A Low PAPR Constellation Mapping Scheme for Rate Compatible Modulation

Ruifeng Duan, Rongke Liu, Member, IEEE, Mahyar Shirvanimoghaddam, Member, IEEE, Yonghui Li, Senior Member, IEEE, Changwen Chen, Fellow, IEEE

Abstract—This paper proposes a novel non-linear constellation mapping (NLCM) scheme for the rate compatible modulation (RCM) approach in low-to-moderate signal to noise ratios (SNRs). In NLCM, each RCM coded symbol is mapped to a complex-valued amplitude phase shift keying (APSK) constellation point. Moreover, we provide a novel method of designing code degree and weight set for a wide SNR range. The proposed NLCM is developed to increase the throughput and decrease the peak to average power ratio (PAPR) at the same time. Both analysis and simulation results show that the NLCM scheme outperforms the original RCM in terms of throughput and PAPR.

Index Terms—rate compatible modulation, constellation mapping, PAPR, throughput.

I. INTRODUCTION

H ow to transmit information effectively and reliably over a time-varying channel has been one of the key research focuses in communication theory. Till now, adaptive coding and modulation (ACM) scheme [1] [2] has been widely used. In ACM, data is transmitted by selecting a proper configuration of modulation and coding scheme according to the channel state information (CSI). However, due to rapid channel variations and transmission delay, accurate CSI at the transmitter is not available. Therefore, ACM cannot precisely adapt to the channel condition, leading to a degradation in the throughput.

To overcome these problems, several physical layer rateless codes [3]–[6] have been recently proposed. Although Spinal codes [3] achieve the near-capacity performance, they have a polynomial decoding complexity in terms of the message block length, limiting their applications in practical communication systems. In RCM [4] [5] and analog fountain codes (AFCs) [6], coded symbols are generated from information bits by weighted sum operation in a rateless manner. AFCs also use a high rate low density parity check (LDPC) code as a precoder to reduce the error probability. Although a near-capacity performance is achieved, AFCs need a delicate design for the LDPC codes and the decoding in AFCs is more complex compared to RCM. Moreover, RCM does not take the PAPR into the code design. Although a small constellation size can be obtained by changing the degree and weight sets, RCM still cannot achieve a high throughput in low-to-moderate SNRs. Some APSK constellation mapping methods [7] [8] are also explored to increase the throughput. However, they do not jointly optimize the coded modulation and constellation mapping. A non-uniform modulation based on iterative polar quantification (IPQ) is proposed in [9] to achieve the channel capacity, but still, the PAPR has not been optimized.

In this paper, we take an step towards the design of a rateless coding strategy based on RCM by using NLCM with a limited PAPR. In the proposed scheme, each RCM coded symbol is mapped to a single complex constellation point. This will increase the codeword distance and thus reduce the decoding error probability. Through analysis, we show how the weight set of RCM and the mapping strategy should be designed to improve the throughput and reduce PAPR in low-to-moderate SNRs. These results are also validated by several simulations, showing the superiority of the proposed scheme.

The remainder of this paper is organized as follows. Section II describes the original RCM. Section III presents the proposed NLCM strategy. The performance of NLCM is analyzed in Section IV. Simulation results are provided in Section V, followed by some concluding remarks in Section VI.

II. ORIGINAL RATE COMPATIBLE MODULATION

RCM can be seen as a special case of AFCs [6], which are characterized by a code degree L, message length k, and a weight set Ws of size D. In RCM, the code degree and weight set size are equal, i.e., D = L, and Ws is assumed to be symmetric, which means that for any wk ∈ Ws, we have −wk ∈ Ws. Let b denote a binary message vector, then a real valued RCM coded symbol Um is calculated as follows:

\[ U_m = \sum_{l=1}^{L} g_{m,l} b_{m_l}, \quad m > 0, \quad (1) \]

where \( g_{m,l} \in W_s \) is the \( l^{th} \) non-zero weight coefficient corresponding to \( b_{m_l} \), which is selected randomly from \( b \). For an RCM truncated at length \( M \), the encoding process can be represented in matrix form \( \mathbf{U} = \mathbf{G} \mathbf{b} \), where \( \mathbf{G} = [g_{i,j}] \) denotes the \( M \) by \( k \) generator matrix, and \( g_{i,j} \in W_s \). In RCM, every two consecutive coded symbols create one complex-valued constellation symbol, \( X_m \), as follows:

\[ X_m = U_{2m-1} + \sqrt{-1} U_{2m}, \quad m > 0, \quad (2) \]
which is sent through the channel.

Let $q_c$ denote the number of all possible RCM coded symbols for given weight set $W_s$, it is calculated as follows:

$$ q_c = \left\{ \sum_{i=1}^{L} b_i w_i \mid b \in \{0, 1\}^L, w_i \in W_s \right\}, $$

where $|A|$ is the cardinality of set $A$, and $b$ represents the vector of information bits of size $L$. For example in [4], the code degree $L$ is 8, $W_s = \{\pm 4, \pm 4, \pm 2, \pm 1\}$, RCM coded symbols take values from -11 to 11; thus $q_c = 23$. It is then clear that the size of the complex constellation, $q_c$, is $q_c^2$.

The average signal power for RCM, $E_{so}$, is given by:

$$ E_{so} = \frac{1}{2} \sum_{i=1}^{L} \omega_i^2, $$

which arises from the fact that $W_s$ in RCM is symmetric and each information bit is either 0 or 1 with equal probability 0.5. The maximum value of an RCM coded symbol is then $U_{\text{max}} = \frac{1}{2} \sum_{i=1}^{L} |w_i|$, which is obtained when exactly half of the selected information bits are 1. The PAPR for original RCM, denoted by $\eta_{so}$, is then given by:

$$ \eta_{so} = 2U_{\text{max}}^2/E_{so} = \left( \sum_{i=1}^{L} |w_i|^2 \right)^2 / \sum_{i=1}^{L} \omega_i^2. $$

III. PROPOSED CONSTELLATION MAPPING SCHEME

A. Coded Symbol Generation

In NLCM, information bits are first encoded using the AFCs strategy presented in Section II, by using a weight set $W_s$ of size $D$. Unlike RCM, the code degree $L$ in NLCM is less than or equal to the weight set size, i.e., $L \leq D$. Moreover, to generate each coded symbol in NLCM, first a size-$L$ symmetric subset is selected from $W_s$ with equal probability and then $L$ information bits are randomly selected from the whole $k$ information bits with equal probability. The selected information bits are multiplied by the weights and added together to generate one NLCM coded symbol.

Let $W^{(j)}$ denote a symmetric subset of $W_s$. Then the coded symbols set generated by this subset can be shown as follows:

$$ U^{(j)} = \left\{ \sum_{i=1}^{L} b_i w_i^{(j)} \mid b \in \{0, 1\}^L, w_i^{(j)} \in W^{(j)} \right\}, $$

and the number of all possible coded symbols of a degree-$L$ NLCM with weight set $W_s$ can be calculated as follows:

$$ q_c = q = \left| \bigcup_{j \mid W^{(j)} \subseteq W_s} U^{(j)} \right|. $$

For example, for an NLCM of degree $L = 8$, weight set size $D = 16$, and weight set $W_{s16} = \{\pm 7, \pm 6, \pm 5, \pm 4, \pm 3, \pm 2, \pm 1\}$, the number of different NLCM coded symbols will be $q_c = 49$.

B. NLCM approach

Unlike RCM, where two consecutive coded symbols are mapped to a constellation point, in NLCM each coded symbol is mapped to an APSK constellation point. Usually, an APSK constellation is characterized by the number of points $q$, the number of rings, $C$, ring radiiuses $R_i$, the number of points on each ring $n_i$ and phase offset $\theta_i$ for $1 \leq i \leq C$. The IPQ constellation [9], as an optimal constellation in terms of mutual information, can be also adopted. However, the NLCM is expected to achieve high throughput and low PAPR simultaneously. In this paper, following strategies are adopted.

First, according to Shannon [10], the capacity of a Gaussian channel is achievable when the probability distribution of input symbols is Gaussian. In NLCM, we place coded symbols which are highly probable on inner rings. The ring radius $R_i$ can be given by follows [7], [8]:

$$ R_i = \sqrt{-\ln(1-P_i)}, $$

where $P_i$ denotes the probability that the constellation symbols are located within the $i$th ring, which is evaluated as follows:

$$ P_i = \sum_{j=1}^{L} P_{ij} + \left( \sum_{j=1}^{L} P_{ij} \right)/2, $$

where $P_{ij}$ represents the probability that the $j$th constellation point is located on the $i$th ring. The joint optimization of PAPR and throughput can be solved by designing $n_i$ and labeling the constellation symbols properly.

Secondly, the number of points on the $i$th ring, $n_i$, is given by $n_i = \min\{2^{|x|+1}, q - 1 - \sum_{j=1}^{i-1} n_j\}$ for $i = 1, 2, \ldots, C$. Then, $n_i$ is adjusted to satisfy $n_{i+1} > n_i$, which leads to a relatively large average power when ring radiiuses are given by Eq. 8. Consequently, it can decrease the PAPR.

Thirdly, constellation labeling is as follows. Coded symbols on the same ring are placed uniformly. Then, we label the constellation symbols to make two constellation points far away from each other with a considerable probability, when respective coded symbols are close. This means that two coded symbols are generated from two information sequences with different bit. As a result, the minimum codeword distance in NLCM which dominates the throughput will probably increase. Moreover, in NLCM, the phase offset $\theta_i$ is chosen to make the constellation points interlaced on adjacent two rings to further enlarge the minimum codeword distance, i.e., $\theta_{i+1} = \theta_i - \pi/n_{i+1}$, $i = 1, 2, \ldots, C - 1$ and $\theta_1 = \pi/n_1$.

Finally, by adjusting the radius, e.g., slightly decreasing the outermost radius, the PAPR can be decreased further without effecting the minimum codeword distance because of the low transmission probabilities of coded symbols on the outermost ring. To sum up, by adjusting the radius and designing the number of constellation points for each ring, and appropriate labeling, a larger minimum codeword distance and a desirable PAPR can be achieved simultaneously.

The constellation point $X_m$, for each real valued NLCM coded symbol $U_m$, can be then written as follows:

$$ X_m = \sum_{i=1}^{C} R_i I(U_m \in U_{\text{sub}_i}) e^{j(\phi(U_m))}, $$

where $I(x) = 1$ if $x$ is true and $U_{\text{sub}_i}$ represents the $i$th ring; $\phi(U_m)$ is the phase determined by $\theta_i$ and constellation labeling, and $X_m = 0$ if $U_m = 0$. An example of the proposed NLCM scheme with $L = 8$ and $W_{s16}$ is depicted in Fig. 1. The average signal power for the proposed NLCM scheme, denoted by $E_{sn}$, can be calculated as follows:

$$ E_{sn} = \sum_{i=1}^{C} R_i^2 \sum_{j=1}^{n_i} P_{ij}, $$

and its PAPR, denoted by $\eta_{sn}$, is given by:

$$ \eta_{sn} = R_i^2/E_{sn}. $$
For the simplicity, we also introduce the following notations: \( R = \{R_i/\sqrt{E_{sn}}|1 \leq i \leq C\} \) and \( n = \{n_i|1 \leq i \leq C\} \).

IV. ANALYSIS OF THE PROPOSED NLCM SCHEME

We consider an additive white Gaussian noise (AWGN) channel, where the received signal is given by \( Y = X + z = f(Gb) + z \), \( X \) denotes the vector of constellation symbols or codeword, and \( z \) represents zero-mean complex AWGN vector with variance \( \sigma^2 \) in each signal dimension. According to the Maximum Likelihood (ML) criterion, the optimal estimation of information sequence, \( \hat{b} \), is given by

\[
\hat{b} = \arg \min_b \|Y - f(Gb)\|^2.
\]  

Let \( b' \) denote an information sequence which is different from \( b \) in at least one place. The pairwise error event probability (PEEP) can then be calculated as follows:

\[
p(b \rightarrow b') = p(\|Y - f(Gb')\|^2 < \|Y - f(Gb)\|^2|G).
\]  

The PEEPs for the original and proposed schemes, denoted by \( p^{(o)} \) and \( p^{(n)} \), respectively, can be found as follows [6]:

\[
p^{(o)} = \frac{1}{2\sigma^2} \sqrt{||Gb - Gb'||^2/E_{so}},
\]  

\[
p^{(n)} = \frac{1}{2\sigma^2} \sqrt{||f(Gb) - f(Gb')|^2/E_{sn}},
\]

where \( Q(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2}dx \). It is clear that PEEP is mainly characterized by the codeword distance. The average throughput, defined as the number of correctly received information bits divided by the number of transmitted symbols [11], is mainly characterized by the frame error probability, \( P_F \), which can be calculated as follows:

\[
P_F = \sum_X p(X) \sum_{X' \neq X} p(X \rightarrow X'),
\]

where \( p(X) = 1/2^C \). By using the minimum codeword distance, \( d_{\min} \), \( P_F \) can be lower bounded as follows [12]:

\[
P_F \geq Q(d_{\min}/2\sigma).
\]

In RCM, the minimum codeword distance is usually given when \( b \) and \( b' \) are different only in one place [5]. Assuming \( b_i \neq b'_i = b_i \), for the original and proposed mapping schemes, the minimum square codeword distance, denoted by \( d_0^n \), can be expressed respectively as Eqs. 18 and 21:

\[
d_0^n = \min_{b \in \{0, 1\}^k} \left\| \sum_{j=1}^{k} G[j]b_j - \sum_{j=1}^{k} G[j]b'_j \right\|^2
\]

\[
= \frac{1}{E_{so}} \|G[i](b_i - b'_i)\|^2 = \frac{2M}{L} \frac{1}{T} \sum_{j=1}^{L} \omega_j^2/E_{so} = \frac{2M}{L},
\]

where \( G[j] \) is the \( j \)th column of \( G \); \( M \) is the number of constellation symbols. For the original and proposed mapping schemes, the number of coded symbols is \( 2M \) and \( M \), respectively, and each information bit is selected \( 2ML/k \) and \( ML/k \) times on average, respectively, when generating all coded symbols. In Eq. 21, \( \alpha \) and \( \beta \) are the non-zero elements in the \( i \)th column, and we define:

\[
U_{bi} = \sum_{l=1}^{L} b_{ml} \omega_l + b_i \omega_n, \omega_n \in W(j),
\]

\[
\alpha = \frac{L}{E_{sn}} \frac{1}{C^L/2} \sum_{j=1}^{L} \sum_{W(j) \in W_s} \int_{W(j)} \omega_n^2 \left| f(U_{bi}) - f(U_{bj}) \right|^2.
\]

Eqs. 18 and 21 arise from the fact that each weight coefficient \( \omega_j \) turns up at each column of \( G \) with equal probability and each subset \( W(j) \) of weight set \( W_s \) is selected with equal probability. For the NLCM, \( |f(U_{bi}) - f(U_{bj})|_{min} \) denotes the minimum distance between two constellation points for given \( W(j) \) and \( \omega_n \), when respective coded symbols contain only one different bit. The lower bound on \( d_0^n \) will be achieved when each \( |f(U_{bi}) - f(U_{bj})| \) takes the minimum value with corresponding \( W(j) \) and \( \omega_n \). As mentioned in Section III, in NLCM, to decrease the PAPR and increase the \( d_{\min} \), we should label constellation points to further separate the close coded symbols subject to PAPR and normalized signal power.

Based on the proposed method, we obtain three mapping schemes denoted by NLCM1, NLCM2, and NLCM3, which are shown in Table I, and the corresponding \( \alpha \) are respectively 5.13, 5.06 and 4.66, all larger than 4. From Eqs. 17, 18 and 21, the NLCM schemes have smaller lower bound on the frame error probability and accordingly a higher throughput.

V. SIMULATION AND DISCUSSION

Simulations are carried out under standard complex AWGN channels at SNRs from -5 dB to 20 dB for both the proposed NLCM and original RCM schemes. At the same time, the proposed NLCM schemes with IPQ constellation, denoted by NLCM-IPQ, are also simulated. Table I lists all the parameters, where \( \eta \) represents the PAPR. Fig. 2 shows the throughput versus the SNR for frame error probability less than \( 10^{-4} \). The number and length of the data frame are \( 10^6 \) and \( 10^5 \), respectively, for both the original and proposed schemes.

As can be seen in Fig. 2, with the same constellation size, NLCM1, NLCM2, and NLCM3 respectively outperform the original RCM1, RCM2, and RCM3 across the entire SNR range. Especially in SNR values below 12 dB, the proposed methods increase the throughput by 21.7%–163.5%, 8.9%–267.2% and 6.5%–469.4% respectively. Moreover, the PAPR in the proposed NLCM is significantly reduced, which is beneficial for practical systems. Besides, saturation effect...
the proposed NLCM scheme improves the throughput and both theoretical analysis and simulation results show that probability performance and obtain a heuristic joint design to effectively increase the minimum codeword distance, thus observed in [5] also appears here. However, the proposed mapping schemes achieve higher maximum throughput with the same constellation size compared to the original RCM which usually have a large PAPR.

In addition, the proposed NLCM3 achieves a better rate performance in SNRs below 14 dB, compared to the original RCM with constellation size of 529 in [4]. Rate saturation in high SNRs in the proposed scheme can be avoided by using larger code degrees and weight coefficients. As mentioned in [5] and as can be seen in Fig. 2, decreasing the degree and weight coefficients cannot improve the throughput of original RCM in the low SNR regime. Moreover, with the cost of a higher PAPR, NLCM3-IPQ provides a slightly better throughput performance compared to the NLCM3 with proposed signal constellation plane. This is due to the optimal design of the IPQ constellation in terms of the mutual information.

VI. CONCLUSION

In this paper, we develop a novel constellation mapping scheme for RCM to achieve higher throughput in low-to-moderate SNRs and reduce the PAPR. This is achieved by mapping each real-valued coded symbol to one complex-value APSK constellation point in a non-linear fashion. Accompanying with a larger code degree and a larger weight set, the novel constellation mapping approach has been designed to effectively increase the minimum codeword distance, thus achieving a higher throughput. Moreover, we analyze the error probability performance and obtain a heuristic joint design for degree and weight set and constellation mapping. Finally, both theoretical analysis and simulation results show that the proposed NLCM scheme improves the throughput and significantly reduces the PAPR of original RCM scheme in a wide range of SNRs.

REFERENCES