmission

$$\begin{aligned} \mathbf{y}_m &= \mathbf{H}\mathbf{x}_m + \mathbf{n}_m, \\ \text{peak power constraint: } |x_{m,i}|^2 &\leq 1/4, \\ \text{average power constraint: } E[|\mathbf{x}_m|^2] &\leq NP_x \end{aligned} \quad (1)$$

where  $\mathbf{x}_m \in \mathbb{C}^{N \times 1}$  is the  $m$ th transmitted block with the  $i$ th entry denoted by  $x_{m,i}$ ;  $\mathbf{y}_m \in \mathbb{C}^{N \times 1}$  is the  $m$ th received block;  $\mathbf{H} \in \mathbb{C}^{N \times N}$  is the channel matrix which is assumed to be constant;  $\mathbf{n}_m \in \mathbb{C}^{N \times 1}$  is the complex additive white Gaussian noise vector at the receiver with zero mean and variance  $\sigma^2$  for all entries;  $P_x$  is the average power constraint per sample at the transmitter. We will show that the vector transmission problem in (1) can be turned into  $N$  parallel scalar transmission problems, each having peak power constraint  $1/4$  and noncausal side information at the transmitter.

In the following we summarize Collings and Clarkson's model. Then we relate this model to the DPC problem with peak power constraint. Assume the channel model is known at the transmitter. To form the parallel scalar problems, Collings and Clarkson first applied the QR decomposition to  $\mathbf{H}$ ,

$$\mathbf{H} = \mathbf{Q}\mathbf{R} = \mathbf{Q}\mathbf{D}\mathbf{U}$$

where  $\mathbf{Q}$  is a unitary matrix and  $\mathbf{D} = \text{diag}(\mathbf{R})$  such that the diagonal entries of the upper triangular matrix  $\mathbf{U}$  are all unity. With  $\tilde{\mathbf{x}}_m \triangleq \mathbf{U}\mathbf{x}_m$  and  $\tilde{\mathbf{y}}_m \triangleq \mathbf{D}^{-1}\mathbf{Q}^H\mathbf{y}_m$ , we then have the  $N$  parallel channels with the  $i$ th channel being

$$\tilde{y}_{m,i} = \tilde{x}_{m,i} + \tilde{n}_{m,i}, \quad 1 \leq i \leq N \quad (2)$$

where  $\tilde{n}_{m,i}$  is the  $i$ th element of  $\mathbf{D}^{-1}\mathbf{Q}^H\mathbf{n}_m$ , which is complex Gaussian with diagonal covariance matrix. With the fact that  $\mathbf{x}_m = (\mathbf{I} - \mathbf{U})\mathbf{x}_m + \tilde{\mathbf{x}}_m$  and  $(\mathbf{I} - \mathbf{U})$  is strictly upper triangular, we have

$$\begin{aligned} x_{m,i} &= \tilde{x}_{m,i} - \sum_{k=i+1}^N x_{m,i} \cdot u_{i,k} \\ &\triangleq \tilde{x}_{m,i} - s_{m,i} \end{aligned} \quad (3)$$

where  $u_{i,k}$  is the entry of  $\mathbf{U}$  in the  $i$ th row and the  $k$ th column. Note that  $x_{m,N} = \tilde{x}_{m,N}$ . In the following, we illustrate how to formulate (3) as a DPC problem with additional peak power constraint. If we encode  $x_{m,N}, \dots, x_{m,i+1}$  before  $x_{m,i}$ , then  $s_{m,i}$  can be treated as the noncausally known interference at the transmitter. With (2) and (3), we can turn the original problem (1) to the following  $N$  subchannels

$$\begin{aligned} \tilde{y}_{m,i} &= x_{m,i} + s_{m,i} + \tilde{n}_{m,i}, \\ \text{peak power constraint: } |x_{m,i}|^2 &\leq 1/4, \\ \text{average power constraint: } E[|x_{m,i}|^2] &\leq P_x \end{aligned} \quad (4)$$

where  $i = 1, \dots, N$ . The variance of  $\tilde{n}_{m,i}$  is  $\sigma_i^2 = \sigma^2/|d_i|^2$ , where  $d_i$  is the  $i$ th diagonal term of  $\mathbf{D}$ . Comparing (1) and (4) we can find that the QR decomposition does transform the vector transmission into  $N$  parallel scalar ones with noncausal side information. Note that the channel in (4) is different from the DPC channel defined in [9] which only specifies the *average power constraint*. However, after a proper modification the code design ideas from the DPC can still be used. This modification will be discussed in Section V. Although  $s_{m,i}$  are correlated with each other, as will be shown in the next section,  $x_{m,i}$  can still be made independent of each other with the aid of a random dither and the modulo operation [14]. When using waterfilling and bit loading to achieve the capacity of parallel channels, the independence between the parallel channels is an important assumption which is violated in Collins and Clarkson's work. This will also be discussed in Section V.

## III. REVIEW OF SEVERAL PRECODING SCHEMES

In this section we briefly introduce the evolution of three important precoding schemes, which are Tomlinson Harashima precoding, trellis precoding, and dirty paper coding. Assume an interference channel which is modeled as

$$Y = X + S + Z$$

where  $S$  is an arbitrarily noncausal interference known at the transmitter and  $Z$  is identically independent distributed Gaussian noise with zero mean and variance  $\sigma_Z^2$ .

Assume the transmit power constraint is  $P_X \triangleq E[X^2]$  and the power of  $S$  is  $P_S$ . Heuristically, we may transmit the presubtracted signal  $X - S$  such that the received signal is  $X + Z$ . But the transmitted power becomes  $P_X + P_S$  which exceeds the power constraint by  $P_S$ . On the other hand, if  $P_X > P_S$ , the encoder may use part of the available power to cancel  $S$ . However, the maximum transmission rate is only at  $\frac{1}{2} \log(1 + \frac{P_X - P_S}{P_Z})$ .

*A. The Tomlinson Harashima Precoding*

To mitigate the power enhancement problem described above, Tomlinson [7] and Harashima [8] used the scalar modulo at the transmitter

$$Y = [X - S] \text{ mod } A + S + Z$$

where the modulo interval is  $(-A/2, A/2]$ . In short, only the quantization error is transmitted. As a result, the dynamic range of the transmitted signal is within the range of  $-A/2$  to  $A/2$  and the problem of power enhancement can be overcome. The modulo operation folds the approximated Gaussian input and makes the output signal approximately uniformly distributed. See [15] for detailed analysis. At the receiver  $X$  can be recovered by another modulo operation:

$$\begin{aligned} \hat{X} &= Y \text{ mod } A \\ &= [X - S + S + Z] \text{ mod } A \\ &= [X + Z] \text{ mod } A \end{aligned} \tag{5}$$

where (5) follows from the distributivity of the modulo operation. Although the problem of power enhancement is solved, there are still three main losses [10]:

- **Shaping loss:** To achieve the capacity, the transmitted sequence should be chosen uniformly in high dimensional spaces and bounded by a hypersphere. The density induced on the constituent constellation is Gaussian, and here it is uniformly distributed.
- **Power loss:** Since the transmitted signal is uniformly distributed in  $(-A/2, A/2]$  and  $A/2$  is larger than the magnitude of the largest alphabet of  $X$ , the transmit power of THP is larger than that without precoding.
- **Modulo loss:** The original decision boundaries of the outmost constellation points are half planes. After modulo operation these decision boundaries diminish. When the SNR is low, these outmost points are prone to be decided as those at the opposite corners of the constellation which incurs performance loss.

*B. The Trellis Precoding*

Trellis precoding [16], in practical, is the precoding method most similar to DPC. Trellis precoding jointly considers THP and signal shaping such that the shaping loss is overcome. The power loss is also dramatically reduced. This is because the distribution of the trellis

precoded signal is approximately Gaussian and the probability of signals falling out of the support of the original constellation caused by precoding is much smaller than that of the uniform one. But trellis precoding still suffers from the modulo loss.

*C. The Dirty Paper Coding*

Costa [9] proved that with proper encoding the capacity of channels with interferences known at the transmitter equals that of an interference free Gaussian channel. He applied the concept of *linear assignment*<sup>1</sup>

$$X = U + \alpha S \tag{6}$$

to Gel'fand and Pinsker's formula [17]

$$C = \max_{p(u,x|s)} I(U; Y) - I(U; S)$$

where  $U$  is a finite alphabet auxiliary random variable chosen from a codebook<sup>2</sup>. Costa proved that if  $\alpha$  in (6) is chosen as the minimum mean-square-error (MMSE) estimator for the real random variable  $X$  given the received signal  $Y = X + Z$ , i.e.

$$\alpha = \frac{P_X}{P_X + \sigma_Z^2}, \tag{7}$$

the AWGN channel capacity can be achieved. The reason is that this additional scaling factor can eliminate the modulo loss. Note that this real channel is the equivalent real interference-free channel of (4), i.e.,  $s_{m,i} = 0$ . The practical coding scheme of DPC had been unknown until Erez, Shamai, and Zamir [18] proposed that the nested lattice can be used. In the nested lattice scheme, the channel and the shaping codewords are selected from the fine and coarse lattices respectively. After that, Bennatan *et al.* [13] proposed superposition coding scheme to implement the DPC. Superposition coding relaxes the restriction in the nested lattice scheme that the shaping codeword must be nested in the channel codeword. Besides, the superposition coding scheme also relaxes the requirement of linear codes used in the construction A to build the lattice. The codes used in the superposition coding can be nonlinear. In the following we give a brief review of the superposition coding which plays an important role in our work.

IV. THE SUPERPOSITION CODING SCHEME

Assume that the subchannels in (4) are separately encoded. A codeword of each subchannels extends  $L$  blocks and  $L$  must be a few times smaller than the channel coherence time to make  $\mathbf{H}$  constant within a codeword transmission. Here a block contains  $N$  symbols. For the  $i$ th channel, we form an equivalent real vector  $\mathbf{x}_i = (\tilde{\mathbf{x}}_{1,i}, \dots, \tilde{\mathbf{x}}_{L,i})^T$ , where  $\tilde{\mathbf{x}}_{m,i} = [\text{Re}\{x_{m,i}\}, \text{Im}\{x_{m,i}\}]$ . The side-information vector  $\mathbf{s}_i$ , the additive noise  $\mathbf{n}_i$ , and the received vector

<sup>1</sup>This term was coined by Bennatan [13].

<sup>2</sup>Readers who are interested in the method of random coding selection for  $U$ , please refer to [17].

$y_i$  are obtained similarly. We define the following two operations for a real vector  $\mathbf{g}$  with length  $2L$ . In addition, the *Crypto lemma* which is important in the following derivation is also restated.

*Definition 1:* The mod  $A$  operation on vector  $\mathbf{g}$ , denoted by  $\mathbf{g}' \triangleq \mathbf{g} \bmod A$ , is defined componentwisely for each element of  $\mathbf{g}$ , such that  $g'_i = g_i - Q_A(g_i)$ ,  $\forall i$ , where  $Q_A(g_i)$  is the nearest multiple of  $A$  to  $g_i$ .

*Definition 2:* The modulo operation associated with a quantization codebook  $C_q$  generated by a code (linear or nonlinear), is defined as

$$\lfloor \mathbf{g} \rfloor_{C_q} = (\mathbf{g} - Q_{C_q}(\mathbf{g})) \bmod A$$

where the quantizer  $Q_{C_q}(\mathbf{g})$  associated with  $C_q$  is defined by:

$$Q_{C_q}(\mathbf{g}) \triangleq \{ \mathbf{c} \in C_q : \mathbf{g}' = (\mathbf{g} - \mathbf{c}) \bmod A, \|\mathbf{g}'\|^2 \leq \|(\mathbf{g} - \mathbf{c}') \bmod A\|^2, \forall \mathbf{c}' \in C_q, \mathbf{c}' \neq \mathbf{c} \}$$

where  $\|\cdot\|$  denotes the Euclidean norm.

*Lemma 1.* (Crypto lemma, [19])

Let  $G$  be a compact Abelian group with group operation  $+$ , and let  $Y = X + N$ , where  $X$  and  $N$  are random variables over  $G$ .  $N$  is independent of  $X$  and uniform over  $G$ . Then  $Y$  is independent of  $X$  and uniform over  $G$ .

In the following the encoder, decoder, and some characteristics of the superposition coding scheme are discussed.

**Encoder:** The encoder selects a codeword  $\mathbf{c}_i^c$  according to the source message and transmits the vector

$$\mathbf{x}_i = \lfloor (\mathbf{c}_i^c - \alpha_i \mathbf{s}_i - \mathbf{u}_i) \bmod A \rfloor_{C_q} \quad (8)$$

where  $\mathbf{u}_i$ , uniformly distributed in the  $2L$ -dimensional cube  $[-A/2, A/2]^{2L}$  and independent of the channel and interference, is a dither signal known to both the transmitter and the receiver. The scaling factor is  $\alpha_i = P_{x_i} / (P_{x_i} + \sigma_{\tilde{n}_i}^2)$  where  $\sigma_{\tilde{n}_i}^2$  is the variance of the additive noise in (4).

In (8), if  $\mathbf{x}$  is white, scalar  $\alpha$  suffices. In vector quantization theory, white  $\mathbf{x}$  can be generated when  $Q_{C_q}$  is a good vector quantizer [20]. The encoder in (8) can be derived from the nested lattice [21]. We give a proof as following.

*Proof:* Here we want to derive the connection between (8) and the modulo operation with respect to a construction A lattice, that is

$$\mathbf{v} \bmod \Lambda_q = \lfloor \mathbf{v} \bmod A \rfloor_{C_q} \quad (9)$$

where  $\mathbf{v} \in \mathbb{R}^n$  is the signal to be shaped and  $\Lambda_q$  is a lattice in the Euclidean space  $\mathbb{R}^n$  which is constructed by

the construction A [22]

$$\Lambda_q = C_q + AZ^n$$

where  $C_q$  is a linear code over  $GF(A)$  and  $A$  is a prime number.

We begin the derivation from the definition of the modulo operation with respect to the lattice  $\Lambda_q$ , that is

$$\mathbf{v} \bmod \Lambda_q = \mathbf{v} - \mathbf{c} - \mathbf{r}$$

where

$$\{\mathbf{c}, \mathbf{r}\} = \arg \min_{\substack{\mathbf{c}' \in C_q \\ \mathbf{r}' \in AZ^n}} \|\mathbf{v} - \mathbf{c}' - \mathbf{r}'\|^2.$$

Actually

$$\min_{\substack{\mathbf{c}' \in C_q \\ \mathbf{r}' \in AZ^n}} \|\mathbf{v} - \mathbf{c}' - \mathbf{r}'\|^2 = \min_{\mathbf{c}' \in C_q} \|(\mathbf{v} - \mathbf{c}') \bmod A\|^2 \quad (10)$$

$$= \min_{\mathbf{c}' \in C_q} \|(\mathbf{v} \bmod A - \mathbf{c}') \bmod A\|^2 \quad (11)$$

$$= \lfloor \mathbf{v} \bmod A \rfloor_{C_q}. \quad (12)$$

The equality of (10) is simply from Definition 1. The equality of (11) is from the distributive property of mod  $A$  operation. (12) is from Definition 2. ■

From the above derivation, we can find that if the constructing code  $C_q$  is linear, the original modulo operation with respect to  $\Lambda_q$  can be simplified as to find a codeword in  $C_q$  with two additional scalar quantizations, i.e. mod  $A$ , as shown in (9). In fact, the DPC transmission model derived from nested lattice theory [21] can be seen as a special case of (8) when  $C_q$  is linear. However, once  $C_q$  is nonlinear, (8) may not be explained in this way since  $\Lambda_q$  is not a lattice.

This encoder in (8) can be further written as

$$\mathbf{x}_i = (\mathbf{c}_i^c + \mathbf{c}_i^q - \alpha_i \mathbf{s}_i - \mathbf{u}_i) \bmod A \quad (13)$$

where  $\mathbf{c}_i^q = -Q_{C_q}(\mathbf{c}_i^c - \alpha_i \mathbf{s}_i - \mathbf{u}_i) \bmod A$ . Using the distributivity of modulo operation again, (13) becomes

$$\begin{aligned} \mathbf{x}_i &= (\mathbf{c}_i^c + \mathbf{c}_i^q - (\alpha_i \mathbf{s}_i + \mathbf{u}_i) \bmod A) \bmod A \\ &= (\mathbf{c}_i^c + \mathbf{c}_i^q - d') \bmod A \end{aligned} \quad (14)$$

where the equality in (14) follows from the Crypto lemma and  $d'$  is uniformly distributed in  $[-\frac{A}{2}, \frac{A}{2}]^{2L}$ . It is clear that  $\mathbf{x}_i$  is independent of  $\mathbf{s}_i$ . Thus  $\mathbf{x}_i$  is independent of  $\mathbf{x}_k$  for all  $k \neq i$  and all subchannels are independent. Note that  $\mathbf{c}_i^q$  depends on  $\mathbf{u}_i$  so we can not use the Crypto lemma to claim that  $\mathbf{x}_i$  is uniformly distributed in  $[-\frac{A}{2}, \frac{A}{2}]^{2L}$ . In fact,  $\mathbf{x}_i$  is truncated Gaussian. As  $A$  approaches infinity,  $\mathbf{x}_i$  becomes Gaussian.

In order to minimize the clipping probability, we can impose an additional constraint on the quantizer  $Q_{C_q}$  such that  $x_{2i-1}^2 + x_{2i}^2 \leq \frac{1}{4}$ ,  $1 \leq i \leq L$ . If we properly design the quantization codebook, the power of  $\mathbf{x}_i$  can be minimized and the power constraint of PA is satisfied.

The mod  $A$  operation also serves to limit the range of  $C_q$  to make the encoding and decoding easier to implement.

**Decoder:** After passing  $\mathbf{x}_i$  through the channel in (4), the decoder first performs some processing on the received signal to get

$$\begin{aligned} \hat{\mathbf{y}}_i &= (\alpha_i \mathbf{y}_i + \mathbf{u}_i) \bmod A \\ &= (\mathbf{c}_i^c + \mathbf{c}_i^q + \mathbf{e}_i) \bmod A \end{aligned} \quad (15)$$

where  $\mathbf{e}_i = -(1 - \alpha_i)\mathbf{x}_i + \alpha_i\tilde{\mathbf{n}}_i$ . (15) comes from the distributive property of the mod  $A$  operation and (13).  $\hat{\mathbf{y}}_i$  is then used to obtain an estimation  $\hat{\mathbf{c}}_i^c$  of  $\mathbf{c}_i^c$ . Due to the effect of the dither,  $\mathbf{e}_i$ ,  $\mathbf{c}_i^q$  and  $\mathbf{c}_i^c$  are independent of one another.

### V. PRACTICAL CODE DESIGN

The overall transmission model is depicted in Fig. 1. Fig. 2 illustrates the transmission of each subcarrier where data over the bold and fine connections are in finite and real field respectively. The quantizer codebook consists of a convolutional code followed by a constellation mapping. In the practical DPC design, we use the sign bit shaping [23] to generate the quantization codebook. The quantization process in Definition 2 is performed by the Viterbi algorithm (V.A.). It searches codewords  $\mathbf{c}_i^q$  with the following two properties. The first is to make the quantization error satisfy the peak power constraint. And the second is to minimize the energy of the quantization error  $\mathbf{x}_i$ . There is a tradeoff between the peak and average power constraints [16]. Thus independent selection of these two constraints may result in no extendable branch in the trellis. In our case we only set the peak value and find the codeword with minimum average power among those satisfying the peak constraint. The output  $\mathbf{c}_i^q$  is used to change the sign bits of  $\mathbf{c}_i^c$  before the constellation mapping.

At the decoder side, we have the equivalent channel (15). To decode  $\mathbf{c}_i^c$ , we choose  $\mathbf{H}_s^T$  in Fig. 2 as the parity check matrix of the quantizer. Then we can do hard decision on the sign bits and eliminate  $\mathbf{c}_i^q$ . However, due to the noise  $\mathbf{e}_i$  and mod  $A$  operation, the performance may not be good enough. Since the convolutional code is used in our system, we can combine the channel code and quantizer trellises to jointly decode  $\mathbf{c}_i^c + \mathbf{c}_i^q$ . In general,  $\mathbf{e}_i$  in (15) is not Gaussian, but we can still use a suboptimal decoder which finds  $\hat{\mathbf{c}}_i^c$  that maximizes the following metric

$$\sum_{\forall \mathbf{c}_i^q} \sum_{\mathbf{z} \in \mathbb{Z}^{2L}} \exp \left[ -\frac{\|\hat{\mathbf{y}}_i - \hat{\mathbf{c}}_i^c - \mathbf{c}_i^q - A\mathbf{z}\|^2}{2\sigma_{\mathbf{e}_i}^2} \right] \quad (16)$$

where  $\sigma_{\mathbf{e}_i}^2$  is the variance of  $\mathbf{e}_i$ . Rigorous analysis of the optimality of the decoding metric can be found in [14]. In practice, only a few neighboring mod  $A$  intervals need to be considered. Thus the summation over all cubic lattice points can be avoided.

Since the noise variance is different in each subchannel due to different subchannel gains, bit loading is necessary

to make the system achieve the capacity. To use the bitloading or waterfilling for parallel Gaussian channels, white transmitted signals and noise are necessary. With the aid of dirty paper coding, the transmission in (4) can be transformed into

$$\begin{aligned} \tilde{\mathbf{y}}_{m,i} &= x_{m,i} + \tilde{\mathbf{n}}_{m,i}, \\ \text{peak power constraint: } &|x_{m,i}|^2 \leq 1/4, \\ \text{average power constraint: } &E[|x_{m,i}|^2] \leq P_x \end{aligned} \quad (17)$$

where  $i = 1, \dots, N$ . In addition, from the Crypto lemma  $x_{m,i}$  are independent for all  $i$ . Thus (17) contains  $N$  parallel Gaussian noise channel. Therefore, the bitloading can be used to achieve the capacity. However, in Collings and Clarkson's case, the transmitted signal is not white and the use of bitloading incurs performance loss.

To prevent loading noninteger or negative number of bits, the loading formula in [6] is used. Let  $B_k$  denote the number of bits that are loaded on the  $k$ th sub-channel. Let  $M$  denote the total number of bits to be transmitted. The  $B_k$  bits are Gray mapped onto a QAM constellation which has  $2^{B_k}$  points evenly distributed in  $[-\frac{1}{2}, \frac{1}{2}]^2$ . The bit loading procedure is summarized as follows:

**Step 1.** Set  $B_p = 0$ .

**Step 2.** Calculate the following three terms

$$\beta_k \triangleq \log_2 d_k + \frac{1}{N} \left( \frac{M}{2} - \sum_{p=1}^N \log_2 d_p \right), \quad (18)$$

$$S \triangleq \sum_{p=1}^N \lceil \beta_p \rceil - \frac{M}{2}, \quad (19)$$

$$\delta_k \triangleq \lceil \beta_k \rceil - \beta_k. \quad (20)$$

**Step 3.** Assign  $B_k = 2(\lceil \beta_k \rceil - 1)$  bits to the subchannels with the  $S$  largest values of  $\delta_k$  and  $B_k = 2\lceil \beta_k \rceil$  to the remainder.

### VI. SIMULATIONS AND DISCUSSIONS

In the following, we compare the coded BER performances of OFDM, QR with THP [6] and QR with DPC by computer simulations. The number of loaded bits per block is set as 250. Note that this is a relatively high loading situation. Most PAPR reduction papers only consider uncoded case but channel coding is inevitable in modern high rate transmission. Thus we compare both coded and uncoded performances. In our simulation the HIPERLAN/2 BRAN A channel model [24] for typical office environments (non line of sight) is used. Several important characteristics of the channel are introduced as follows. The channel is modeled by a finite impulse response filter with 18 taps. The root mean square delay spread of the channel is 50 ns and the maximum delay is 390 ns. The number of subchannels is set to 64. The power delay profile of the channel is listed in Table I.

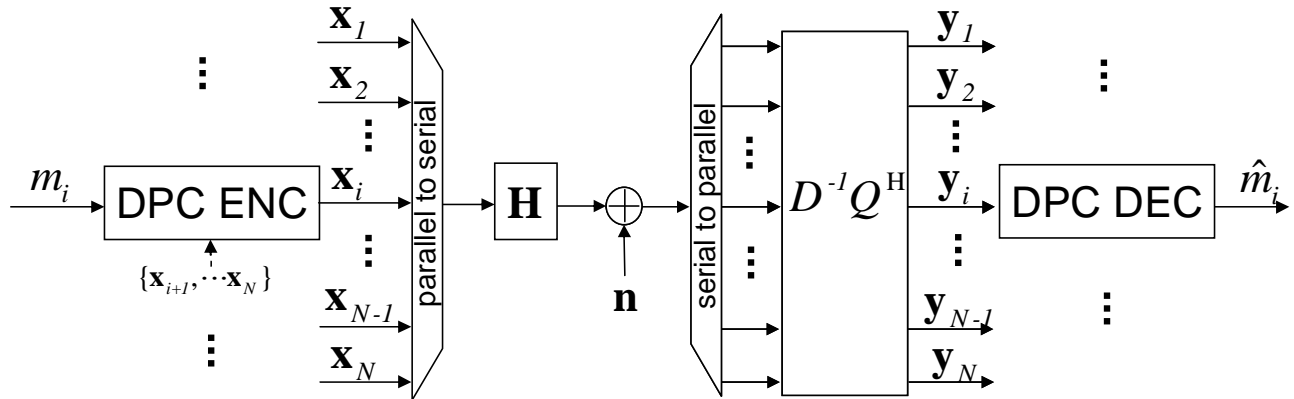


Figure 1. The proposed transmission scheme.

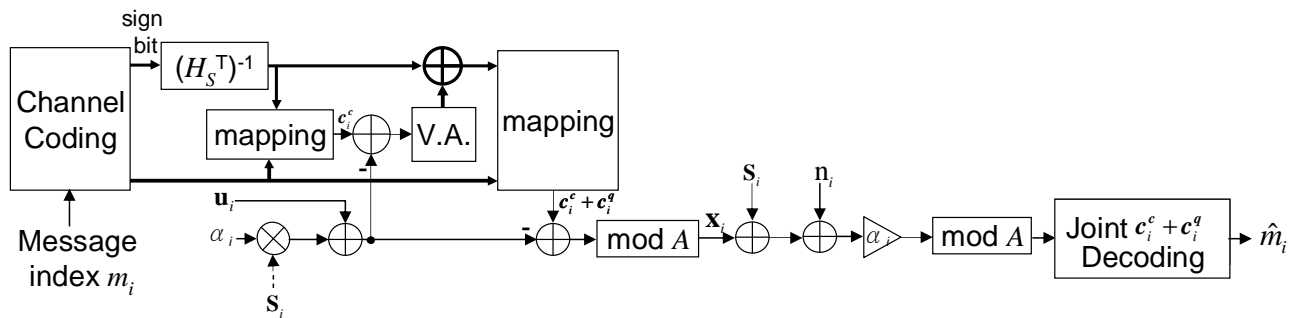


Figure 2. The transmission model of each subcarrier.

Each codeword of the shaping code covers  $L=32$  blocks in both cases. The modulo size  $A$  is set to 1 for our proposed method. Nonsystematic convolutional codes are used for both the shaping code and channel code. The generator polynomials are listed in Table II.

When channel coding is not considered, the transmission model can be obtained by removing the channel coding block in Fig. 2. Fig. 3 shows the simulation result without channel coding. Since the considered SNR range is high, the effect of noise is small. The method in [6] can prevent clipping by just scaling the signals into the linear range with very little performance degradation. Note that the clip-to-average ratio is set to 5 dB in [6], which is a small value. As the loading is high, the OFDM signals are prone to be clipped. Under such an unfair condition, OFDM is obviously worse than the method in [6]. The gain of the proposed scheme is obvious from Fig. 3. Fig. 4 compares the performance when channel coding is considered. All three systems use soft decoding. From Fig. 3 and Fig. 4, we can find that the three curves become much closer to one another in coded case. Intuitively, most errors caused by clipping can be recovered by the channel coding. In addition, OFDM performs better than QR decomposition with THP. The proposed method still performs best although we use a suboptimal metric to do the joint decoding. It is easy to verify that the method proposed in [6] is indeed a special case of our approach. In particular, that method transmits signals on the  $i$ th subchannel with  $\alpha_i = 1$ ,  $u_i = 0$  and there is no  $[\cdot]_{C_q}$  in

(8) and  $A$  is set as  $1/\sqrt{2}$ .

Finally we summarize the advantages of the proposed method over the one in [6]:

- From information-theoretic point of view, bit loading operation maximizes the system throughput under the assumption that subchannels are independent. Without the dither  $u_i$ , the transmitted signal  $x_i$  in Collings and Clarkson's model will depend on  $s_i$ . Then all subchannel signals are correlated. Thus bit loading may not achieve the desired capacity. When DPC is applied, due to the dither and modulo operation, these subchannels are independent.
- The MMSE scaling factor  $\alpha_i$  can minimize the variance of additive noise at the receiver. The proposed method suffers less noise power and operates better at low SNR.
- The sign-bit shaping makes the transmitted signal distributed more like Gaussian and has lower power, thus introducing the shaping gain.

## VII. CONCLUSION

In this paper, we use the concept of DPC with peak power constraint to replace the THP in [6]. BER performances for conventional OFDM, Collings and Clarkson's method, and the proposed method with bit loading are obtained through simulation. When channel coding is considered, Collings and Clarkson's method is the worst. The proposed method performs best when the number

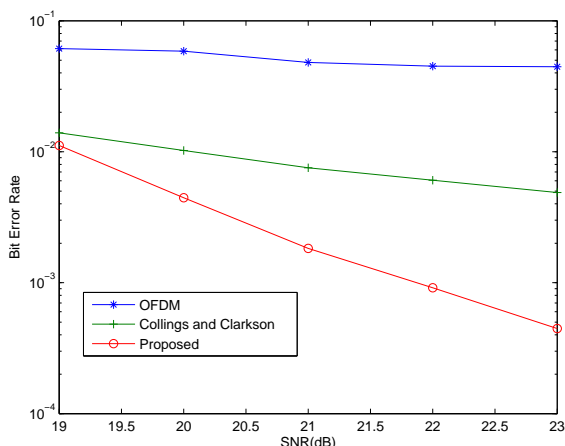


Figure 3. Performance comparison for 250-bit loading, without channel coding.

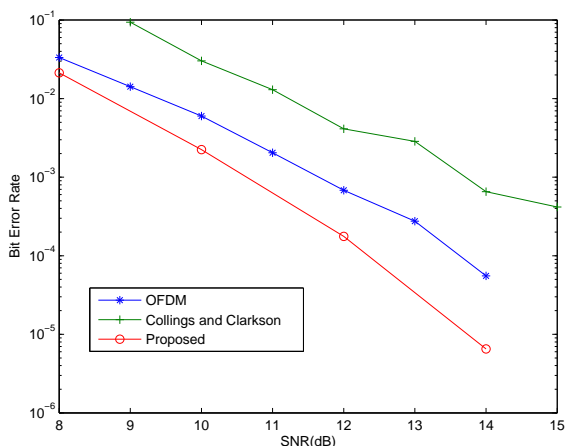


Figure 4. Performance comparison for 250-bit loading, with channel coding.

of bit is large and hence is suitable for high speed transmission.

REFERENCES

[1] X. Li and Jr. L. J. Cimini, "Effects of clipping and filtering on the performance of OFDM," in *Proc. IEEE vehicular technology conference*, May 1997, vol. 3, pp. 1634–1638.  
 [2] R. Fisher, R. Bauml, and J. B. Huber, "Reducing the peak-to-average power ratio of multicarrier modulation by selected mapping," in *Elec. Letters*, Oct 1996, pp. 2056–2057.  
 [3] S. H. Muller and J. B. Huber, "A comparison of peak reduction schemes for OFDM," in *Proc. IEEE GLOBECOM Conf.*, Nov. 1997.  
 [4] J. Tellado and J. M. Cioffi, "Peak power reduction for multicarrier transmission," in *Proc. IEEE GLOBECOM Conf.*, Nov. 1998.  
 [5] A. Narula, F. Kschischang, and V. Eyuboglu, "A new approach to PAR control in DMT systems," in *ITU*, May 1998, vol. Q4/15(NF-083).  
 [6] I. B. Collings and I. V. L. Clarkson, "A low-complexity lattice-based low-PAR transmission scheme for DSL channels," *IEEE Trans. Commun.*, vol. 52, no. 5, pp. 755–763, May 2004.

TABLE I.  
THE POWER DELAY PROFILE OF THE CHANNEL MODEL USED IN THE SIMULATION.

Tap number	Delay [ns]	Relative power [dB]
1	0	0.0
2	10	-0.9
3	20	-1.7
4	30	-2.6
5	40	-3.5
6	50	-4.3
7	60	-5.2
8	70	-6.1
9	80	-6.9
10	90	-7.8
11	110	-4.7
12	140	-7.3
13	170	-9.9
14	200	-12.5
15	240	-13.7
16	290	-18.0
17	340	-22.4
18	390	-26.7

TABLE II.  
THE PARAMETERS OF THE CODES USED IN THE SIMULATION.

	Generator polynomial	Code rate	Number of states
Shaping code	$[7, 5]_8$	1/2	4
Syndrome inverse	$[1, 3]_8$	1/2	4
Channel code	$[71, 65, 57]_8$	1/3	32

[7] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electr. Lett.*, vol. 7, pp. 138–139, Mar. 1971.  
 [8] M. Miyakawa and H. Harashima, "A method of code conversion for a digital communication channel with intersymbol interference," *Trans. Inst. Elec. Comm. Eng. Japan*, vol. 52-A, pp. 272–273, Jun. 1969.  
 [9] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, pp. 439–441, May 1983.  
 [10] S. Shamai and R. Laroia, "The intersymbol interference channel: lower bounds on capacity and channel precoding loss," *IEEE Trans. Inform. Theory*, vol. 42, no. 9, pp. 1388–1404, Sep. 1996.  
 [11] T. Philosof, "Precoding for interference cancellation at low SNR using nested codes," M.S. thesis, Tel-Aviv University, 2003.  
 [12] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Inform. Theory*, vol. 51, no. 10, pp. 3417 – 3432, Oct 2005.  
 [13] A. Bennatan, D. Burshtein, G. Caire, and S. Shamai, "Superposition coding for side-information channels," vol. 52, no. 5, pp. 1872–1889, May 2006.  
 [14] U. Erez and R. Zamir, "Achieving  $\frac{1}{2} \log(1 + \text{SNR})$  on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct 2004.  
 [15] R. F. H. Fischer, *Precoding and signal shaping for digital transmission*, IEEE and Wiley interscience, 2002.  
 [16] M.V. Eyuboglu and Jr. G.D. Forney, "Trellis precoding: combined coding, precoding and shaping for intersymbol interference channels," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 301–314, Mar 1992.

- [17] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of control and information theory*, vol. 9, no. 1, pp. 19–31, 1980.
- [18] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," in *Proc. ISITA*, 2000.
- [19] G. D. Forney, "On the role of MMSE estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener," in *Allerton conference*, Oct 2003, pp. 430–439.
- [20] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct 1998.
- [21] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002.
- [22] J. H. Conway and N. J. A. Sloane, *Sphere packings, lattices, and groups*, New York: Springer-Verlag, 1988.
- [23] G. D. Forney, "Trellis shaping," *IEEE Trans. Inform. Theory*, vol. 38, pp. 281–300, Mar. 1992.
- [24] J. Medbo and P. Schramm, "Channel models for HIPERLAN/2 in different indoor scenarios," *ETSI EP BRAN3ERI085B*, Mar 1998.

coding/modulation, fast Hybrid-ARQ, scheduling, and Radio Link Control protocol for 3G wireless networks. In 2003, Dr. Su joined the Department of Electrical Engineering, National Taiwan University, where he is currently an Associate Professor. His research interests cover coding, modulation, signal processing, power control and synchronization of narrowband and wideband communication systems.

**Pin-Hsun Lin** received his B.S. degree from the Department of Electrical Engineering at the National Taiwan University, Taipei, Taiwan, in 2000. Since 2004, he has been working towards his Ph.D. degree in the Graduate Institute of Communication Engineering at the National Taiwan University. He was with the Graduate Institute of Electronics Engineering at the same university between 2000-2004.

His research interests include signal processing, communication architecture, communication, information theory.

**Shih-Chun Lin** received his B.S. degree from the Department of Electrical Engineering at the National Taiwan University, Taipei, Taiwan, in 2000. Since 2004, he has been working towards his Ph.D. degree in the Graduate Institute of Communication Engineering at the National Taiwan University. He was with the Graduate Institute of Electronics Engineering at the same university between 2000-2004. From Jan. 2007, he has been holding a visiting research appointment at the Department of Electrical Engineering, Ohio state University, Columbus, OH., USA.

His research interests include signal processing, coding, communication, information theory.

**Hsuan-Tien Liu** received his B.S. degree from the Department of Electrical Engineering at the National Chiao Tung University, Taiwan, in 2003, and the M.S. degree in communication engineering from National Taiwan University in 2005. He is now on military service.

**Hsuang-Jung Su** received the B.S. degree in Electronics Engineering from the National Chiao-Tung University, Taiwan, in 1992, and the M.S. and Ph.D. degrees in Electrical Engineering from the University of Maryland, College Park, in 1996 and 1999, respectively.

From 1999 to 2000, he was a Postdoctoral Research Associate with the Institute for Systems Research, University of Maryland. From 2000 to 2003, he was with the Bell Laboratories, Lucent Technologies, Holmdel, New Jersey, where he was involved in the design and performance evaluation of adaptive