

# Superposition Coded Modulation With Peak-Power Limitation

Jun Tong, *Student Member, IEEE*, Li Ping, *Senior Member, IEEE*, and Xiao Ma

**Abstract**—We apply clipping to superposition coded modulation (SCM) systems to reduce the peak-to-average power ratio (PAPR) of the transmitted signal. The impact on performance is investigated by evaluating the mutual information driven by the induced peak-power-limited input signals. It is shown that the rate loss is marginal for moderate clipping thresholds if optimal encoding/decoding is used. This fact is confirmed in examples where capacity-approaching component codes are used together with the maximum *a posteriori* probability (MAP) detection. In order to reduce the detection complexity of SCM with a large number of layers, we develop a suboptimal soft compensation (SC) method that is combined with soft-input soft-output (SISO) decoding algorithms in an iterative manner. A variety of simulation results for additive white Gaussian noise (AWGN) and fading channels are presented. It is shown that with the proposed method, the effect of clipping can be efficiently compensated and a good tradeoff between PAPR and bit-error rate (BER) can be achieved. Comparisons with other coded modulation schemes demonstrate that SCM offers significant advantages for high-rate transmissions over fading channels.

**Index Terms**—Capacity, clipping, iterative decoding, peak-to-average power ratio (PAPR), soft compensation, superposition coded modulation (SCM).

## I. INTRODUCTION

TRADITIONAL trellis-coded modulation (TCM) [1] is based on uniformly spaced constellations with equal probability for every signaling point. In an additive white Gaussian noise (AWGN) channel, there is an asymptotic 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 a Gaussian distribution) can be applied using shaping techniques, e.g., by assigning nonuniform probabilities on different signaling points [4]–[8]. The resulting advantage is referred to as the shaping gain [5]. Usually, special shaping codes and algorithms are needed.

Manuscript received October 28, 2005; revised March 05, 2009. Current version published May 20, 2009. This work was supported by a grant from the Research Grant Council of the Hong Kong Special Administrative Region, China, under project CityU 117305. The work of X. Ma was supported by the National Natural Science Foundation of China (NSFC 60572059). The material in this paper was presented in part at the 2006 IEEE International Conference on Communications, Istanbul, Turkey, June 2006.

J. Tong and L. Ping are with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong SAR (e-mail: jun.tong@student.cityu.edu.hk; eeliping@cityu.edu.hk).

X. Ma is with the Department of Electronics and Communication Engineering, Sun Yat-sen University, Guangzhou 510275, China (e-mail: maxiao@mail.sysu.edu.cn).

Communicated by M. P. Fossorier, Associate Editor for Coding Techniques. Digital Object Identifier 10.1109/TIT.2009.2018224

Recently, superposition coded modulation (SCM) has been studied as an alternative approach to other bandwidth-efficient coded modulation techniques [9], [10]. With SCM, several coded sequences (each referred to as a layer) are linearly superimposed before transmission. Consequently, when the number of layers is large, the transmitted signal exhibits an approximate Gaussian distribution that matches to an AWGN channel. This provides a more straightforward approach for achieving shaping gain [9], [11]. The work in [9]–[12] shows that such a concept is realizable with practical encoding and decoding methods. Simulation results show that an SCM scheme can operate within the shaping gap over AWGN channels [11], surpassing the theoretical limit of the uniform signaling based methods. (In this paper, it will be shown that SCM also provides a simple and effective means of high-rate transmission over fading channels. By using low-rate component codes, significant diversity gains can be achieved with SCM.)

SCM also finds use in many other contexts, e.g., in the achievability proof of multi-user channel capacity in [13]–[15]. In [10], SCM has also been studied for practical broadcasting channel applications, where it is shown that SCM can provide a significant gain over traditional time-division schemes. Another application is adaptive modulation through adjusting the number of layers (and so rate) according to channel condition [16]. This is more flexible than traditional approaches, such as switching among, say, TCM using 8-PSK (8-ary phase shift keying), 16-QAM (16-ary quadrature amplitude modulation), 32-QAM, etc., for channel adaptation [17]. The latter has the drawbacks of abrupt rate change and high receiver cost due to the need of many different TCM decoders. With SCM, rate change can be achieved smoothly by using a low-rate code for each layer. The receiver cost can be kept low by using the same code for all the layers and time-sharing a common decoder.

However, there is a practical concern with SCM: the Gaussian-like transmitted signal has a relatively high peak-to-average power ratio (PAPR), which may cause a problem for radio frequency amplifier efficiency [18]. The same PAPR problem also exists in other shaped coded modulation schemes [5]–[8] and orthogonal frequency-division multiplexing (OFDM) systems. For OFDM systems, a number of PAPR reduction techniques have been studied (see [18] and references therein). Among these techniques, clipping is the most straightforward but may lead to substantial degradation in the bit-error-rate (BER) performance [18]–[27], especially for high-rate applications.

In this paper, we investigate the use of clipping in SCM schemes. We show by mutual information analysis that the theoretical penalty due to clipping is marginal for practical PAPR values. (Actually, with the same PAPR, the capacity

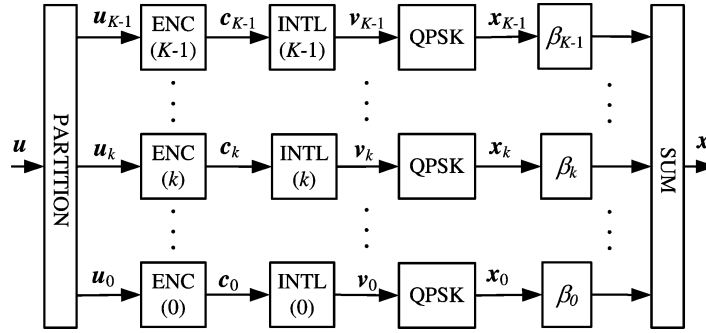


Fig. 1. Encoder of a superposition coded modulation system.

of clipped SCM signaling is higher than uniform signaling in low-to-medium rate region, see Fig. 4 later.) This is also verified by simulation results.

We devise a practical soft compensation method to recover the performance loss due to clipping by exploiting the characteristics of clipping noise. This method can provide (at much lower complexity) performance close to the optimal maximum *a posteriori* probability (MAP) method. It can be easily incorporated into the overall iterative receiver structure based on the low-cost multiuser detection principles developed in [29], [30]. A variety of numerical results for AWGN and fading channels are provided. It is shown that the proposed method can effectively recover the performance loss incurred by clipping.

We also consider the clipping issue for SCM with transform-domain transmission (e.g., OFDM). The soft compensation technique is extended to such cases and we provide performance comparisons between SCM and the well-known bit-interleaved coded modulation (BICM) scheme [31]–[36]. We show that when rate is relatively low, the two schemes have similar performance. However, as the rate increases, the advantage of SCM becomes more significant. In particular, when an underlying OFDM layer is involved, SCM can offer a better tradeoff between PAPR and BER at high rates than alternative schemes.

This paper is organized as follows: Section II introduces the basic system model. Section III contains an information-theoretic analysis of the clipping effect. Iterative decoding techniques are developed in Section IV. Section V addresses the clipping issue for SCM with transform-domain transmission. Comparisons of SCM and BICM are presented in Section VI. Finally, we summarize our main results in Section VII.

## II. SYSTEM MODEL

### A. Encoding

We consider a  $K$ -layer SCM system. The encoding scheme is shown in Fig. 1. A binary data sequence  $\mathbf{u}$  is partitioned into  $K$  subsequences  $\{\mathbf{u}_k\}$ . The  $k$ th subsequence  $\mathbf{u}_k$  is encoded by a binary encoder (ENC- $k$ ) at the  $k$ th layer, resulting in a coded bit sequence  $\mathbf{c}_k = \{c_k(j)\}$  of length  $2J$ , where  $c_k(j) \in \{0, 1\}$  and  $J$  is the frame length. The randomly interleaved version  $\mathbf{v}_k$  of  $\mathbf{c}_k$ , from interleaver- $k$  (INTL- $k$ ), is then mapped to a quadrature phase shift keying (QPSK) sequence  $x_k(j) = x_k^{Re}(j) + ix_k^{Im}(j)$ , where  $i = \sqrt{-1}$ , the superscripts “*Re*” and “*Im*” are used to denote the real and imaginary parts of

complex numbers, respectively,  $x_k^{Re}(j) = 1 - 2v_k(2j)$  and  $x_k^{Im}(j) = 1 - 2v_k(2j + 1)$ . It is clear that  $x_k^{Re}(j) \in \{+1, -1\}$  and so does  $x_k^{Im}(j)$ .

The output signal at time  $j$  is a linear superposition of  $K$  independently coded symbols

$$x(j) = \sum_{k=0}^{K-1} \beta_k x_k(j), \quad j = 0, 1, \dots, J-1 \quad (1)$$

where  $\{\beta_k\}$  are constant weighting factors. The overall rate is  $R = 2 \sum_{k=0}^{K-1} R_k$  in bits/symbol, where  $R_k$  is the rate of the  $k$ th binary component code. The selection of  $\{\beta_k\}$  will be discussed in Section II-B.

### B. Peak-to-Average Power Ratio

Let  $E[\cdot]$  denote the mathematical expectation and  $|\cdot|$  the amplitude. The PAPR (in decibel) of  $x(j)$  is defined as

$$\text{PAPR} = 10 \log_{10} \left( \frac{\max\{|x(j)|^2\}}{E[|x(j)|^2]} \right). \quad (2)$$

We assume that all the interleaved coded bits  $\{v_k(j)\}$  are independent and identically distributed (i.i.d.) random variables with  $\Pr(v_k(j) = 0) = \Pr(v_k(j) = 1) = 1/2$ . The PAPR can be very high when  $K$  is large. For an SCM scheme with  $\{\beta_0 = \dots = \beta_{K-1}\}$ , the PAPR is  $K$ . In order to suppress PAPR, we can clip  $x(j)$  to  $\bar{x}(j)$  before transmission according to the following rule:

$$\bar{x}(j) = \begin{cases} x(j), & |x(j)| < A \\ Ax(j)/|x(j)|, & |x(j)| \geq A \end{cases} \quad (3)$$

where  $A > 0$  is the clipping threshold. We define the clipping ratio (CR,  $\gamma$ ) in decibel as

$$\gamma = 10 \log_{10} \left( \frac{A^2}{E[|x(j)|^2]} \right). \quad (4)$$

The PAPR of the transmitted signal is given by  $\text{PAPR} = 10 \log_{10} (A^2/E[|\bar{x}(j)|^2])$ . We select  $A$  according to the desired PAPR value.

The performance of an SCM scheme can be improved by properly choosing the power allocation factors  $\{|\beta_k|^2\}$ . In this paper,  $\{|\beta_k|^2\}$  are determined using the simulation-based power allocation method [11], [12] for small  $K$  and the linear programming method [29] for large  $K$ . The phase angles  $\{\angle \beta_k\}$  can be used to shape the signal constellations and adjust PAPR. For example, for a 4-layer SCM with  $\{|\beta_k| = 1, \forall k\}$ ,

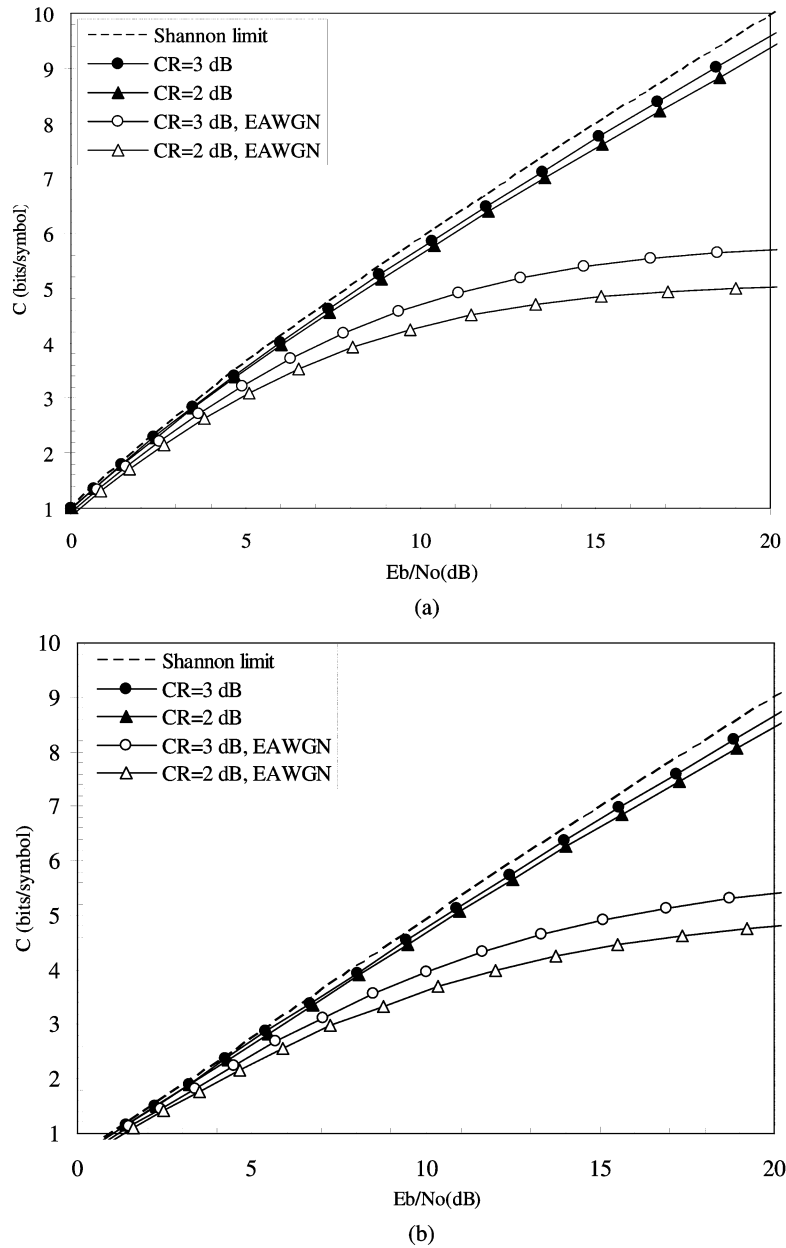


Fig. 2. Capacities of clipped Gaussian input signals over (a) AWGN and (b) Rayleigh fading channels.

the maximum PAPR = 6.02 dB is reached at  $\{\angle\beta_k = 0, \forall k\}$ . On the other hand, if we set  $\{\angle\beta_k = k\pi/8, \forall k\}$ , then the PAPR is reduced to 5.16 dB.

### C. Received Signal Model

The clipped signal  $\bar{x}(j)$  is then transmitted over a memoryless channel. The received signal is given by

$$y(j) = h(j)\bar{x}(j) + w(j), \quad j = 0, 1, \dots, J-1 \quad (5)$$

where  $h(j)$  is the channel coefficient and  $w(j)$  is a complex, zero-mean white Gaussian noise with variance  $\sigma^2$  per dimension. The ratio of energy per bit ( $E_b$ ) to the noise power spectral density ( $N_0 = 2\sigma^2$ ) is given by  $E_b/N_0 = E[|\bar{x}(j)|^2]/(2R\sigma^2)$ .

When  $K$  is large,  $x(j)$  can be approximated by a Gaussian random variable from the central limit theorem. Using Price's

theorem for nonlinear systems with Gaussian inputs [28], we can model the clipping operation in (3) as a linear process

$$\bar{x}(j) = \alpha x(j) + d(j) \quad (6)$$

where  $\alpha$  is a constant attenuation factor, and  $d(j)$  is a Gaussian-distributed distortion term with zero-mean and variance  $\sigma_d^2$  per dimension, which is statistically uncorrelated with  $x(j)$ . In general,  $\alpha$  and  $\sigma_d^2$  depend on  $\gamma$  and the statistics of  $x(j)$ , and can be calculated as [20], [25]

$$\alpha = \frac{E[x^*(j)\bar{x}(j)]}{E[|x(j)|^2]} \quad (7)$$

$$\sigma_d^2 = \frac{E[|\bar{x}(j)|^2] - \alpha^2 E[|x(j)|^2]}{2} \quad (8)$$

where  $*$  denotes complex conjugate. Then the received signal can be alternatively written as

$$y(j) = \alpha h(j)x(j) + h(j)d(j) + w(j). \quad (9)$$

The above modeling will be used in Section IV-E.

### III. EFFECT OF CLIPPING ON THE ACHIEVABLE RATES

We now investigate the impact of clipping on the performance limits of SCM systems. Consider a memoryless channel characterized by  $Y = HX + W$ , where  $H$ ,  $X$  and  $W$  are, respectively, the channel coefficient, transmitted signal and AWGN. Assume that the receiver has perfect knowledge of  $H$ . Then for a given distribution of  $X$ , the capacity is quantified by the average mutual information [13]

$$\begin{aligned} C &= E_H [I(X; Y|H)] \\ &= E_H [-E_Y [\log(p(Y|H))] - \log(2\pi e\sigma^2)] \end{aligned} \quad (10)$$

where  $E_H[\cdot]$  and  $E_Y[\cdot]$  denote expectation with respect to  $H$  and  $Y$ , respectively. We can apply numerical methods to evaluate  $C$ . We examine the cases of continuous and discrete  $X$  separately below.

#### A. Continuous Input Signal

We first consider  $X$  as a clipped version of a complex, Gaussian random variable with zero mean and variance  $1/2$  per dimension. The clipping rule given by (3) is used. The probability density function of  $X$  is given by

$$p(x) = \begin{cases} \exp(-|x|^2)/\pi, & |x| < A \\ \exp(-|A|^2)\delta(|x| - A)/(2\pi A), & |x| = A \end{cases} \quad (11)$$

where  $\delta(\cdot)$  is the Dirac delta function.

The numerical results based on (10) for the clipped Gaussian signaling and the Shannon limit are shown in Fig. 2. We set  $\gamma = 3$  and  $2$  dB, and the resultant PAPR is 3.64 and 3 dB, respectively. The performance of the clipped Gaussian signaling is close to the Shannon limit if optimal decoding is used. We have also included in Fig. 2 the results (marked by ‘‘EAWGN’’) for a suboptimal strategy [20] based on (9), in which the clipping distortion  $d(j)$  is approximated by an equivalent AWGN. It is seen from Fig. 2 that this suboptimal approach leads to significant performance loss. This motivates us to develop improved techniques.

#### B. Discrete Input Signal

Next we examine the SCM systems where  $X$  is a discrete variable. For the  $K$ -layer scheme described in Fig. 1, the input signal before clipping is the summation of  $K$  complex random variables. When  $K$  is large, the distribution of the unclipped signal is approximately complex Gaussian. Hence, the above analysis on clipped Gaussian signaling can provide insights into such cases.

We now focus on the impact of clipping on SCM schemes with a small-to-medium  $K$ . We take a five-layer scheme as an example. It employs a nonequispaced  $4^5 = 1024$ -ary signal constellation that is fully determined by  $\{\beta_k\}$ , as depicted in Fig. 3. The related PAPR without clipping is 5.39 dB. We set

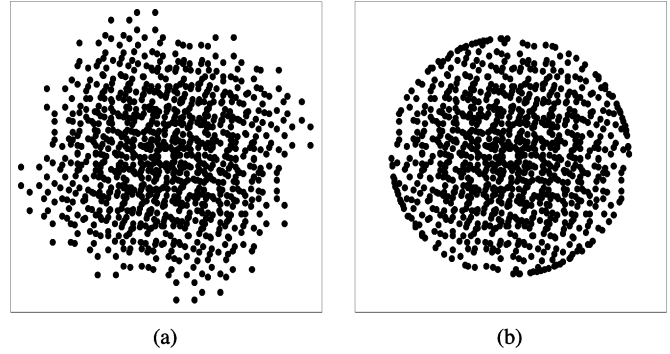


Fig. 3. SCM constellations (a) without and (b) with clipping.  $K = 5$ ,  $\{|\beta_k|\} = \{1, 1.4565, 2.1218, 3.0912, 4.5031\}$  and  $\{\angle\beta_k = k\pi/10, \forall k\}$ .

$\gamma = 3.5$  dB, and the resultant PAPR = 3.68 dB. The capacities achieved over AWGN and Rayleigh fading channels are shown in Fig. 4. The performance of the conventional 1024-QAM<sup>1</sup> signaling (PAPR  $\approx 4.5$  dB) and the Shannon limit are also included for comparison. From Fig. 4, we can make the following observations:

- Without clipping, the differences between the SCM capacity, 1024-QAM capacity and Shannon limit are not significant for low-to-medium rates (e.g., up to about 9 bits/symbol). The SCM capacity is also higher than the 1024-QAM capacity for the most part in this region. For a target rate of 5 bits/symbol, the difference between the required  $E_b/N_0$  of SCM and the Shannon limit is only 0.21 dB for AWGN channels, representing a shaping gain of about 0.7 dB over the 1024-QAM signaling.
- Although clipping degrades the achievable rate, the effect is not serious if optimal decoding is applied. We see that in AWGN channels, at a rate of 5 bits/symbol, about 0.2 dB loss in  $E_b/N_0$  is introduced by clipping at PAPR = 3.68 dB. The clipped SCM, when compared with the conventional 1024-QAM, has lower PAPR but higher capacity in the vicinity of 5 bits/symbol. This reveals that the clipped SCM scheme can provide a good trade-off between PAPR and achievable rate.

### IV. ITERATIVE DECODING

Now we turn our attention to practical SCM systems. For simplicity, we only discuss AWGN channels with  $h(j) = 1, \forall j$ , and assume that the weighting factors  $\{\beta_k\}$  are real numbers. The discussions below can be easily extended to more general cases.

In Section III-A, we have shown that, from the point of view of capacity, the performance penalty incurred by clipping is not severe for a reasonable clipping ratio. But, as seen from the ‘‘EAWGN’’ curves in Fig. 2, clipping can cause serious problems if treated improperly.

The following observations suggest a possible approach to this issue.

- If  $|y(j)|$ , the amplitude of the received signal, is large, then the clipping probability (i.e., the probability of  $x(j)$  being clipped) is high.

<sup>1</sup>In this paper, a ‘‘conventional QAM’’ constellation represents a square QAM constellation in which signal points are equispaced and utilized with equal probabilities [5].

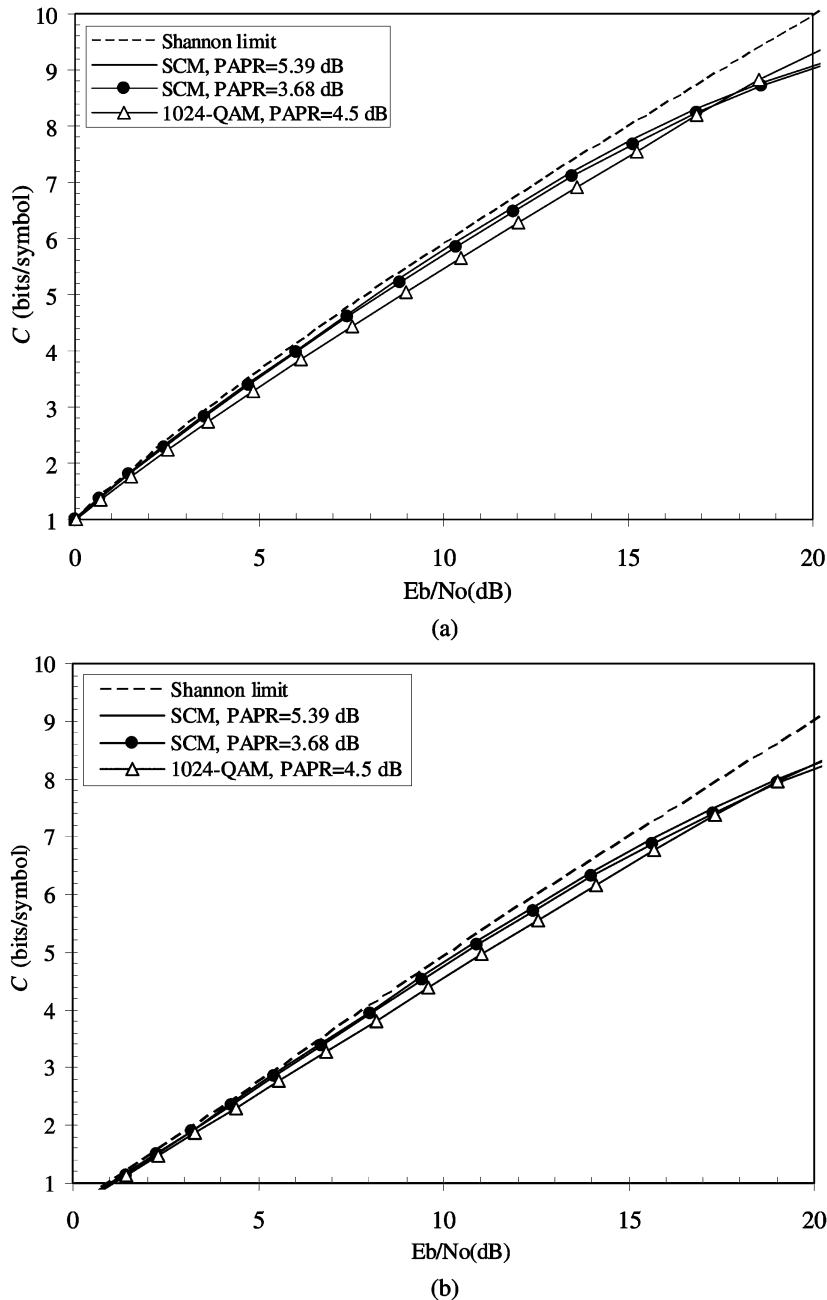


Fig. 4. Capacities of clipped SCM schemes over (a) AWGN and (b) Rayleigh fading channels.

- If  $|y(j)|$  is small, then the clipping probability is small. We may exploit these facts to compensate for the clipping effect. This is the underlying rationale for the soft compensation method presented here.

#### A. Overall Iterative Detection Principle

We first outline the basic receiver structure. SCM 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 SCM systems and the interleave-division multiple-access systems, we employ a suboptimal iterative receiver similar to that in [29] and [30].

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 (deinterleavers), operating iteratively. The messages passing between the ESE and the DECs are the so-called *extrinsic* information values. The turbo-type iterative process basically follows the discussion in [11] and [29] with the ESE outputs used as the DEC inputs and vice versa. We will not discuss the DEC function in detail since it only involves standard *a posteriori* probability (APP) decoding. We will focus on the ESE that handles the interlayer interference, as well as the clipping distortion.

#### B. Optimal Realization of the ESE

The function of the ESE is to estimate the transmitted signals  $\{x_k(j)\}$  ignoring the coding constraint (i.e., as if  $\{x_k(j)\}$  is an

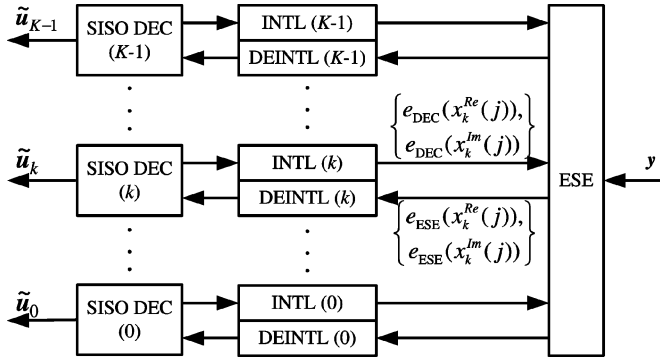


Fig. 5. Block diagram of the iterative decoding/detection algorithm.

uncoded sequence). The optimal MAP principle can be applied here to compute the following log-likelihood ratio (LLR) for the coded bit  $x_k^{Re}(j)$  (we handle  $x_k^{Im}(j)$  in a similar way)

$$e_{ESE}(x_k^{Re}(j)) = \ln \left( \frac{\Pr(x_k^{Re}(j) = +1|y(j))}{\Pr(x_k^{Re}(j) = -1|y(j))} \right) - e_a(x_k^{Re}(j)) \quad (12)$$

where  $e_a(x) = \ln(\Pr(x = +1)/\Pr(x = -1))$  is the a priori LLR for  $x$ . During iterative decoding,  $e_a(x)$  is approximated by the feedback LLRs from the DECs. See Fig. 5 and [29]. Given the signal constellation (e.g., Fig. 3) at the transmitter, (12) can be evaluated following the standard procedure [33], [37]. The related complexity is proportional to the constellation size.

The constellation size of the SCM (with and without clipping) grows exponentially with the number of layers  $K$ , which can be a serious concern for large  $K$ . In the following, we present several suboptimal, low-cost alternatives.

### C. Suboptimal Realization of the ESE Based on Gaussian Approximation

We first consider unclipped SCM. The following Gaussian approximation (GA) detector forms the basis for all the suboptimal methods discussed in this section. Let us concentrate on layer- $k$ . Rewrite (5) as

$$y(j) = x(j) + w(j) = \beta_k x_k(j) + \zeta_k(j) \quad (13)$$

where

$$\zeta_k(j) = \sum_{k' \neq k} \beta_{k'} x_{k'}(j) + w(j) \quad (14)$$

represents the interference-plus-noise with respect to  $x_k(j)$ . We approximate  $\zeta_k(j)$  by an additive complex Gaussian variable. Then we can compute the *extrinsic* LLRs by (15) below (with much lower complexity than the optimal MAP method)

$$\begin{aligned} e_{ESE}(x_k^{Re}(j)) &= \ln \left( \frac{(2\pi V[\zeta_k^{Re}(j)])^{-1/2} \exp \left\{ -\frac{(y^{Re}(j) - \beta_k - E[\zeta_k^{Re}(j)])^2}{2V[\zeta_k^{Re}(j)]} \right\}}{(2\pi V[\zeta_k^{Re}(j)])^{-1/2} \exp \left\{ -\frac{(y^{Re}(j) + \beta_k - E[\zeta_k^{Re}(j)])^2}{2V[\zeta_k^{Re}(j)]} \right\}} \right) \\ &= 2\beta_k \frac{y^{Re}(j) - E[\zeta_k^{Re}(j)]}{V[\zeta_k^{Re}(j)]} \end{aligned} \quad (15)$$

where  $V[\cdot]$  denotes the variance function and we have assumed that  $\{\beta_k\}$  are real numbers. The computational details for  $E[\zeta_k^{Re}(j)]$  and  $V[\zeta_k^{Re}(j)]$ , as well as the treatments for complex  $\{\beta_k\}$ , are given in Appendix.

### D. SNR Evolution

The following signal-to-noise ratio (SNR) evolution technique [29] can be used to evaluate the performance of the GA detector. It also provides insight into the convergence properties of the iterative decoding. Denote by  $\{\text{snr}_k^{(q)}, k = 0, 1, \dots, K-1\}$  and  $\{v_k^{(q)}, k = 0, 1, \dots, K-1\}$ , respectively the average SNR contained in the ESE outputs and the average variance at the output of the DECs, both in the  $q$ th iteration. Let  $Q$  be the number of iterations between the ESE and the DECs. It can be shown that, during the iterative process, these two groups of quantities determine each other recursively as follows.

**Initialization:** Set  $\{v_k^{(0)} = 1, k = 0, 1, \dots, K-1\}$

**Recursion:** For  $q = 1, 2, \dots, Q$ , compute

$$\text{snr}_k^{(q)} = \frac{|\beta_k|^2}{\sum_{k' \neq k} |\beta_{k'}|^2 v_{k'}^{(q-1)} + \sigma^2}, \quad k = 0, 1, \dots, K-1 \quad (16)$$

$$v_k^{(q)} = f_k(\text{snr}_k^{(q)}), \quad k = 0, 1, \dots, K-1. \quad (17)$$

An intuitive explanation of (16) is that  $|\beta_k|^2$  and  $\sum_{k' \neq k} |\beta_{k'}|^2 v_{k'}^{(q-1)} + \sigma^2$ , respectively, represents the received power and uncertainty with respect to the  $k$ th layer. The uncertainty is caused by the joint effect of the interfering signals from other layers and the channel noise, with the contribution from the  $k$ th layer measured by the variance at the output of DEC- $k$  scaled by a power factor  $|\beta_k|^2$ . The function  $f_k(\cdot)$  in (17) characterizes the DEC operation and can be obtained using the Monte Carlo method (similar to the treatment of an *extrinsic* information transfer (EXIT) function [38]). We also define the BER performance of DEC- $k$  as a function of  $\text{snr}_k^{(q)}$  as

$$\text{BER}_k = g_k(\text{snr}_k^{(q)}). \quad (18)$$

The BER performance can be estimated by substituting  $\text{snr}_k^{(Q)}$  from the final step in (16) into  $g_k(\cdot)$ . We show later that this semi-analytic SNR evolution technique provides a fast and reasonably accurate way to predict the performance of the GA detection procedure outlined above. Furthermore, this technique can be used to optimize the weighting factors  $\{\beta_k\}$  following the procedures developed in [29].

### E. Modified GA Method (MGA)

We now proceed to consider clipped SCM systems. In this case, the basic GA method in Section IV-C deteriorates if the clipping distortion is ignored. Based on the approximation in (9), an improved method is to treat the clipping distortion  $d(j)$  as an equivalent AWGN. Again, we focus on layer- $k$ . Rewrite (9) as

$$y(j) = \alpha \beta_k x_k(j) + \tilde{\zeta}_k(j) \quad (19)$$

with

$$\tilde{\zeta}_k(j) = \alpha \sum_{k' \neq k} \beta_{k'} x_{k'}(j) + d(j) + w(j). \quad (20)$$

Now the GA method can be applied. The performance of this modified GA (MGA) method is still not satisfactory, but it can be used in the initial stage for our simulation results presented below, because it can start without accurate feedback information.

#### F. Soft Compensation

Next, we derive a soft compensation (SC) technique that uses a joint process to treat the interlayer interference and clipping noise. We rewrite (5) as

$$\begin{aligned} y(j) &= x(j) + w(j) + z(j) \\ &= \beta_k x_k(j) + \zeta_k(j) + z(j) \end{aligned} \quad (21)$$

where  $\zeta_k(j)$  is given by (14) and  $z(j) = \bar{x}(j) - x(j)$  represents the clipping noise. Consider the detection of  $x_k(j)$  based on (21). We propose the following suboptimal SC method which has complexity only slightly higher than the GA method. We again approximate  $\zeta_k(j) + z(j)$  by an additive complex Gaussian variable. It is important to note that here the statistics of  $\zeta_k(j) + z(j)$  are different for different hypotheses on  $x_k(j)$ , since the clipping effect depends on the hypothesis. For simplicity, we omit the time index  $j$  from now on and rewrite (21) as  $y = \beta_k x_k + \zeta_k + z$ , and introduce the following notations.

- $E^+$  and  $E^-$  are, respectively, the means of  $\zeta_k^{Re} + z^{Re}$  under the hypothesis  $x_k^{Re} = +1$  and  $-1$ .
- $V^+$  and  $V^-$  are, respectively, the variances of  $\zeta_k^{Re} + z^{Re}$  under the hypothesis  $x_k^{Re} = +1$  and  $-1$ .

Similar to (15), we have

$$\begin{aligned} e_{\text{ESE}}(x_k^{Re}) &= \ln \left( \frac{(2\pi V^+)^{-1/2} \exp \left\{ -\frac{(y^{Re} - \beta_k - E^+)^2}{2V^+} \right\}}{(2\pi V^-)^{-1/2} \exp \left\{ -\frac{(y^{Re} + \beta_k - E^-)^2}{2V^-} \right\}} \right) \\ &= -\frac{1}{2} \ln \left( \frac{V^+}{V^-} \right) - \frac{(y^{Re} - \beta_k - E^+)^2}{2V^+} \\ &\quad + \frac{(y^{Re} + \beta_k - E^-)^2}{2V^-}. \end{aligned} \quad (22)$$

The conditional means  $E^\pm$  and variances  $V^\pm$  in (22) can be estimated using the method in Section IV-G. Notice that (22) is only slightly more complicated than (15).

#### G. Efficient Techniques to Evaluate the Conditional Means and Variances

We approximately treat  $\zeta_k$  and  $z$  in (21) as independent variables. Since  $\zeta_k$  is exactly the same as that in (14), the key is to find the statistics of  $z$ . From (3),  $z$  is given by

$$z = \begin{cases} 0, & |x| < A \\ Ax/|x| - x, & |x| \geq A. \end{cases} \quad (23)$$

If the conditional probability density function  $p(x|x_k^{Re} = \pm 1)$  is available, then  $E^\pm$  and  $V^\pm$  can be evaluated using numerical integration. However, this is difficult in practice due to the excessive computational cost. We propose the following suboptimal strategy based on Gaussian approximation. For clarity, we will omit the superscripts “ $\pm$ ” and the method below is applied to both the hypotheses  $x_k^{Re} = +1$  and  $-1$ . Let  $\mathbf{C}[x]$  be the covariance matrix of  $x$ , as defined in Appendix. From (21), we can express  $x$  as

$$x = \zeta_k - w + \beta_k x_k. \quad (24)$$

Then  $\mu \equiv \mathbb{E}[x]$  and  $\mathbf{V} \equiv \mathbf{C}[x]$  can be obtained from (24), if  $\mathbb{E}[x_k]$  and  $\mathbf{C}[x_k]$  are given and  $\mathbb{E}[\zeta_k]$  and  $\mathbf{C}[\zeta_k]$  are available. We now treat  $x$  as a complex Gaussian random variable. Then  $\mathbb{E}[z]$  and  $\mathbf{C}[z]$  are fully determined by  $(\mu, \mathbf{V})$  from (23). We denote these relationships using the two functions below:

$$\mathbb{E}[z] = \phi(\mu, \mathbf{V}) \quad (25a)$$

$$\mathbf{C}[z] = \varphi(\mu, \mathbf{V}). \quad (25b)$$

In general, the functions in (25) can be generated numerically using the Monte Carlo method. We can create two look-up tables to characterize them. Assuming that these two tables are available, then the SC cost is only slightly higher than the GA method in Section IV-C.

Since  $(\mu, \mathbf{V})$  involves five parameters, we need two five-dimensional (5-D) tables. We now consider an approximate technique to reduce memory cost using

$$\mathbf{V} \approx v\mathbf{I} \quad (26)$$

where  $v = (\mathbb{V}[x^{Re}] + \mathbb{V}[x^{Im}])/2$  and  $\mathbf{I}$  denotes the  $2 \times 2$  identity matrix. This is to approximately characterize  $x$  using a symmetric complex Gaussian distribution  $\mathcal{CN}(\mu, v\mathbf{I})$ . Now there are only three parameters involved. Furthermore, it is easily shown that

$$\phi(\mu, v\mathbf{I}) = \frac{\mu}{|\mu|} \phi(|\mu|, v\mathbf{I}) \quad (27a)$$

$$\varphi(\mu, v\mathbf{I}) = \Psi \varphi(|\mu|, v\mathbf{I}) \Psi^T \quad (27b)$$

where

$$\Psi = \frac{1}{|\mu|} \begin{bmatrix} \mu^{Re} & -\mu^{Im} \\ \mu^{Im} & \mu^{Re} \end{bmatrix}. \quad (28)$$

Therefore we only need two two-dimensional (2-D) tables to characterize  $\phi(|\mu|, v\mathbf{I})$  and  $\varphi(|\mu|, v\mathbf{I})$  and then use (27) to find  $\phi(\mu, v\mathbf{I})$  and  $\varphi(\mu, v\mathbf{I})$  for the SC method.

The SC method is essentially a turbo-type clipping noise cancellation technique based on the *extrinsic* information produced by the SISO decoders. This distinguishes it from the decision-aided clipping noise cancellation techniques in [25], [27] and the signal reconstruction techniques in [24], [26].

#### H. Examples

We now provide several examples of clipped SCM schemes. We will always assume that the interleavers are randomly generated. We observe that the MGA method based on the approxima-

TABLE I  
PARAMETERS OF THE SCM SCHEMES IN FIG. 6

$K$	$R$	$\{ \beta_k \}$	$\{\angle\beta_k\}$	PAPR(dB)
4	4	1, 1.4523, 2.1086, 3.0626	$k\pi/8, \forall k$	4.74
5	5	1, 1.4565, 2.1218, 3.0912, 4.5031	$k\pi/10, \forall k$	5.39

tion in (9) (see Section IV-E) provides improved performance at the starting stage of the iterative process. In this case, the feedback is not reliable and the SC method (see Section IV-F) is not effective. However, after a few iterations, the feedback from the DEC becomes reliable and the SC method becomes more effective. In the following, we adopt a hybrid strategy in which the MGA method is executed for  $Q_M$  iterations, followed by the SC procedure for  $Q_S$  iterations. The values of  $Q_M$  and  $Q_S$  are obtained experimentally.

1) *AWGN Channels*: We first consider AWGN channels. The rate-1/2 doped code introduced in [39] is chosen as the component code for each layer. We set  $K = 4$  and 5. The parameters and the PAPR without clipping are given in Table I. Clipping with  $\gamma = 3.5$  dB is applied to both systems to reduce the PAPR to about 3.68 dB. (Note that the conventional 64-QAM signaling has the same PAPR of 3.68 dB.) The maximum number of iterations in the component DECs [39] is set to 200, and the maximum number of iterations between the DECs and the ESE is set to 6. The entropy-based stopping criterion introduced in [11] is used to terminate the iterations.

The simulated BER performance with different detection methods is shown in Fig. 6(a) and (b) respectively for  $K = 4$  and 5. For comparison, the Shannon limit, the 1024-QAM capacity, and the capacity limits of the related SCM signaling are also included.

We can see that when clipping is not used, the GA method and the optimal MAP method perform similarly, and the achieved performance is quite close to the channel capacity. At BER of  $10^{-5}$ , the difference between the required  $E_b/N_0$  and the Shannon limit is only 0.9 and 1.2 dB for  $R = 4$  and 5 bits/symbol, respectively.<sup>2</sup> Also, we can see that the results predicted by SNR evolution are in good agreement with the simulation results.

On the other hand, from Fig. 6, the SCM performance deteriorates when clipping is used. With the optimal MAP detection, the loss is within 0.2 dB (at BER =  $10^{-5}$ ) for the two examples. This is roughly in line with the loss in capacity due to clipping.

For the suboptimal GA and MGA methods, the performance loss in comparison with the unclipped performance increases significantly, with the MGA method performing better. Using the SC method, however, the differences in performance are reduced to about 0.3 and 0.5 dB for  $K = 4$  and 5, respectively. This implies that the SC method can recover most of the performance loss due to clipping and hence provide an efficient solution.

The performance loss of the SC method due to the approximation in (26) is also shown in Fig. 6. Each dimension is quantized to 20 levels. We can see that the difference between the

<sup>2</sup>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 [7], [8]. However, the associated PAPRs in [7], [8] are relatively high (7.26 and 8.93 dB for  $R = 4$  and 5 bits/symbol, respectively).

5-D and 2-D methods is marginal (within 0.1 dB). For the latter, two small tables of size  $20 \times 20$  are used for  $\phi(\cdot, \cdot)$  and  $\varphi(\cdot, \cdot)$ .

Based on the above observations, we only consider the sub-optimal detection methods and use 2D tables for SC in the rest of this paper.

2) *Fading Channels*: Next we consider SCM over fully interleaved Rayleigh fading channels. The detection methods in Sections IV-A–G can be easily extended here. For simplicity, we only consider real  $\{\beta_k\}$ , and the component code is realized by the serial concatenation of a rate-1/2 nonsystematic convolutional code with generator polynomials  $(23, 35)_8$  and a length- $S$  repetition code. The repetition coding is introduced here to average out the fading effect. We fix the total rate  $R = 2$  bits/symbol (i.e.,  $K = 2S$ ) and the frame length  $J = 2048$ . Three schemes with different  $S$  as tabulated in Table II are compared.

From Fig. 7, we can see that the SCM performance in fading channels can be significantly improved by introducing repetition coding. A gain of about 2.8 dB at BER =  $10^{-5}$  can be achieved by increasing  $S$  from 1 to 4. Note that the receiver cost increases slightly with  $S$ , which can be seen as follows.

- The normalized cost (per information bit) of a convolutional decoder is independent of  $S$ .
- The cost related to the repetition code is negligible [29].
- The ESE cost grows linearly with  $S$ .
- The dominant factors of the receiver cost are usually related to the convolutional decoders.

In Fig. 7, the results are also shown for the clipped schemes with  $S = 4$  and 8 at a PAPR of 2.96 dB, the same as that of the unclipped scheme with  $S = 1$ . (The clipping ratios are  $\gamma = 2.1$  and 2.02 dB, respectively, for  $S = 4$  and 8.) In order to reduce the SC cost, we have adopted the following two strategies with marginal performance loss. First, since the number of layers is large, we assume that different hypotheses of the bit to be estimated (e.g.,  $x_k^{Re} = +1$  or  $-1$ ) have negligible effect on the estimation of clipping noise. Thus, we use *extrinsic* information only in evaluating (25). Secondly, we update the soft estimation of clipping noise only when all the *extrinsic* information from the DECs is renewed.

It is seen that the SC technique can reduce the error floor. The clipped schemes with SC outperform the unclipped scheme with  $S = 1$  by about 2 dB at BER =  $10^{-5}$ . Note that with  $S = 8$ , a large portion of the transmitted symbols are clipped and the PAPR is reduced by about 9 dB, but the performance loss is within 1 dB. This demonstrates that the SC technique can work effectively for deep clipping.

## V. SUPERPOSITION CODED MODULATION WITH TRANSFORM-DOMAIN TRANSMISSION

Define the output signal vector of an SCM encoder as

$$\mathbf{x} = [x(0), x(1), \dots, x(J-1)] \quad (29)$$

where  $x(j)$ ,  $j = 0, 1, \dots, J-1$ , is the coded symbol given by (1). We now consider the PAPR issue for  $\mathbf{X} = [X(0), X(1), \dots, X(J-1)] = \mathbf{F}(\mathbf{x})$ , where  $\mathbf{F}(\cdot)$  is a linear transform function with inverse  $\mathbf{F}^{-1}(\cdot)$ . The problem arises, e.g., when the SCM signal is transmitted using an

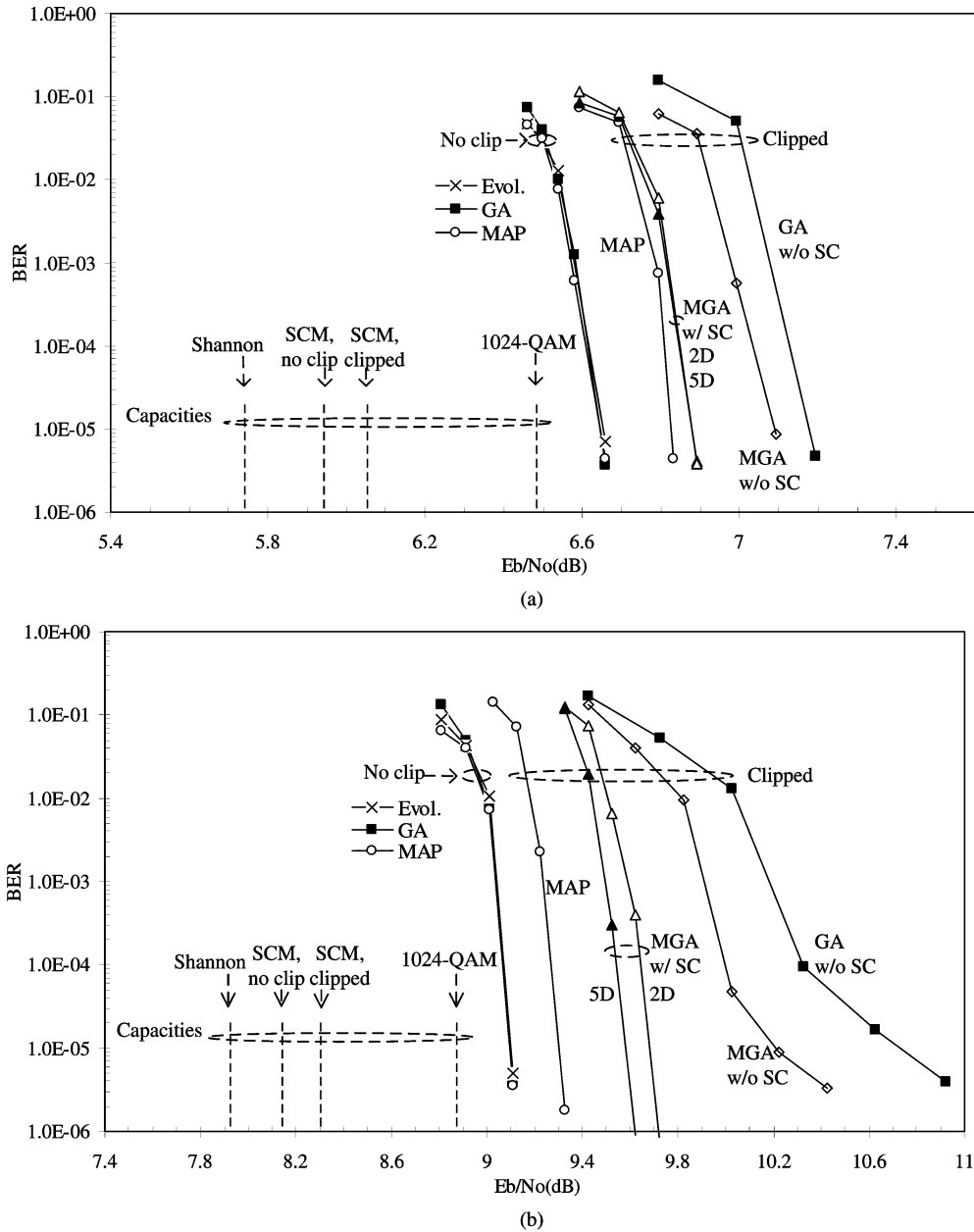


Fig. 6. Performance of the clipped SCM with a doped code and different detection methods over AWGN channels. (a)  $K = 4$ , (b)  $K = 5$ .  $J = 10^5$ . The number of iterations is 6. For curves marked by “MGA w/ SC”, the MGA method is first used for  $Q_M = 1$  iteration and then the SC method is used for  $Q_S = 5$  iterations.

TABLE II  
PARAMETERS OF THE SCM SCHEMES AT  $R = 2$  BITS/SYMBOL IN FIG. 7

$S$	$K$	$\{\beta_k \text{ value} \times \text{layer number}\}$	PAPR(dB)
1	2	$1 \times 1, 1.25 \times 1$	2.96
4	8	$1 \times 6, 1.44 \times 2$	8.90
8	16	$1 \times 12, 1.22 \times 4$	11.92

OFDM scheme where  $F(\cdot)$  represents the inverse discrete Fourier transform (IDFT).

The transmitted signal is obtained by clipping  $X(m)$ ,  $m = 0, 1, \dots, J - 1$ , to

$$\bar{X}(m) = \begin{cases} X(m), & |X(m)| < A \\ AX(m)/|X(m)|, & |X(m)| \geq A. \end{cases} \quad (30)$$

Let  $Z(m) = \bar{X}(m) - X(m)$ ,  $m = 0, 1, \dots, J - 1$ , and

$$\mathbf{z} = [z(0), z(1), \dots, z(J - 1)] = F^{-1}(\mathbf{Z}) \quad (31)$$

where  $\mathbf{Z}$  is defined similarly to (29). The received signal is first transformed back (using  $F^{-1}(\cdot)$ ) to the original domain, resulting in

$$y(j) = x(j) + w(j) + z(j), \quad j = 0, 1, \dots, J - 1 \quad (32)$$

where  $\{w(j)\}$  is the transform of the channel noise. Note that the counterparts of (30)–(32) in the original clipping system are (3), (23) and (21), respectively.

Eqns. (32) and (21) appear the same, but there is a subtle difference. In (21),  $z(j)$  is correlated with  $x(j)$  since a larger amplitude of  $x(j)$  implies a larger clipping probability. On the other hand, in (32),  $z(j)$  is the weighted sum of  $J$  clipping noise samples (see (31)) and is approximately independent of  $x(j)$ . Therefore, we will ignore the impact of the hypothesis

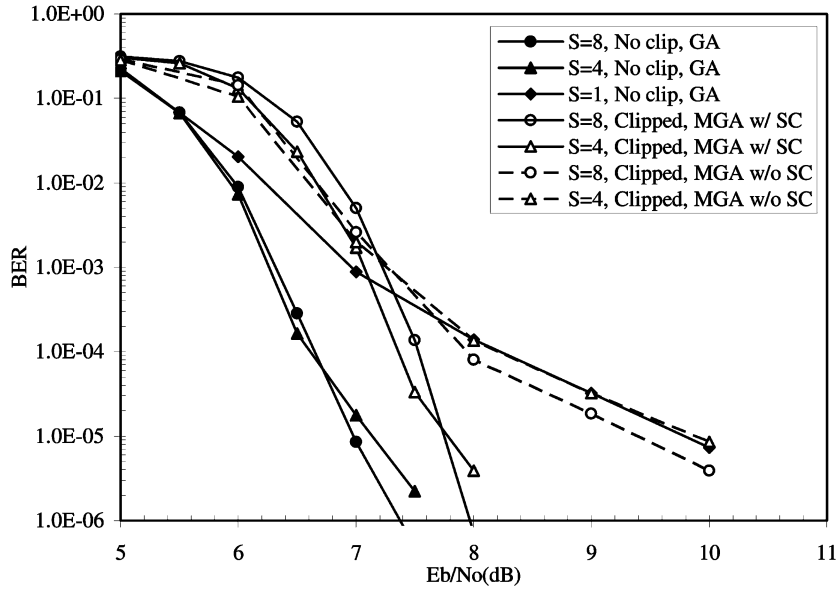


Fig. 7. Performance of SCM at  $R = 2$  bits/symbol over fully interleaved Rayleigh fading channels.  $J = 2048$ . The number of iterations is 10. For the clipped cases with SC, we set  $Q_M = 6$ , and  $Q_S = 4$ .

of  $\{x_k(j)\}$  in estimating clipping noise. We approximate  $z(j)$  by an additive Gaussian random variable that can be treated together with  $w(j)$  provided that  $E[z(j)]$  and  $C[z(j)]$  are known. The detection rule in (15) (instead of (22), since the influence of the hypothesis is ignored) can then be applied. We adopt the following steps to estimate  $\{E[z(j)]\}$  and  $\{C[z(j)]\}$ .

- 1) Estimate  $\{E[X(m)]\}$  and  $\{C[X(m)]\}$  from  $\{E[x(j)]\}$  and  $\{C[x(j)]\}$ .
- 2) Find  $\{E[Z(m)]\}$  and  $\{C[Z(m)]\}$  using the look-up table method in Section IV.
- 3) Generate  $\{E[z(j)]\}$  and  $\{C[z(j)]\}$  from the statistics of  $\{Z(m)\}$ .

Here,  $\{E[x(j)]\}$  and  $\{C[x(j)]\}$  are computed using the feedbacks from the DECs as in Appendix . Steps 1 and 3 can be performed based on  $\mathbf{X} = \mathbf{F}(\mathbf{x})$  and  $\mathbf{z} = \mathbf{F}^{-1}(\mathbf{Z})$ .

Note that with OFDM (a typical transform-domain transmission technique), other coded modulation methods also suffer from the PAPR problem. The SCM has a significant advantage in this case, since the SC method can be efficiently applied to SCM to alleviate the clipping effect. For other coded modulation techniques, compensation methods are not so effective [26].

## VI. COMPARISON WITH BICM

It is interesting to compare SCM with other alternative coded modulation schemes. In the following, we will focus on comparison with BICM [31]–[36] that has attracted much attention recently for its performance advantages in fading channels.

SCM and BICM are closely related. Denote  $\mathbf{v}(j) = \{v_0(j), v_1(j), \dots, v_{M-1}(j)\}$  where  $v_m(j) \in \{0, 1\}$  is the  $m$ th coded bit carried by  $x(j)$ . With BICM, the transmitted signal  $x(j)$  is generated using, instead of (1), a more general mapping rule:

$$x(j) = \mu(\mathbf{v}(j)). \quad (33)$$

The image of  $\mu(\cdot)$  is usually a  $2^M$ -ary constellation of uniformly distributed signaling points, but the principle can be generalized to nonuniform constellations. With this view, some comments are in order.

- SCM is a special case of BICM since (1) is a special case of (33). As such, SCM may not outperform optimized BICM.
- The SCM in Fig. 1 involves multiple encoders while a BICM scheme usually involves only one overall encoder. We have observed that the performance of SCM with multiple encoders is better than that of SCM with a single encoder. (For the latter case, the signals in  $K$  layers are generated by interleaving and segmenting the outputs of a single encoder.)
- For very long codes in AWGN channels, SCM is as good as BICM, since SCM can achieve near-capacity performance (see Fig. 6). Later, in Fig. 8, we will show that for short codes, BICM does have advantages in certain cases.
- With QPSK modulation at each layer, SCM optimization only involves  $K = M/2$  weighting factors  $\{\beta_k\}$  (see Section II and [12]). BICM optimization is a much more complicated issue involving  $2^M$  constellation points [34].
- The detection complexity for SCM is  $O(M)$  while that for BICM is  $O(2^M)$ . Therefore SCM has a complexity advantage for large  $M$ .
- As demonstrated in Fig. 7, given a target rate, we can achieve diversity gain in SCM by decreasing the rate of each layer (and increasing  $K$  accordingly). The design and detection complexities of SCM grow linearly with  $K$ . For BICM, we can increase diversity gain by using larger constellations or rotating the signal constellations [35] but the design and detection complexities of these methods increase very quickly.
- As explained below, both SCM and BICM suffer from the high PAPR problem when OFDM is involved. We show that SCM is more robust to clipping effect compared to BICM.

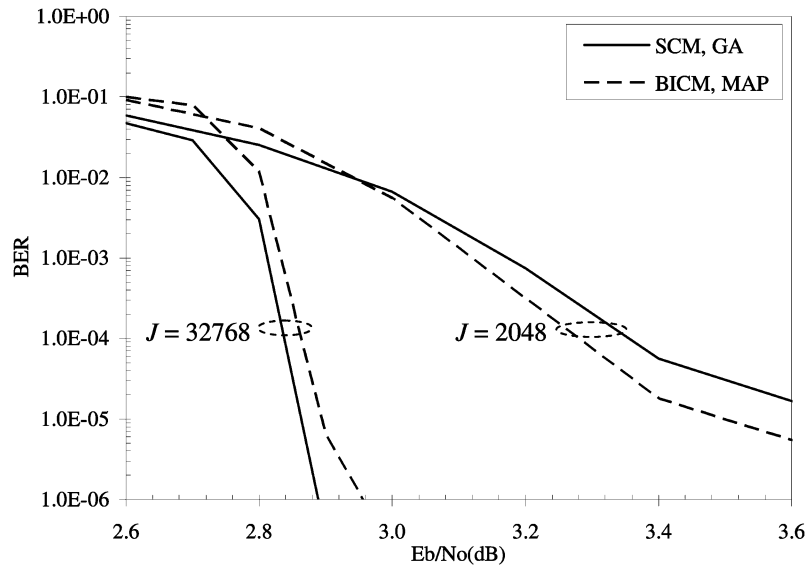


Fig. 8. Performance of turbo-coded  $((23, 35)_8)$  SCM and BICM at  $R = 2$  bits/symbol over AWGN channels.

In the following, we present several comparison examples based on turbo and convolutional codes.

#### A. Comparison in AWGN Channels

We first compare SCM and BICM over AWGN channels. We consider  $R = 2$  bits/symbol. The rate-1/2 turbo code [41]  $(23, 35)_8$  is employed in both schemes. For the SCM, we set  $K = 2$ ,  $\{\beta_k\} = \{1, 1.51\}$ ; PAPR = 2.83 dB; the number of iterations is 9 in the DEC and 2 between the DEC and the ESE. For the BICM, the Gray mapping is applied to the 16-QAM constellation;<sup>3</sup> PAPR = 2.55 dB; the number of iterations in the DEC is 18. The two schemes have nearly the same complexity. Clipping is not considered here since the PAPRs are not significant.

The performance with  $J = 2048$  and 32768 is shown in Fig. 8. The BICM performance is better when  $J$  is small while the SCM performance surpasses that of BICM for a large  $J$ . (Similar observations have been made for convolutional codes-based schemes over AWGN channels.) One reason for this is that the suboptimal GA detection is used in SCM while the optimal MAP demapping is used in BICM. Another reason is that the interleaver length for SCM is only  $1/K$  of that of BICM at fixed  $J$ , which affects interleaving gain [40]. (However, when  $J$  is large, the impact of interleaver length becomes less significant.)

#### B. Comparison in Fading Channels

Fig. 9 compares SCM with BICM over fading channels at rates  $R = 2, 3$  and 5 bits/symbol. The same component code with  $S = 4$  as that in Fig. 7 is again used here for SCM. For comparison, three BICM schemes with iterative decoding (BICM-ID) reported in [33], [34] and [36] are also simulated.

<sup>3</sup>Gray mapping yields the best known performance for BICM with turbo codes and QAM constellations [34].

For BICM-ID, the  $(23, 35)_8$  convolutional code is directly used for  $R = 2$ , and punctured to rate 3/4 and 5/6 (using the optimal puncture patterns in [42]) for  $R = 3$  and 5, respectively. The related parameters, including signaling schemes and PAPRs without clipping, are given in Table III, where SP denotes the set-partitioning mapping [1] and MSP the modified SP mapping [33]. Clipping (with  $\gamma = 3$  dB, PAPR  $\approx 3.55$  dB) is applied to SCM, but not used for BICM-ID since the related PAPRs are small.

From Fig. 9, we can see that at  $R = 2$ , BICM-ID exhibits better performance and lower PAPR (2.55 dB). For higher rates, clipped SCM is inferior to BICM-ID at low SNRs. However, SCM provides a lower error floor at high SNRs, since it can achieve higher diversity gains with low-rate component codes. Here we would like to point out that the performance provided in Fig. 9 are based on the best schemes known to us, although it may be possible to improve the BICM-ID performance through further optimization.

With real  $\{\beta_k\}$  and the GA method, the detection complexity of SCM is about 6 real multiplications, 6 real additions, and a  $\tanh(\cdot)$  operation per coded bit [29]. As a comparison, in BICM-ID schemes employing  $2^M$ -ary constellations and the log-MAP demapping [33], the demapping complexity is about  $2^M$  comparisons,  $3 \times 2^M$  real additions and  $2^M$  table look-ups per coded bit, which can be very high when  $M$  is large (e.g.,  $M = 6$  for the 64-QAM signaling). We can show that, taking into account the APP decoding cost and the numbers of iterations needed, the overall complexities of the two schemes are comparable.

#### C. Comparison in Channels Involving OFDM Modulation

Finally, we briefly discuss the channels with OFDM modulation. Both SCM and BICM suffer from the high PAPR problem in this case. We apply clipping with  $\gamma = 3$  dB to the two

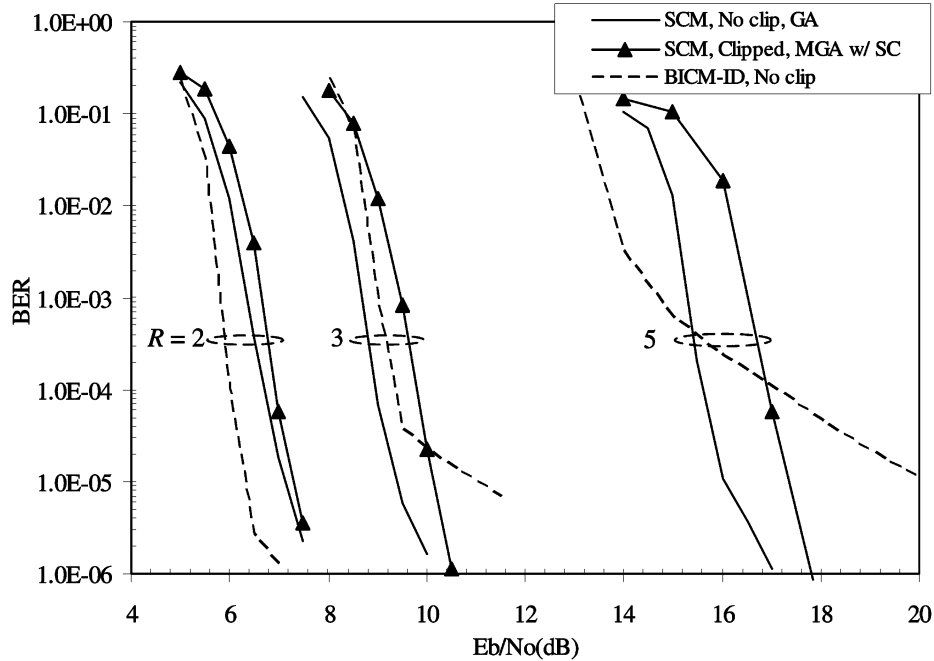


Fig. 9. Performance of SCM and BICM-ID over fully-interleaved Rayleigh fading channels.  $J = 2048$ . For the SCM without clipping, the number of iterations is 12 for  $R = 2$  and 3, and 16 for  $R = 5$ . For the clipped SCM, we set  $Q_M = 6$ ,  $Q_S = 6$  for  $R = 2$  and 3, and  $Q_M = 6$ ,  $Q_S = 10$  for  $R = 5$ . The number of iterations is 10 for all the BICM-ID results.

TABLE III  
PARAMETERS OF THE SCM AND BICM-ID SCHEMES IN FIGS. 9 AND 10

$R$	SCM			BICM-ID	
	$K$	$\{\beta_k \text{ value} \times \text{layer number}\}$	PAPR (dB)	Signaling Scheme	PAPR (dB)
2	8	$1 \times 6, 1.44 \times 2$	9.03	16-QAM, MSP [33]	2.55
3	12	$1 \times 6, 1.58 \times 3, 2.07 \times 2, 2.27 \times 1$	10.49	16-QAM, MSP [33]	2.55
5	20	$1 \times 6, 1.58 \times 3, 2.07 \times 1, 2.27 \times 1, 2.73 \times 1, 2.99 \times 1, 3.27 \times 1, 3.58 \times 1, 4.30 \times 2, 5.65 \times 1, 6.19 \times 2$	11.79	64-QAM, SP [1]	3.68

schemes and compare their performance.<sup>4</sup> The frame length  $J$  is set to 2048 and the number of subcarriers  $N$  is set to 256. (This means that each frame contains  $J/N = 8$  OFDM blocks.) For simplicity, the channel gains over subcarriers are assumed to be independent, Rayleigh-distributed [20]. The coding schemes are the same as those in Fig. 9. The operations related to the cyclic prefix of OFDM are ignored in our simulations.

We observe that the soft compensation strategy is not effective for BICM-ID with the SP and MSP mappings. A similar observation is made in [26] and an explanation is provided there for this observation using the EXIT chart technique. On the other hand, compensation techniques are more effective for BICM with Gray mapping, but the resultant performance is still not satisfactory. Based on this, we adopt the signal model given by (9) and treat the clipping distortion as an equivalent AWGN for clipped BICM-ID. The results are compared with SCM applying SC in Fig. 10. For reference, we have also included in Fig. 10 the results for Gray-mapped BICM with clipping ( $\gamma = 3$  dB) and SC.

It is seen that in general, SCM provides the best solution for clipped systems. The difference becomes significant when rate

is high ( $R = 3$  and 5). Notice the serious performance degradation of BICM-ID with clipping at  $R = 5$ . At this rate, the channel noise level is very low (due to high working SNR) and, consequently, clipping distortion becomes a dominating problem. Clearly, considering the PAPR problem, SCM provides a better choice than BICM-ID in high-rate OFDM applications.

## VII. CONCLUSION

We have investigated a peak-power-limited SCM system based on clipping. By evaluating the mutual information achieved with the clipped input signal, we have shown that significant 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 method. Simulation results show that a good tradeoff between PAPR and BER can be achieved with the proposed method. We have also compared SCM with BICM over various types of channels. It has been shown that, compared to BICM, SCM provides a simpler and more efficient means of achieving diversity gains for high-rate applications, especially when OFDM modulation with clipping is applied.

<sup>4</sup>In this paper, we only consider Nyquist-rate-sampled systems. The oversampled OFDM system [19]–[21] is a more complicated issue and we are currently conducting a comprehensive study on the application of SCM in such systems.

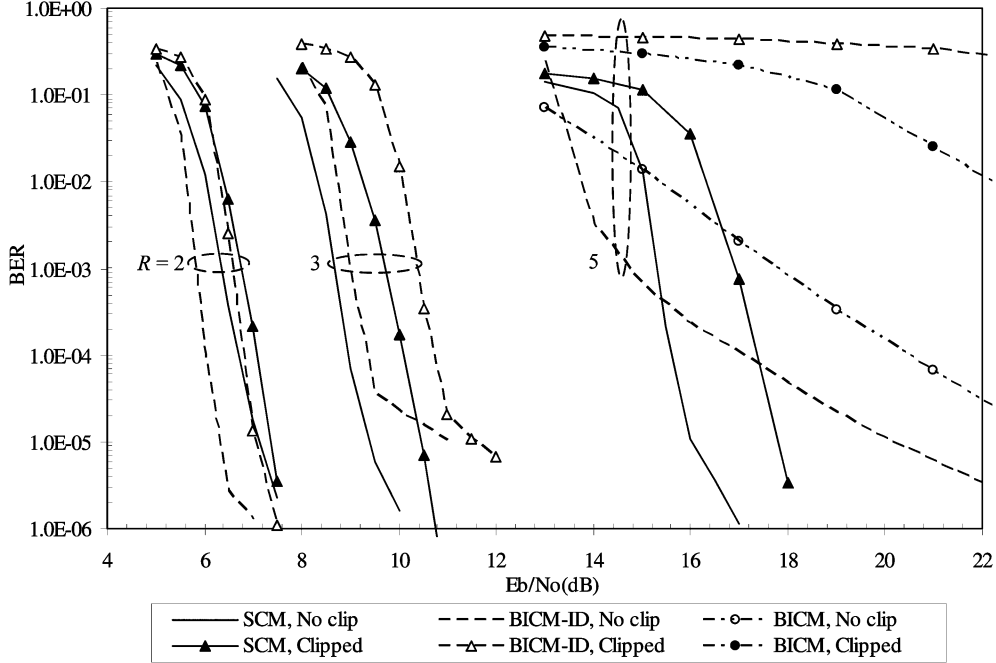


Fig. 10. Performance comparison of SCM and BICM-ID with OFDM modulation over Rayleigh fading channels;  $J = 2048$ .  $N = 256$ . The numbers of iterations are the same as those in Fig. 9 for SCM and BICM-ID. For the unclipped BICM, the number of iterations is 1. For the clipped BICM,  $Q_M = 1$  and  $Q_S = 15$ .

#### APPENDIX

##### THE GA METHOD FOR SCM WITH COMPLEX WEIGHTING FACTORS

In this appendix, we extend the Gaussian approximation (GA) detection method to SCM with complex weighting factors  $\{\beta_k\}$  and analyze the complexity. The statistics  $E[\zeta_k^{Re}(j)]$  and  $V[\zeta_k^{Re}(j)]$  in (15) [for the case of real  $\{\beta_k\}$ ] can also be evaluated by the methods below.

When  $\{\beta_k\}$  are complex, we first generate

$$\hat{y}(j) = \frac{\beta_k^*}{|\beta_k|} y(j) = |\beta_k| x_k(j) + \hat{\zeta}_k(j) \quad (34)$$

where  $\hat{\zeta}_k(j) = \beta_k^* \zeta_k(j) / |\beta_k|$  with  $\zeta_k(j)$  given by (14). We approximate  $\hat{\zeta}_k(j)$  by a complex, Gaussian random variable. Referring to (15), we can see that the key to estimate  $x_k^{Re}(j)$  is to find  $E[\hat{\zeta}_k^{Re}(j)]$  and  $V[\hat{\zeta}_k^{Re}(j)]$ . First, a definition. Let  $x$  be a complex random variable and  $E[x]$  be its mean. Define the covariance matrix of  $x$  as

$$\mathbf{C}[x] = \begin{bmatrix} V[x^{Re}] & C[x^{Re}, x^{Im}] \\ C[x^{Re}, x^{Im}] & V[x^{Im}] \end{bmatrix} \quad (35)$$

where  $C[x^{Re}, x^{Im}] = E[x^{Re} x^{Im}] - E[x^{Re}]E[x^{Im}]$ . Following [29],  $E[x_k(j)]$  and  $\mathbf{C}[x_k(j)]$  can be estimated as

$$E[x_k(j)] = \tanh\left(\frac{e_{\text{DEC}}(x_k^{Re}(j))}{2}\right) + i \tanh\left(\frac{e_{\text{DEC}}(x_k^{Im}(j))}{2}\right) \quad (36a)$$

$$\mathbf{C}[x_k(j)] = \begin{bmatrix} 1 - (E[x_k^{Re}(j)])^2 & 0 \\ 0 & 1 - (E[x_k^{Im}(j)])^2 \end{bmatrix} \quad (36b)$$

where  $e_{\text{DEC}}(x_k^{Re}(j))$  and  $e_{\text{DEC}}(x_k^{Im}(j))$  are the *extrinsic* LLRs generated by the DECs, we have assumed that the real and imaginary parts of  $x_k(j)$  are uncorrelated, and thus the off-diagonal entries of  $\mathbf{C}[x_k(j)]$  are zeros. The initial values of  $e_{\text{DEC}}(x_k^{Re}(j))$  and  $e_{\text{DEC}}(x_k^{Im}(j))$  are set to zeros, implying no *a priori* information. From (1) and (13), we have

$$E[y(j)] = \sum_{k=0}^{K-1} \beta_k E[x_k(j)] \quad (37a)$$

$$\mathbf{C}[y(j)] = \sum_{k=0}^{K-1} \mathbf{B}_k \mathbf{C}[x_k(j)] \mathbf{B}_k^T + \sigma^2 \mathbf{I} \quad (37b)$$

where

$$\mathbf{B}_k = \begin{bmatrix} \beta_k^{Re} & -\beta_k^{Im} \\ \beta_k^{Im} & \beta_k^{Re} \end{bmatrix} \quad (38)$$

the superscript “ $T$ ” denotes transpose of matrixes and  $\mathbf{I}$  is the  $2 \times 2$  identity matrix. Then, from (13), we have

$$E[\zeta_k(j)] = E[y(j)] - \beta_k E[x_k(j)] \quad (39a)$$

$$\mathbf{C}[\zeta_k(j)] = \mathbf{C}[y(j)] - \mathbf{B}_k \mathbf{C}[x_k(j)] \mathbf{B}_k^T. \quad (39b)$$

Now, we can generate

$$E[\hat{\zeta}_k(j)] = \frac{\beta_k^*}{|\beta_k|} E[\zeta_k(j)] \quad (40a)$$

$$\mathbf{C}[\hat{\zeta}_k(j)] = \frac{1}{|\beta_k|^2} \mathbf{B}_k^T \mathbf{C}[\zeta_k(j)] \mathbf{B}_k. \quad (40b)$$

Finally,  $E[\hat{\zeta}_k^{Re}(j)]$  and  $V[\hat{\zeta}_k^{Re}(j)]$  can be obtained from (40) and  $x_k^{Re}(j)$  is estimated as

$$e_{\text{ESE}}(x_k^{Re}(j)) = 2|\beta_k| \frac{\hat{y}_k^{Re}(j) - E[\hat{\zeta}_k^{Re}(j)]}{V[\hat{\zeta}_k^{Re}(j)]}. \quad (41)$$

We now consider the computational cost for the GA method as described by (34)–(41). Some simple methods, such as sharing the results of (37) for all layers, can be applied to reduce the computational cost. Also, some intermediate results can be reused to speed up the computations. It can be shown that the total computational cost of the ESE is about 15 real multiplications, 13 real additions and a few other operations [e.g.,  $\tanh(x)$  function] per coded bit [Note: Each  $x_k(j)$  carries two coded bits], which is independent of the number of layers  $K$ . For the case of real  $\{\beta_k\}$ , the complexity of the GA method can be further reduced. In contrast, the complexity of the MAP method is linear with  $4^K$ , i.e., the size of the signal constellation.

Compared with the DEC cost, the above ESE cost is quite moderate. As an example, the APP decoding [43] for a 16-state rate-1/2 convolutional code requires about 64 real multiplications and 32 real additions per coded bit. In this case, the overall complexity of the receiver is dominated by the DEC's. Considering that the APP decoding has become a standard function in modern systems employing turbo-like codes, we expect that the cost of the above GA method is acceptable for real applications.

#### ACKNOWLEDGMENT

The authors wish to thank the Associate Editor and the anonymous reviewers for their valuable comments. The authors are also grateful to Prof. Ezio Biglieri for enlightening discussions.

#### REFERENCES

- [1] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.
- [2] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. IT-23, pp. 371–377, May 1977.
- [3] L. Ping, B. Bai, and X. M. Wang, "Low complexity concatenated two-state TCM schemes with near capacity performance," *IEEE Trans. Inf. Theory*, vol. 49, pp. 3225–3234, Dec. 2003.
- [4] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1361–1391, Jul. 1999.
- [5] G. D. Forney, Jr., "Trellis shaping," *IEEE Trans. Inf. Theory*, vol. 38, pp. 281–300, Mar. 1992.
- [6] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*. New York: Wiley, 2002.
- [7] P. Limpaphayom and K. A. Winick, "Power- and bandwidth-efficient communications using LDPC codes," *IEEE Trans. Commun.*, vol. 52, pp. 350–354, Mar. 2004.
- [8] N. Varnica, X. Ma, and A. Kavčić, "Iteratively decodable codes for bridging the shaping gap in communication channels," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2002.
- [9] L. Duan, B. Rimoldi, and R. Urbanke, "Approaching the AWGN channel capacity without active shaping," in *Proc. IEEE Int. Symp. Information Theory*, Ulm, Germany, Jun./Jul. 1997, pp. 374–374.
- [10] S. Gadkari and K. Rose, "Time-division versus superposition coded modulation schemes for unequal error protection," *IEEE Trans. Commun.*, vol. 47, pp. 370–379, Mar. 1999.
- [11] X. Ma and L. Ping, "Coded modulation using superimposed binary codes," *IEEE Trans. Inf. Theory*, vol. 50, pp. 3331–3343, Dec. 2004.
- [12] X. Ma and L. Ping, "Power allocations for multilevel coding with sigma mapping," *Electron. Lett.*, vol. 40, no. 10, pp. 609–611, May 2004.
- [13] T. M. Cover and J. M. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [14] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [15] P. Wang, J. Xiao, and L. Ping, "Comparison of orthogonal and nonorthogonal approaches to future wireless cellular systems," *IEEE Veh. Technol. Mag.*, vol. 1, no. 3, pp. 4–11, Sep. 2006.
- [16] H. Schoeneich and P. A. Hoeher, "Adaptive interleave-division multiple access-A potential air interface for 4G bearer services and wireless LANs," in *Proc. 1st IEEE and IFIP Int. Conf. Wireless and Opt. Commun. Netw. (WOCN 2004)*, Muscat, Oman, Jun. 2004, pp. 179–182.
- [17] A. Goldsmith and S.-G. Chua, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 595–602, May 1998.
- [18] S. H. Han and J. H. Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," *IEEE Wireless Commun.*, vol. 12, no. 2, pp. 56–65, Apr. 2005.
- [19] X. Li and L. J. Cimini, "Effects of clipping and filtering on the performance of OFDM," *IEEE Commun. Lett.*, vol. 2, pp. 131–133, May 1998.
- [20] H. Ochiai and H. Imai, "Performance analysis of deliberately clipped OFDM signals," *IEEE Trans. Commun.*, vol. 50, pp. 89–101, Jan. 1999.
- [21] J. Armstrong, "Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering," *Electron. Lett.*, vol. 38, pp. 246–247, Feb. 2002.
- [22] J. Tellado, L. M. C. Hoo, and J. M. Cioffi, "Maximum-likelihood detection of nonlinearly distorted multicarrier symbols by iterative decoding," *IEEE Trans. Commun.*, vol. 51, pp. 218–228, Feb. 2003.
- [23] A. R. S. Bahai, M. Singh, A. J. Goldsmith, and B. R. Saltzberg, "A new approach for evaluating clipping distortion in multicarrier systems," *IEEE J. Sel. Areas Commun.*, vol. 20, pp. 1037–1046, Jun. 2002.
- [24] D. Kim and G. L. Stüber, "Clipping noise mitigation for OFDM by decision-aided reconstruction," *IEEE Commun. Lett.*, vol. 3, pp. 4–6, Jan. 1999.
- [25] H. Chen and A. M. Haimovich, "Iterative estimation and cancellation of clipping noise for OFDM signals," *IEEE Commun. Lett.*, vol. 7, pp. 305–307, Jul. 2003.
- [26] M. Colas, G. Gelle, and D. Declercq, "Analysis of iterative receivers for clipped COFDM signaling based on soft turbo-DAR," in *Proc. Int. Symp. Wireless Commun. Syst.*, 2004.
- [27] H. Nikopour, A. K. Khandani, and S. H. Jamali, "Turbo-coded OFDM transmission over a nonlinear channel," *IEEE Trans. Veh. Technol.*, vol. 54, no. 4, pp. 1361–1371, Jul. 2005.
- [28] R. Price, "A useful theorem for nonlinear devices having gaussian inputs," *IRE Trans. Inf. Theory*, vol. IT-4, pp. 69–72, Jun. 1958.
- [29] L. H. Liu, J. Tong, and L. Ping, "Analysis and optimization of CDMA systems with chip-level interleavers," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 141–150, Jan. 2006.
- [30] L. Ping, L. H. Liu, K. Y. Wu, and W. K. Leung, "Interleave-division multiple-access," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 938–947, Apr. 2006.
- [31] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, pp. 873–884, May 1992.
- [32] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, pp. 927–946, May 1998.
- [33] A. Chindapol and J. A. Ritcey, "Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels," *IEEE J. Sel. Areas Commun.*, vol. 19, pp. 944–957, May 2001.
- [34] J. Tan and G. Stüber, "Analysis and design of symbol mappers for iteratively decoded BICM," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 662–672, Mar. 2005.
- [35] N. H. Tran, H. H. Nguyen, and T. Le-Ngoc, "Performance of BICM-ID with signal space diversity," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1732–1742, May 2007.
- [36] S. X. Ng, T. H. Liew, and L. Hanzo, "Comparative study of TCM, TTCM, BICM and BICM-ID schemes," in *Proc. IEEE VTC'2001*, Rhodes, Greece, May 2001, pp. 2450–2454.
- [37] S. t. Brink, J. Speidel, and R. Yan, "Iterative demapping and decoding for multilevel modulation," in *Proc. IEEE GLOBECOM*, Nov. 1998, pp. 579–584.
- [38] S. t. Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, pp. 1727–1737, Oct. 2001.
- [39] S. ten Brink, "Rate one-half code for approaching the Shannon limit by 0.1 dB," *Electron. Lett.*, vol. 36, no. 15, pp. 1293–1294, Jul. 2000.
- [40] S. Benedetto and G. Montorsi, "Unveiling turbo codes: Some results on parallel concatenated coding schemes," *IEEE Trans. Inf. Theory*, vol. 42, pp. 409–428, Mar. 1996.
- [41] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," in *Proc. IEEE Int. Conf. Commun.*, Geneva, Switzerland, May 1993, pp. 1064–1070.
- [42] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2000.

- [43] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. IT-20, pp. 284–287, Mar. 1974.

**Jun Tong** (S'06) received the B.E. degree in information engineering and the M.E. degree in signal and information processing from the University of Electronic Science and Technology of China, Chengdu, in 2001 and 2004, respectively.

He is currently working toward the Ph.D. degree at City University of Hong Kong. His research interests are signal processing and coding techniques in communication systems.

**Li Ping** (S'87–M'91–SM'06) received the Ph.D. degree from Glasgow University, Scotland, in 1990.

He lectured at the Department of Electronic Engineering, Melbourne University, Australia, from 1990 to 1992, and worked as a member of Research Staff at Telecom Australia Research Laboratories from 1993 to 1995. He has been

with the Department of Electronic Engineering, City University of Hong Kong, since January 1996, where he is now a Chair Professor. His research interests are communications systems and coding theory.

Dr. Li Ping was awarded a British Telecom-Royal Society Fellowship in 1986, the IEE J. J. Thomson premium in 1993, and a Croucher Senior Research Fellowship in 2005.

**Xiao Ma** received the Ph.D. degree in communication and information systems from Xidian University, China, in 2000.

From 2000 to 2002, he was a Postdoctoral Fellow with Harvard University, Cambridge, MA. From 2002 to 2004, he was a Research Fellow with City University of Hong Kong. He is now a Professor with the Department of Electronics and Communication Engineering, Sun Yat-sen University, Guangzhou, China. His research interests include information theory, channel coding theory and their applications to communication systems and digital recording systems.

Dr. Ma is a corecipient, with A. Kavcic and N. Varnica, of the 2005 IEEE Best Paper Award in Signal Processing and Coding for Data Storage.