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
Graduate Institute of Communication Engineering
National Taiwan University, Taipei, Taiwan, 10617
E-mail:{f89921145, f89921044, r92942098}@ntu.edu.tw, hjsu@cc.ee.ntu.edu.tw

ABSTRACT

In this paper, we improve the peak to average power ratio (PAPR) 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 (BER) performance among conventional orthogonal frequency division multiplexing (OFDM), 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 independent of the number of bits loaded. The proposed method performs the best when the number of bits is large and hence is suitable for high speed transmission.

1. INTRODUCTION

The performance of multicarrier systems is often affected by their high peak to average power ratio (PAPR) when the power amplifier (PA) is not ideal. To mitigate the problem of high PAPR, Collings and Clarkson proposed in [1] using the QR decomposition of the channel matrix instead of the eigenvalue decomposition used in orthogonal frequency division multiplexing (OFDM) systems. They then employed a block Tomlinson Harashima precoding (THP) [2] to remove the interference caused by other subchannels. The modulo size of the THP was set to be less than the linear range of the PA, therefore clipping of the transmitted signal by the PA was avoided. The method in [1], however, only works at 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. Channel coding plays an important role since it can correct the error produced from the clipping effect.

In this paper, we use the dirty paper coding (DPC) [3] with peak power constraint to replace the THP used in [1]. The information-theoretic results revealed in [3] 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 [4]. The practical code design issues can be found in [4, 5, 6]. 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 peak power constraint and noncausal side information at the transmitter. Although the original DPC problem focuses on average power constraint, we can still use the same concept with additional peak power constraint. Bit error rate (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 independent of the number of bits loaded. 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 2 we introduce the system model and explain how to obtain the scalar problems. Next, the general code design concept is presented in Section 3. Section 4 provides a practical code design. Simulations and discussions are given in Section 5. Finally, Section 6 concludes this paper.

2. SYSTEM MODEL AND PROBLEM FORMULATIONS

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 $1/2$. To avoid the clipping effect of PA, we consider the following vector channel

$$\mathbf{y}_m = \mathbf{H}\mathbf{x}_m + \mathbf{n}_m, \;\; |x_{m,i}|^2 \leq 1/4, \tag{1}$$

where $\mathbf{x}_m \in \mathbb{C}^{N\times 1}$ is the mth transmitted block with the ith element denoted by $x_{m,i}$; $\mathbf{y}_m \in \mathbb{C}^{N\times 1}$ is the mth received block; $\mathbf{H} \in \mathbb{C}^{N\times N}$ is the channel matrix which is assumed to be constant, and $\mathbf{n}_m \in \mathbb{C}^{N\times 1}$ is the complex additive white Gaussian noise vector at the receiver with zero mean and

This work was supported in part by the NTU-MediaTek Wireless Research Laboratory, and the National Science Council, Taiwan, R.O.C., under grant NSC 94-2219-E-002-005.

variance σ^2 for all components. 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 Collings and Clrarkson's model is given. Then we relate this model to the DPC problem with peak power constraint. To form the parallel scalar problems, Collings and Clarkson first applied the QR decomposition to $\mathbf{H}$,

$$\mathbf{H} = \mathbf{QR} = \mathbf{QDU},$$

where $\mathbf{D}$ is a diagonal matrix such that $diag(\mathbf{D}) = diag(\mathbf{R})$; $\mathbf{Q}$ is an unitary matrix and the diagonal terms of the upper triangular matrix $\mathbf{U}$ are all unity. With $\tilde{\mathbf{x}}_m \triangleq \mathbf{U}\mathbf{x}_m$ and $\tilde{\mathbf{y}}_m \triangleq \mathbf{D}^{-1}\mathbf{Q}^H\mathbf{y}_m$, where the superscript H denotes complex conjugate and transpose, we have the N parallel channels with the ith channel being

$$\tilde{y}_{m,i} = \tilde{x}_{m,i} + \tilde{n}_{m,i}, \;\; 1 \leq i \leq N \tag{2}$$

where $\tilde{n}_{m,i}$ is the ith 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

$$x_{m,i} = \tilde{x}_{m,i} - \sum_{k=i+1}^{N} x_{m,i} \cdot u_{i,k} \triangleq \tilde{x}_{m,i} - s_{m,i}, \tag{3}$$

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

$$\tilde{y}_{m,i} = x_{m,i} + s_{m,i} + \tilde{n}_{m,i}, \;\; i = N, \ldots, 1, \tag{4}$$

with *peak power constraint* $|x_{m,i}|^2 \leq 1/4$. And the variance of $\tilde{n}_{m,i}$ is $\sigma_i^2 = \sigma^2/|d_i|^2$, where d_i is the ith diagonal term of $\mathbf{D}$. Note that this channel is different from the DPC channel defined in [3] which only specifies the *average power constraint*. However, we can borrow the code design ideas from the DPC. Although $s_{m,i}$ are correlated with each other, as will be shown in the next section, $x_{m,i}$ can be made independent of each other with the aid of random dither and modulo operation [7].

3. THEORETICAL CODING SCHEME

We now introduce the coding scheme which can also be found in [8]. Assume that the subchannels (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 ith channel, we form an equivalent real vector $\mathbf{x}_i = (\check{\mathbf{x}}_{1,i}, \ldots, \check{\mathbf{x}}_{L,i})^{\mathrm{T}}$, where $\check{\mathbf{x}}_{m,i} = [\mathrm{Re}\{x_{m,i}\}, \mathrm{Im}\{x_{m,i}\}]$. The side-information vector $\mathbf{s}_i$, the additive noise $\mathbf{n}_i$, and the received vector $\mathbf{y}_i$ are obtained similarly. We also define the following two operations for a real vector $\mathbf{g}$ with length $2L$.

Definition 1 *The* mod A *operation on vector* $\mathbf{g}$*, denoted by* $\mathbf{g}' \triangleq \mathbf{g} \bmod A$*, is defined componentwise 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* $\mathcal{C}_q$ *generated by a linear code, is defined as*

$$\lfloor \mathbf{g} \rfloor_{\mathcal{C}_q} = (\mathbf{g} - Q_{\mathcal{C}_q}(\mathbf{g})) \bmod A,$$

where the quantizer $Q_{\mathcal{C}_q}(\mathbf{g})$ *associated with* $\mathcal{C}_q$ *is defined by:*

$$Q_{\mathcal{C}_q}(\mathbf{g}) \triangleq \{\mathbf{c} \in \mathcal{C}_q : \mathbf{g}' = (\mathbf{g}-\mathbf{c}) \bmod A, \textit{and}$$
$$\|\mathbf{g}'\|^2 \leq \|(\mathbf{g}-\mathbf{c}') \bmod A\|^2, \; \forall \mathbf{c}' \in \mathcal{C}_q, \; \mathbf{c}' \neq \mathbf{c}\},$$

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

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

$$\mathbf{x}_i = \lfloor (\mathbf{c}_i^c - \alpha_i \mathbf{s}_i - \mathbf{u}_i) \bmod A \rfloor_{\mathcal{C}_q} \tag{5}$$

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. In order to minimize the clipping probability, we can impose an additional constraint on the quantizer $Q_{\mathcal{C}_q}$ such that $x_{2i-1}^2 + x_{2i}^2 \leq \frac{1}{4}$, $1 \leq i \leq L$. Together with the mod A operation, the dither makes $\mathbf{x}_i$ independent of $\mathbf{c}_i^c$ and $\mathbf{s}_i$ [6]. If we properly design the quantization codebook, the power of $\mathbf{x}_i$ can be minimized and the power constraint of PA satisfied. The mod A operation also serves to limit the range of $\mathcal{C}_q$ to make encoding/decoding easier to implement. The role of the scaling factor α_i will be clarified in the next section.

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

$$\begin{aligned} \hat{\mathbf{y}}_i &= (\alpha_i \mathbf{y}_i + \mathbf{u}_i) \bmod A \\ &= (\mathbf{c}_i^c + \mathbf{c}_i^q + \mathbf{e}_i) \bmod A, \end{aligned} \tag{6}$$

where $\mathbf{e}_i = -(1-\alpha_i)\mathbf{x}_i + \alpha_i \tilde{\mathbf{n}}_i$ and $\mathbf{c}_i^q = -Q_{\mathcal{C}_q}(\mathbf{c}_i^c - \alpha_i \mathbf{s}_i - \mathbf{u}_i) \bmod A$. The second equality of (6) comes from the distributive property of the mod-A operation and the fact that

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

$\hat{\mathbf{y}}_i$ 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.

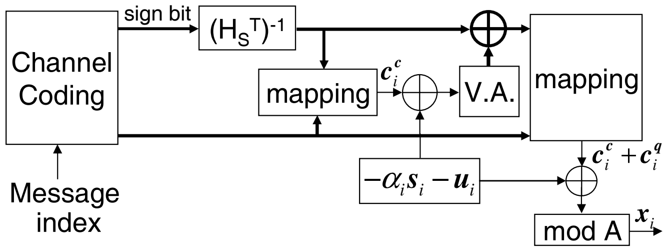

Fig. 1. Block diagrams of the proposed encoder.

4. PRACTICAL CODE DESIGN

Borrowing from the DPC design example [6], we use the sign bit shaping to design the quantization codebook. The overall encoder is depicted in Fig. 1, in which data over the bold connections are in finite field while data over the fine connections are real. The quantizer codebook consists of a convolutional code followed by a constellation mapping. The quantization process in Definition 2 is performed with the aid of the Viterbi algorithm (V.A.). It searches codewords $\mathbf{c}_i^q$ with the following two properties. The first is to minimize the energy of the quantization error $\mathbf{x}_i$. And the second is to make the quantization error (5) satisfy the peak power 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 (6). It is proved in [5] that if we choose α_i as the MMSE estimator for the real random variable x given the received signal $y = x + \tilde{n}_i$, where $\tilde{n}_i$ is white Gaussian with variance $\sigma_i^2/2$, the power of $\mathbf{e}_i$ will be minimized. If x is white, scalar α_i suffices. In theory, white x can be generated when Q_{C_q} is a good vector quantizer [9]. Note that this real channel is the equivalent real interference-free channel of (4), i.e., $\mathbf{s}_{m,i} = 0$. The MMSE scaling indeed increases the effective SNR. To decode $\mathbf{c}_i^c$, we choose $\mathbf{H}_s^T$ in Fig. 1 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.

If the channel code is a convolutional code, 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 (6) is not Gaussian, but we can still use a suboptimal decoder which finds $\hat{\mathbf{c}}_i^c$ that maximizes the following metric

$$\sum_{\forall \mathbf{c}_i^q} \sum_{\mathbf{z} \in \mathbb{Z}^{2L}} \exp\left[-\frac{||\hat{\mathbf{y}}_i - \hat{\mathbf{c}}_i^c - \mathbf{c}_i^q - A\mathbf{z}||^2}{2\sigma_{\mathbf{e}_i}^2} \right], \qquad (7)$$

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 [7]. In practice, only a few neighboring mod A intervals need to be considered. Thus 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 prevent loading noninteger or negative number of bits, the loading formula [1] is used here. Let B_k denote the number of bits that are loaded on the kth sub-channel. The B_k bits are Gray mapped onto a QAM constellation which has 2^{B_k} points evenly distributed in $\left[\frac{-1}{2}, \frac{1}{2}\right)^2$.

5. SIMULATIONS AND DISCUSSIONS

In our simulation the HIPERLAN/2 BRAN A channel model [10] for typical office environments (non line of sight) is used. The root mean square delay spread of the channel is 50 ns and the maximum delay is 390 ns. There are 18 taps. The number of subchannels is set to 64. In the following, we compare the BER performances of OFDM, QR using THP [1] and QR using nested lattice precoding for the situation when 250 bits are loaded per block. 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. The shaping code is a 4-state, rate-1/2 nonsystematic convolutional code with polynomials $[7, 5]_8$ and each codeword covers L=32 blocks in both cases. A is set to 1 for our proposed method.

When channel coding is not considered, Fig. 1 is still used with the channel coding block removed. Fig. 2 shows the simulation result without channel coding. Since the considered SNR range is high, the effect of noise is small. The method in [1] 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 [1], 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 [1]. The gain of the proposed scheme is obvious from Fig. 2. When a coded system is considered, a 32-state, rate-1/3 nonsystematic convolutional code with polynomials $[71, 65, 57]_8$ is used. All three systems use soft decoding. Comparing Fig. 3 and Fig. 2, we can find that the three curves become much closer to one another. Intuitively, most errors caused by clipping can be recovered by the channel coding. In addition, OFDM performs better than QR with THP. The proposed method still performs best although we use an suboptimal metric to do the joint decoding. Due to limited space, the simulation results with 150-bit loading are not shown here. We however find that the BER performance with 150-bit loading has the same trend as 250-bit loading. When channel coding is considered, Collings and Clarkson's method is significantly worse than the other two systems which perform similarly. This is because under low bit rate the PAPR is low, and clipping seldom happens. In addition, shaping can further decreases the clipping probability. Then the few errors caused by clipping can be effectively recovered by channel coding. Similar phenomenon has also been observed in [11]. We conclude that the proposed method is the most suitable for high speed transmission. It is easy to verify that the method proposed in [1] is indeed a special case of our approach. In particular, that method transmits signals on the ith subchannel with $\alpha_i = 1$, $\mathbf{u}_i = 0$ and there is no $\lfloor . \rceil_{C_q}$ in (5). Indeed, it applies THP with $A = 1/\sqrt{2}$ to avoid clipping and considers the uncoded case. Finally we list the advantages of the proposed method

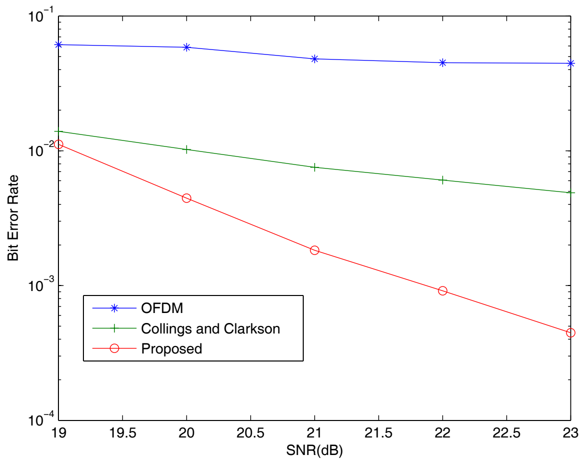

Fig. 2. Performance comparison for 250-bit loading, without channel coding.

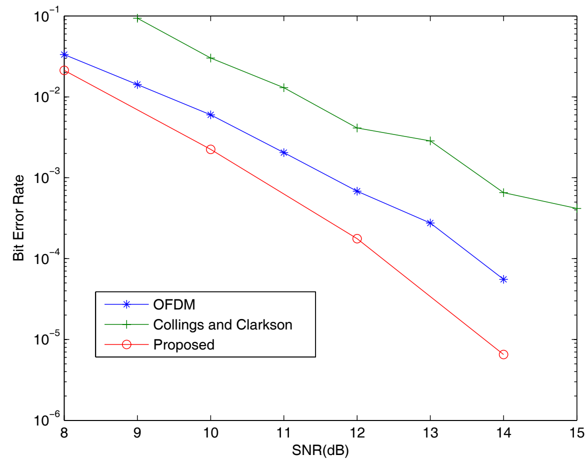

Fig. 3. Performance comparison for 250-bit loading, with channel coding.

over the method in [1]:

- From information-theoretic point of view, bit loading operation maximizes the system throughput under the assumption that subchannels are independent. Without the dither $\mathbf{u}_i$, the transmitted signal $\mathbf{x}_i$ in Collings and Clarkson's model will depend on $\mathbf{s}_i$. Then all subchannel signals are correlated. Thus bit loading may not achieve the desired capacity.

- The MMSE scaling factor α_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 [4].

6. CONCLUSION

In this paper, we use the concept of DPC with peak power constraint to replace the THP in [1]. 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 independent of the number of bits loaded. The proposed method performs best when the number of bit is large and hence is suitable for high speed transmission.

7. REFERENCES

[1] 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.

[2] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electr. Let.*, vol. 7, pp. 138–139, Mar. 1971.

[3] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, pp. 439–441, May 1983.

[4] W. Yu, D. Varodayan, and J.M. Cioffi, "Trellis and convolutional precoding for transmitter-based interference pre-subtraction," *IEEE Trans. Commun.*, vol. 53, no. 7, pp. 1220–1230, July 2005.

[5] 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.

[6] T. Philosof, "Precoding for interference cancellation at low SNR using nested codes," M.S. thesis, Tel-Aviv University, 2003.

[7] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct 2004.

[8] A. Bennatan, D. Burshtein, G. Caire, and S. Shamai, "Superposition coding for side-information channels," submitted to IEEE Transactions on Information Theory, 2004.

[9] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct 1998.

[10] J. Medbo and P. Schramm, "Channel models for HIPERLAN/2 in different indoor scenarios," *ETSI EP BRAN3ERI085B*, Mar 1998.

[11] H. Ochiai and H. Imai, "Performance analysis of deliberately clipped OFDM signals," *IEEE Trans. Commun.*, vol. 50, no. 1, pp. 89–101, Jan 2002.