Iterative Decoding of Superposition Coding[1]

Jun Tong and Li Ping
Department of Electronic Engineering, City University of Hong Kong, jun.tong@student.cityu.edu.hk, eeliping@cityu.edu.hk

Abstract

Clipping is applied to superposition coding systems to reduce the peak-to-average power ratio (PAPR) of the transmitted signal. The performance limit is investigated through evaluating the mutual information driven by the induced peak-power-limited input signals. It is shown that the channel capacity can be approached by clipped superposition coding systems. To alleviate the performance degradation due to clipping noises, we develop a soft compensation algorithm that is combined with soft-input-soft-output (SISO) decoding algorithms in an iterative manner. Simulation results show that with the proposed algorithm, most performance loss can be recovered.

1 Introduction

The traditional trellis coded modulation (TCM) [1] is based on uniformly spaced constellation with equal probability for every signaling point. In an additive white Gaussian noise (AWGN) channel, there is an asymptotical gap of about 1.53 dB (the so-called shaping gap) between the achievable performance of TCM (and other schemes based on uniform signaling [2][3]) and the channel capacity [4]. To narrow this gap, Gaussian signaling (that produces signals with Gaussian distribution) can be applied using shaping techniques, such as assigning non-uniform probabilities on different signaling points [4]-[9]. The related advantage is referred to as shaping gain [9].

Superposition coding is an alternative approach to bandwidth efficient coded modulation [10]-[13]. With superposition coding, several coded sequences (each referred to as a layer) are linearly superimposed before transmission. Consequently, the transmitted signal exhibits an approximately Gaussian distribution that matches to an AWGN channel. This provides an alternative means to achieve shaping gain. The work in [10] [11] shows that such a concept is realizable with practical encoding and decoding methods. Simulation results show that a superposition coding scheme can operate within the shaping gap [10], surpassing the theoretical limit of the uniform signaling based methods.

However, there is a practical concern on superposition coding: an ideal Gaussian distributed signal has an infinite peak power even its average power is finite. The resultant high peak-to-average power ratio (PAPR) may cause a problem for radio frequency amplifier efficiency [14]. Clipping is a straightforward technique for PAPR reduction but it may lead to performance degradation [14]-[16]. The same PAPR problem also exists in other schemes for achieving shaping gain [9]. In this paper, we will investigate the clipping issue for superposition codes. We will show by mutual information analysis that the theoretical penalty due to clipping is marginal for practical PAPR values. Furthermore, with the same PAPR, clipped Gaussian signaling can actually outperform uniform signaling. This demonstrates the theoretical advantage of the superposition codes.

We will then devise a practical method to recover the performance loss due to clipping by exploiting the characteristics of clipping noises. It can be easily incorporated into the overall iterative receiver structure based on the low-cost multiuser detection principles developed in [17][18]. Compared with other clipping noise mitigation techniques (such as the hard-decision-aided methods [14][15]), the proposed method can achieve noticeable performance improvement.

2 Transmitter and PAPR

2.1 Encoding

We consider a K-layer superposition coding system. The encoding scheme is shown in Fig. 1. A binary data sequence $\boldsymbol{u}$ is partitioned into K subsequences $\{\boldsymbol{u}^{(k)}\}$. The kth subsequence $\boldsymbol{u}^{(k)}$ is encoded by a binary encoder (ENC-k) at the kth layer, resulting in a coded bit sequence $\boldsymbol{c}^{(k)} = \{ c_j^{(k)} \}$ of length J with $c_j^{(k)} \in \{0, 1\}$. The randomly interleaved version $\boldsymbol{v}^{(k)}$ of $\boldsymbol{c}^{(k)}$, from interleaver-k (INTL-k), is then mapped to a quadrature phase shifting keying (QPSK) sequence $x_j^{(k)} = x_{\mathrm{Re},j}^{(k)} + i x_{\mathrm{Im},j}^{(k)}$ where $i = \sqrt{-1}$. The subscripts $_{\mathrm{Re}}$ and $_{\mathrm{Im}}$ are used to denote the real and imaginary parts of complex numbers, respectively. We will assume that $x_{\mathrm{Re},j}^{(k)} \in \{+1, -1\}$ and so does $x_{\mathrm{Im},j}^{(k)}$.

[1] This work was fully supported by a grant from the Research Grant Council of the Hong Kong Special Administrative Region, China, under project CityU 117305.

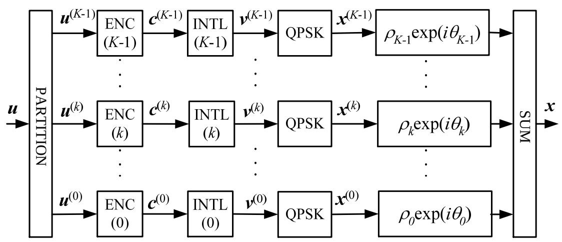

Fig. 1. Encoder of a superposition coding system.

The output signal at time j is a linear superposition of independent coded symbols:

$$x_j \equiv \sum_{k=0}^{K-1} r_k e^{iq_k} x_j^{(k)} \tag{1}$$

where $r_k\,(r_k>0)$ is an amplitude factor and q_k $(0 \le q_k < p/2)$ a rotation angle for layer-k. The overall rate is $R = 2\sum_{k=0}^{K-1} R_k$ in bits/symbol, where R_k is the rate of the kth binary component code. Appropriate r_k's, which are necessary to facilitate the detection at the receiver, can be determined by power allocation methods [10][11].

2.2 Peak-to-Average Power Ratio

Let $\mathrm{E}(\cdot)$ denote the mathematical expectation and $|\cdot|$ the absolute value. The PAPR of x_j is defined as

$$\mathrm{PAPR} = \frac{\max_j (|x_j|^2)}{\mathrm{E}(|x_j|^2)}. \tag{2}$$

We assume that all coded bits $c_j^{(k)}$ are independent, identically distributed (i.i.d.) random variables. It can be verified that, the PAPR is maximized when all q_k's are equal. As an example, for a 5-layer system with $\{r_k\}$ = {0.1634, 0.2380, 0.3467, 0.5051, 0.7358}, the maximum PAPR = 5.97 dB is reached at $q_k = 0$, $\forall k$. On the other hand, if we set $q_k = k\pi/10$, $\forall k$, then the PAPR is reduced to 5.39 dB.

In order to further suppress PAPR, we can clip x_j to $\bar{x}_j$ before transmission according to the following rule:

$$\bar{x}_j \equiv \begin{cases} x_j, & |x_j| \le A \\ Ax_j/|x_j|, & |x_j| > A \end{cases} \tag{3}$$

where $A > 0$ is the clipping threshold (a real value). The clipped signal $\bar{x}_j$ is then transmitted over an AWGN channel, and the PAPR of the transmitted signal is $A^2/\mathrm{E}(|\bar{x}_j|^2)$. The received signal can be written as

$$y_j = \bar{x}_j + w_j, \tag{4}$$

where $\{w_j\}$ are samples of a circularly symmetric complex Gaussian random process with zero mean and variance s^2 per dimension. The ratio of energy per bit (E_b) to the noise power spectral density (N_0 $=2s^2$) is given by $E_b/N_0 \equiv \mathrm{E}(|\bar{x}_j|^2)/(2Rs^2)$. The task of the receiver is to estimate $\boldsymbol{u}$ after observing $\boldsymbol{y} = \{y_j\}$. We will discuss this later in Section 4.

3 Effect of Clipping on the Achievable Rates

We now investigate the impact of clipping on the performance limits of superposition coding systems. Consider a complex AWGN channel $Y = X + W$, where X and Y are the input and output signal, respectively, and W an independent circularly symmetric complex Gaussian random variable with zero mean and variance s^2 per dimension. For a given distribution of X, the maximum reliable transmission rate can be quantified by the mutual information [12],

$$I(X;Y) = h(Y) - h(W) = -\mathrm{E}(\log p(Y)) - \log(2pes^2) \tag{5}$$

where $h(\cdot)$ is the entropy function. For most cases of interest, $\mathrm{E}(\log p(Y))$ (and hence $I(X; Y)$) can be computed by numerical integration methods. We will examine continuous and discrete X separately below.

3.1 Continuous Input Signal

We first consider X as a clipped version of a circularly symmetric complex Gaussian random variable with zero mean and variance 1/2 per dimension. The clipping rule given by (3) is used, and the resulting PAPR is

$$\mathrm{PAPR} = \frac{A^2}{\int_{|x|\le A} |x|^2\, p(x)dx + A^2 \int_{|x|>A} p(x)dx}, \tag{6}$$

where $p(x)=1/\pi\exp(-|x|^2)$. Specifically, when $A=\infty$, the PAPR is infinite and

$$I(X;Y) = \log(1+\mathrm{E}(|X|^2)/(2s^2))$$

which is the channel capacity.

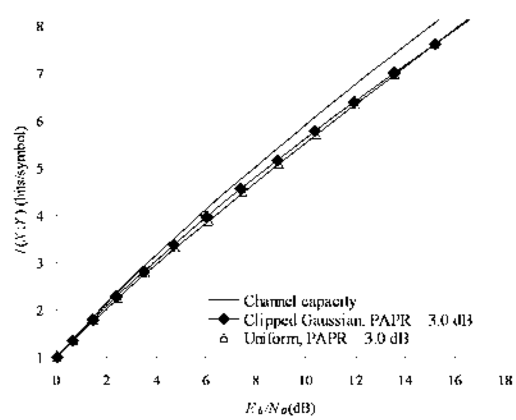

Fig. 2. Mutual information versus E_b/N_0 curves for different continuous inputs.

The achieved mutual information for the above clipped Gaussian signaling is shown in Fig. 2. We set A = 1.26, and the resultant PAPR is 3.0 dB. We can see that the performance of the clipped Gaussian signaling is close to the channel capacity. As a comparison, we also examine a case where X is uniformly distributed within a circle centering at the origin. This is a limiting case for systems based on circular QAM constellations [7], and the associated PAPR is also 3.0 dB. It is seen that the clipped Gaussian signaling is slightly better than the uniform signaling for up to about 7 bits/symbol and the situation is reversed for higher rates.

3.2 Discrete Input Signal

We now examine the superposition coding systems where X is a discrete variable. For the K-layer scheme described in Fig. 1, X is the summation of K complex random variables, and so the distribution of X is approximately Gaussian before clipping.

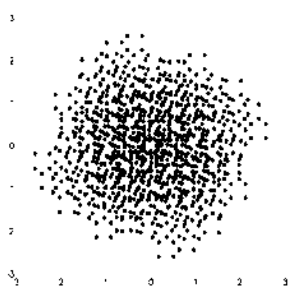

Fig. 3. Signal constellation of a superposition coding system parameterized by $K = 5$, $\{\theta_k = k\pi/10, \forall\ k\}$ and $\{\rho_k\} = \{0.1634, 0.2380, 0.3467, 0.5051, 0.7358\}$.

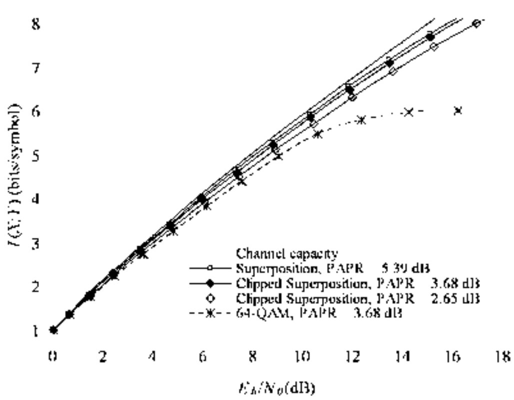

Fig. 4. Mutual information versus E_b/N_0 curves for clipped superposition coding systems. The parameters are the same as those in Fig. 3.

We take a $K = 5$-layer scheme as an example. The signal constellation of X before clipping is depicted in Fig. 3. The related PAPR without clipping is 5.39 dB. The achievable $I(X;\ Y)$ for X with and without clipping is shown in Fig. 4. The performance limit of the conventional 64-QAM[2] systems (PAPR = 3.68 dB) and the channel capacity are also included for comparison. (In this paper, we are most interested in transmission rates of 4 and 5 bits/symbol. Hence, the 64-QAM signaling, which is often applied in traditional coded modulation schemes for such rates [3]-[5], is used for comparison.) From Fig. 4, we can make the following observations:

- Without clipping, the superposition signaling can provide noticeable shaping gains over the conventional 64-QAM signaling, and hence can closely approach the channel capacity. For a target rate of 5 bits/symbol, the gap between the required E_b/N_0 and the channel capacity is only 0.21 dB, representing a shaping gain of more than 1 dB.
- Although clipping degrades the achievable rate, the effect is not serious. For example, at $I(X;\ Y) = 5$ bits/symbol, only 0.16 dB loss in E_b/N_0 is introduced by clipping with PAPR = 3.68 dB. In addition, even at a low PAPR = 2.65 dB, the clipped superposition signaling still performs better than the conventional 64-QAM signaling with PAPR = 3.68 dB. This reveals that the clipped superposition coding scheme can provide a good trade-off between PAPR and achievable rate.

[2] Throughout this paper, a "conventional QAM" constellation represents a square QAM constellation in which signal points are equispaced and utilized with equal probabilities [7].

4 Iterative Decoding

Now we turn our attention to practical superposition coding systems. So far, we have shown that, theoretically, the performance penalty incurred by clipping is not severe for reasonable clipping thresholds A. However, if the clipping effect is not treated, performance would degrade significantly even when A is moderate, as will be shown by simulation examples later.

The following observations suggest a possible approach towards a solution:

- If $|y_j|$, the amplitude of the received signal, is large, then the probability is high that x_j has been clipped.
- If $|y_j|$ is small, then the clipping probability is small.

We may exploit this fact to compensate for the clipping effect. This is the underlying rationale for the soft compensation algorithm (SCA) presented below.

4.1 The Basic Receiver Structure

We first outline the basic receiver structure without the compensation for clipping effect. The superposition coding systems can be treated as perfectly coordinated multiple-access systems by viewing one layer as one user. Hence multiuser detection principles can be applied. Specifically, due to the similarity between the superposition coding system and the interleave-division multiple-access system, we employ a sub-optimal iterative receiver similar to that in [17]-[19].

As illustrated in Fig. 5, the receiver consists of one elementary signal estimator (ESE) and K soft-input soft-output (SISO) decoders (DECs). They are connected by the INTLs and DEINTLs (de-interleavers), operating iteratively. The messages passing between the ESE and the DECs are the so-called *extrinsic* information values. Since the SISO decoders perform standard *a posteriori* probability (APP) decoding, we will only focus on the ESE which performs SISO detection to generate the *a posteriori* log-likelihood ratios (LLRs) for all coded bits [17].

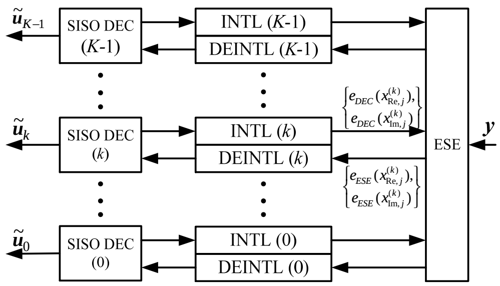

Fig. 5. Block diagram of the iterative decoding / detection algorithm.

We first ignore the clipping effect and concentrate on the detection for layer-*k*. Rewrite (4) as

$$y_j = x_j + w_j = r_k e^{iq_k} x_j^{(k)} + z_j^{(k)} \tag{7}$$

where

$$z_j^{(k)} = \sum_{k' \neq k} r_{k'} e^{iq_{k'}} x_j^{(k')} + w_j \tag{8}$$

represents the interference-plus-noise with respect to $x_j^{(k)}$. In order to detect $x_j^{(k)}$, we generate

$$\hat{y}_j^{(k)} = e^{-iq_k} y_j = r_k x_j^{(k)} + \hat{z}_j^{(k)}, \tag{9}$$

where $\hat{z}_j^{(k)} = e^{-iq_k} z_j^{(k)}$. We approximate $\hat{z}_j^{(k)}$ as an additive complex Gaussian variable. Then the output LLR ($e_{\mathrm{ESE}}(x^{(k)}_{\mathrm{Re},j})$) for $x^{(k)}_{\mathrm{Re},j} \in \{+1, -1\}$ can be computed as follows. (We can handle $x^{(k)}_{\mathrm{Im},j}$ in a similar way.)

$$\begin{aligned} e_{\mathrm{ESE}}(x^{(k)}_{\mathrm{Re},j}) &\equiv \ln\left(\frac{\Pr(\hat{y}^{(k)}_{\mathrm{Re},j} \mid x^{(k)}_{\mathrm{Re},j} = +1)}{\Pr(\hat{y}^{(k)}_{\mathrm{Re},j} \mid x^{(k)}_{\mathrm{Re},j} = -1)}\right) \\ &= \ln\left(\frac{\left(\mathrm{Var}(\hat{z}^{(k)}_{\mathrm{Re},j})\right)^{-1/2} \exp\left(-\left(\hat{y}^{(k)}_{\mathrm{Re},j} - r_k - \mathrm{E}(\hat{z}^{(k)}_{\mathrm{Re},j})\right)^2 \Big/ \left(2\mathrm{Var}(\hat{z}^{(k)}_{\mathrm{Re},j})\right)\right)}{\left(\mathrm{Var}(\hat{z}^{(k)}_{\mathrm{Re},j})\right)^{-1/2} \exp\left(-\left(\hat{y}^{(k)}_{\mathrm{Re},j} + r_k - \mathrm{E}(\hat{z}^{(k)}_{\mathrm{Re},j})\right)^2 \Big/ \left(2\mathrm{Var}(\hat{z}^{(k)}_{\mathrm{Re},j})\right)\right)}\right) \\ &= 2 r_k \cdot \frac{\hat{y}^{(k)}_{\mathrm{Re},j} - \mathrm{E}(\hat{z}^{(k)}_{\mathrm{Re},j})}{\mathrm{Var}(\hat{z}^{(k)}_{\mathrm{Re},j})} \end{aligned} \tag{10}$$

where $\mathrm{Var}(\cdot)$ denotes the variance function. The computational details for $\mathrm{E}(\hat{z}^{(k)}_{\mathrm{Re},j})$ and $\mathrm{Var}(\hat{z}^{(k)}_{\mathrm{Re},j})$ are given in Appendix A.

4.2 Soft Compensation Algorithm

Next we derive the SCA which performs a joint process to treat the inter-layer interference and clipping noise. Again, we focus on layer-*k*. We rewrite (4) as

$$y_j = x_j + z_j + w_j = r_k e^{iq_k} x_j^{(k)} + x_j^{(k)} \tag{11}$$

where $x_j^{(k)}$ denotes the distortion consisting of two terms,

$$x_j^{(k)} \equiv z_j^{(k)} + z_j \tag{12}$$

with $z_j^{(k)}$ given by (8) and z_j the clipping noise. From (3), z_j is given by

$$z_j = \begin{cases} 0, & |x_j| \leq A \\ A x_j / |x_j| - x_j, & |x_j| > A \end{cases}. \tag{13}$$

Consider the detection of $x^{(k)}_{\mathrm{Re},j}$ based on (11) and (12). The optimal solution has an exponential complexity and hence is not practical (especially when K is large). Instead, we can adopt the following sub-optimal SCA with complexity only slightly higher than the basic detection algorithm in Section 4.1. As will be shown later, most of the performance loss can be recovered by SCA.

Similar to (9), we generate

$$\hat{y}_j^{(k)} = e^{-iq_k} y_j = r_k x_j^{(k)} + \hat{x}_j^{(k)}, \tag{14}$$

where $\hat{x}_j^{(k)} = e^{-iq_k} x_j^{(k)}$. We again approximate $\hat{x}_j^{(k)}$ by an additive complex Gaussian variable. Similar to (10), we have

$$\begin{aligned} & e_{\mathrm{ESE}}(x^{(k)}_{\mathrm{Re},j}) \\ &= \ln\left(\frac{\left(\mathrm{Var}^+(\hat{x}^{(k)}_{\mathrm{Re},j})\right)^{-1/2} \exp\left(-\left(\hat{y}^{(k)}_{\mathrm{Re},j} - r_k - \mathrm{E}^+(\hat{x}^{(k)}_{\mathrm{Re},j})\right)^2 \Big/ \left(2\mathrm{Var}^+(\hat{x}^{(k)}_{\mathrm{Re},j})\right)\right)}{\left(\mathrm{Var}^-(\hat{x}^{(k)}_{\mathrm{Re},j})\right)^{-1/2} \exp\left(-\left(\hat{y}^{(k)}_{\mathrm{Re},j} + r_k - \mathrm{E}^-(\hat{x}^{(k)}_{\mathrm{Re},j})\right)^2 \Big/ \left(2\mathrm{Var}^-(\hat{x}^{(k)}_{\mathrm{Re},j})\right)\right)}\right) \\ &= -\frac{1}{2}\ln\left(\frac{\mathrm{Var}^+(\hat{x}^{(k)}_{\mathrm{Re},j})}{\mathrm{Var}^-(\hat{x}^{(k)}_{\mathrm{Re},j})}\right) - \frac{\left(\hat{y}^{(k)}_{\mathrm{Re},j} - r_k - \mathrm{E}^+(\hat{x}^{(k)}_{\mathrm{Re},j})\right)^2}{2\mathrm{Var}^+(\hat{x}^{(k)}_{\mathrm{Re},j})} \\ &\quad + \frac{\left(\hat{y}^{(k)}_{\mathrm{Re},j} + r_k - \mathrm{E}^-(\hat{x}^{(k)}_{\mathrm{Re},j})\right)^2}{2\mathrm{Var}^-(\hat{x}^{(k)}_{\mathrm{Re},j})} \end{aligned} \tag{15}$$

where $\mathrm{E}^+(\hat{x}^{(k)}_{\mathrm{Re},j})$ and $\mathrm{E}^-(\hat{x}^{(k)}_{\mathrm{Re},j})$ denote the means of $\hat{x}^{(k)}_{\mathrm{Re},j}$ under the hypothesis that $x^{(k)}_{\mathrm{Re},j}$ is either +1 or −1, respectively. $\mathrm{Var}^+(\hat{x}^{(k)}_{\mathrm{Re},j})$ and $\mathrm{Var}^-(\hat{x}^{(k)}_{\mathrm{Re},j})$ are defined similarly. Note that here the statistics of $\hat{x}_j^{(k)}$ are different for different hypothesis, since the clipping effect depends on hypothesis. The details on the generation of the mean and variance of $\hat{x}^{(k)}_{\mathrm{Re},j}$ can be found in Appendix B.

As a summary, we have two comments on SCA.

- Since the statistics of $\hat{x}_j^{(k)}$ is related to the hypothesis on $x^{(k)}_{\mathrm{Re},j}$, (15) appears slightly more complicated than (10).
- SCA is essentially a turbo-type clipping noise cancellation technique based on the *extrinsic* information produced by the SISO decoders. This distinguishes it from the hard-decision-aided clipping noise cancellation technique in [15] and the signal reconstruction techniques in [14][16].

5 Numerical Results

In this section, we present simulation results to demonstrate the performance of clipped superposition coding systems. The rate-1/2 doped code [20] with data length 10^5 is chosen as the component code for each layer. We set K = 4 and 5. The corresponding total rates R = 4 and 5 bits/symbol, and the PAPR without clipping are 4.74 and 5.39 dB, respectively. Clipping is applied to both systems to reduce the PAPR to 3.68 dB. The iteration number in the component decoders is set to 200, and the number of iterations between detection and decoding is set to 6. The entropy based stop criterion introduced in [10] is used.

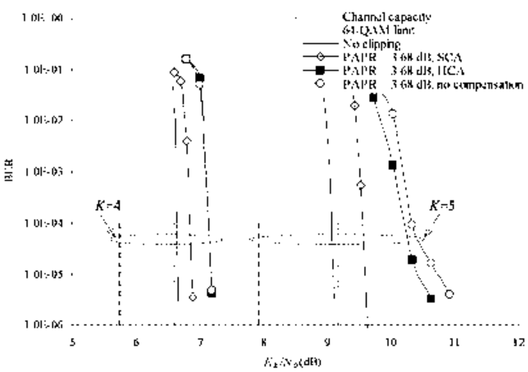

Fig. 6. Performances of clipped superposition coding systems with different algorithms. For K = 4, $\{\rho_k\}$ = {0.2430, 0.3529, 0.5124, 0.7442} and $\{\theta_k = k\pi/8, \forall k\}$. The parameters for K = 5 are the same as those in Fig. 3

The simulated BER performances for the above examples are shown in Fig. 6. For comparison, we also include the performance of a hard-decision-aided compensation algorithm (HCA), which is closely related to the method devised in [15] for OFDM systems, as briefly discussed in Appendix C.

From Fig. 6, we can see that the performance of superposition coding schemes without clipping is quite close to the channel capacity. At BER of 10^{-5}, the gaps between the required E_b/N_0 and the channel capacity are only 0.9 and 1.2 dB for R = 4 and 5 bits/symbol, respectively. Notice that for R = 5 bits/symbol, the shaping gap with respect to the conventional 64-QAM constellation is about 1.25 dB.

On the other hand, seen from Fig. 6, the performance of the clipped systems deteriorates when no compensation is applied. For example, at PAPR = 3.68 dB, the performance gaps to the channel capacity increase to about 1.5 and 3.0 dB for K = 4 and 5, respectively. Using SCA, however, the performance gaps are reduced to 1.2 and 1.7 dB, respectively. This indicates a possible direction towards the theoretical limit given in Section 3 for clipped systems.

For comparison, the best simulation results (to the authors' knowledge) based on trellis shaping and equispaced QAM constellations for both R = 4 and 5 bits/symbol are about 0.8 dB away from the channel capacity, as reported in [5][6]. However, the associated PAPRs in [5] and [6] are relatively high (7.26 and 8.93 dB for R = 4 and 5 bits/symbol, respectively).

It is seen that SCA is superior to HCA. The reason is that hard decisions based HCA suffers from error propagations and hence is less robust than SCA. As read from Fig. 6, at PAPR = 3.68 dB and BER=10^{-5}, SCA outperforms HCA by about 0.3 dB for K = 4 and 1 dB for K = 5.

6 Conclusions

We have investigated a peak-power-limited superposition coding system based on clipping. By computing the mutual information achieved with the clipped input signal, we have shown that noticeable shaping gains can be achieved with reasonable clipping thresholds. To combat the clipping effect for practically coded systems, we have derived an efficient iterative soft compensation algorithm. The simulation results show that a good trade-off between PAPR and performance can be achieved with the proposed algorithm.

7 Appendix A

First, a definition. Let x be a complex random variable and E(x) be its mean. Define its covariance matrix as

$$\boldsymbol{Cov}(x)=\begin{pmatrix} \mathrm{Var}(x_{\mathrm{Re}}) & \mathrm{E}(x_{\mathrm{Re}}x_{\mathrm{Im}})-\mathrm{E}(x_{\mathrm{Re}})\mathrm{E}(x_{\mathrm{Im}}) \\ \mathrm{E}(x_{\mathrm{Re}}x_{\mathrm{Im}})-\mathrm{E}(x_{\mathrm{Re}})\mathrm{E}(x_{\mathrm{Im}}) & \mathrm{Var}(x_{\mathrm{Im}}) \end{pmatrix}$$

Let $e_{\mathrm{DEC}}(x)=\ln\left(\Pr(x=+1)/\Pr(x=-1)\right)$ be the extrinsic LLR for coded bit x from the DECs (see Fig. 5). Following [17][18], $\mathrm{E}(x_j^{(k)})$ and $\boldsymbol{Cov}(x_j^{(k)})$ can be estimated by

$$\mathrm{E}(x_j^{(k)})=\tanh\left(e_{DEC}(x_{\mathrm{Re},j}^{(k)})/2\right)+i\tanh\left(e_{DEC}(x_{\mathrm{Im},j}^{(k)})/2\right) \quad (16a)$$

$$\boldsymbol{Cov}(x_j^{(k)})=\begin{pmatrix} 1-\left(\mathrm{E}(x_{\mathrm{Re},j}^{(k)})\right)^2 & 0 \\ 0 & 1-\left(\mathrm{E}(x_{\mathrm{Im},j}^{(k)})\right)^2 \end{pmatrix}, \quad (16b)$$

where we have assumed that the extrinsic LLRs for the real and imaginary parts of $x_j^{(k)}$ are uncorrelated, and thus the off diagonal entries of $\boldsymbol{Cov}(x_j^{(k)})$ are zeros. (Initially, we set $\mathrm{E}(x_j^{(k)})=0$ and $\boldsymbol{Cov}(x_j^{(k)})=\boldsymbol{I}$, implying no feedback from the DECs.)

From (1) and (4),

$$\mathrm{E}(y_j)=\sum_{k=0}^{K-1} r_k e^{iq_k}\,\mathrm{E}(x_j^{(k)})\,, \quad (17a)$$

$$\boldsymbol{Cov}(y_j)=\sum_{k=0}^{K-1} r_k^2 \boldsymbol{R}^{(k)}\boldsymbol{Cov}(x_j^{(k)})(\boldsymbol{R}^{(k)})^T+s^2\boldsymbol{I}\,, \quad (17b)$$

where $\boldsymbol{R}^{(k)}=\begin{pmatrix} \cos(q_k) & -\sin(q_k) \\ \sin(q_k) & \cos(q_k) \end{pmatrix}$, the superscript ^{T} denotes transpose of matrixes and $\boldsymbol{I}$ a 2×2 identity matrix. Then, from (7), we have

$$\mathrm{E}(z_j^{(k)})=\mathrm{E}(y_j)-r_k e^{iq_k}\,\mathrm{E}(x_j^{(k)})\,, \quad (18a)$$

$$\boldsymbol{Cov}(z_j^{(k)})=\boldsymbol{Cov}(y_j)-r_k^2\boldsymbol{R}^{(k)}\boldsymbol{Cov}(x_j^{(k)})(\boldsymbol{R}^{(k)})^T\,. \quad (18b)$$

Finally, we can generate

$$\mathrm{E}(\hat{z}_j^{(k)})=e^{-iq_k}\,\mathrm{E}(z_j^{(k)})\,, \quad (19a)$$

$$\boldsymbol{Cov}(\hat{z}_j^{(k)})=(\boldsymbol{R}^{(k)})^T\boldsymbol{Cov}(z_j^{(k)})\boldsymbol{R}^{(k)}\,. \quad (19b)$$

8 Appendix B

The key to evaluate $\mathrm{E}(\hat{x}_j^{(k)})$ and $\boldsymbol{Cov}(\hat{x}_j^{(k)})$ is to find $\mathrm{E}(z_j)$ and $\boldsymbol{Cov}(z_j)$. We treat z_j and z_j in (12) as independent variables. From (11) and (12), we can express x_j as

$$x_j = r_k e^{i\varphi_k} x_j^{(k)} + z_j^{(k)} - w_j \,. \tag{20}$$

Then $m_j \equiv \mathrm{E}(x_j)$ and $\boldsymbol{V}_j \equiv \boldsymbol{Cov}(x_j)$ can be obtained from (20) if $\mathrm{E}(x_j^{(k)})$ and $\boldsymbol{Cov}(x_j^{(k)})$ are given and $\mathrm{E}(z_j^{(k)})$ and $\boldsymbol{Cov}(z_j^{(k)})$ are available (see Appendix A). We treat x_j as a complex Gaussian random variable. Then $\mathrm{E}(z_j)$ and $\boldsymbol{Cov}(z_j)$ are fully determined by (μ_j, $\boldsymbol{V}_j$) from (13). We denote these two relationships using two functions below:

$$\mathrm{E}(z_j) = f(m_j, \boldsymbol{V}_j)\,, \tag{21a}$$

$$\boldsymbol{Cov}(z_j) = g(m_j, \boldsymbol{V}_j)\,. \tag{21b}$$

In general, the two functions in (21) can be generated numerically using the Monte Carlo method. We can create two look-up tables to characterize them.

9 Appendix C

Recalling (11) and assuming that A is known at the receiver, we can derive the following simple HCA:

(a) Initialization:

Set the estimation of z_j as $\hat{z}_j = 0$, $\forall\, j$.

(b) Main iteration:

i) Taking $\{\hat{y}_j = y_j - \hat{z}_j\}$ as the received sequence, generate a hard estimation $\tilde{\boldsymbol{u}}$ of $\boldsymbol{u}$ based on (10).

ii) Taking $\tilde{\boldsymbol{u}}$ as the data sequence, construct an estimate $\{\hat{x}_j\}$ of $\{x_j\}$ using the encoder shown in Fig. 1.

iii) Compute $\{\hat{z}_j\}$ using (13), and then go back to step i.

10 Acknowledgment

The authors are grateful to Prof. Xiao Ma for enlightening discussions and for help in developing the software used in this work.

11 References

[1] Ungerboeck, G.: Channel coding with multilevel/ phase signals. *IEEE Trans. Inform. Theory*, vol. IT−28, Jan. 1982, pp. 55−67

[2] Imai, H.; Hirakawa, S.: A new multilevel coding method using error-correcting codes. *IEEE Trans. Inform. Theory*, vol. IT−23, May 1977, pp. 371− 377

[3] Li, P.; Bai, B.; Wang, X.: Low complexity concatenated two-state TCM schemes with near capacity performance. *IEEE Trans. Inform. Theory*, vol.49, no.2, Dec. 2003. pp. 3225−3234

[4] Wachsmann, U.; Fischer, R. ; Huber, J.: Multilevel codes: theoretical concepts and practical design rules. *IEEE Trans. Inform. Theory*, vol. 45, no.5, July 1999, pp. 1361−1391

[5] Limpaphayom, P.; Winich, K.: Power- and bandwidth-efficient communications using LDPC codes. *IEEE Trans. Commun.*, vol. 52, no. 3, Mar. 2004, pp. 350−354

[6] Varnica, N.; Ma, X.; Kavčić, A.: Iteratively decodable codes for bridging the shaping gap in communication channels. *Proc. Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, CA, Nov. 2002

[7] Forney, G.: Trellis shaping. *IEEE Trans. Inform. Theory*, vol. 38, no. 2, Mar. 1992, pp. 281−300

[8] Goff., S.; Sharif, B.; Jimaa, S.: Bit-interleaved turbo-coded modulation using shaping coding. *IEEE Commun. Lett*., vol. 9, no. 3, Mar. 2005, pp. 246–248

[9] Fischer, R. *Precoding and Signal Shaping for Digital Transmission*. New York: Wiley, 2002

[10] Ma, X.; Li, P.: Coded modulation using superimposed binary codes. *IEEE Trans. Inform. Theory*, vol. 50, no. 12, Dec. 2004, pp. 3331−3343

[11] Ma, X.; Li, P.: Power allocations for multilevel coding with sigma mapping. *Electron. Lett*., vol. 40, no. 10, May 2004, pp. 609−611

[12] Cover, T.; Thomas, J.: *Elements of Information Theory*. New York: Wiley, 1991

[13] Tse, D.; Viswanath, P.: *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005

[14] Kim, D.; Stüber, G.:Clipping noise mitigation for OFDM by decision–aided reconstruction, *IEEE Commun. Lett*., vol. 3, no.1, Jan. 1999, pp. 4–6

[15] Chen, H.; Haimovich, A.: Iterative estimation and cancellation of clipping noise for OFDM signals, *IEEE Commun. Lett*., vol. 7, no. 7, July 2003, pp. 305−307

[16] Gelle, G.; Colas, M.; Declercq, D.: Turbo decision aided reconstruction of clipping noise in coded OFDM. *Proc. IEEE SPAWC 2004*, 2004

[17] Li, P.; Liu, L.; Wu, K.; Leung, W.: Interleave-division multiple-access. Accepted by *IEEE Trans. Wireless Commun*., Available at: www.cityu.edu.hk/~liping

[18] Liu, L.; Tong, J.; Li, P.: Analysis and optimization of CDMA systems with chip-level interleavers. *IEEE J. Select. Areas Commun.*, vol. 24, no. 1, Jan. 2006, pp. 141−150

[19] Schoeneich, H.; Hoeher, P.: Adaptive interleave-division multiple access-A potential air interference for 4G bearer services and wireless LANs. *Proc. WOCN 2004,* Muscat, Oman, June 2004, pp. 179–182

[20] Brink, S.: Rate one-half code for approaching the Shannon limit by 0.1 dB. *Electron. Lett.*, vol. 36, no. 15, July 2000, pp. 1293−1294