Analysis and Design of Symbol Mappers for Iteratively Decoded BICM

Jun Tan, *Member, IEEE,* Gordon L. Stüber, *Fellow, IEEE*

Abstract—Iterative decoding and demodulation of bit-interleaved coded modulation (BICM) is investigated. The union bound on the error probability of BICM with maximum-likelihood (ML) decoding is derived with the assumption of a uniform interleaver. Based on the union bound, it is shown that an interleaving gain is achievable for BICM at high SNR. The design rules on the symbol mapper for optimal asymptotic performance at high SNR are derived and a new symbol mapper, called the maximum squared Euclidean weight (MSEW) symbol mapper, is proposed with examples given for 8-PSK and 16-QAM. The MSEW mapper is shown by means of an extrinsic information transfer (EXIT) chart to provide the best error-floor performance for BICM at high SNR. At low SNR, a mapper doping technique is proposed that combines the MSEW mapper with Gray mapper to provide a good *match* to a given outer code. Simulation results are also presented to confirm our analysis.

Index Terms—bit-interleaved coded modulation (BICM), iterative decoding, Turbo codes

I. Introduction

Trellis coded modulation (TCM) [1] uses mapping by set partitioning to maximize the free Euclidean distance, thus providing good performance on AWGN channels. When TCM is applied on flat fading channels with symbol-by-symbol interleaving we try to maximize the minimum built-in time diversity (or shortest error event path) of the code. This is because the code performance on fading channels depends more strongly on the minimum Hamming distance between coded symbol sequences than on the minimum Euclidean distance. However, the Hamming distance can be increased further by using bit-by-bit interleaving of the code bits prior to symbol mapping rather than symbol-by-symbol interleaving of the code symbols after symbol mapping. The former approach is called bit-interleaved coded modulation (BICM), and has been shown to outperform symbol-interleaved TCM on fading channels [2], [3].

The bit interleaver in the BICM system serves as a channel interleaver to permute the coded bit stream for the purpose of achieving time diversity on fading channels. It concatenates the encoder and the symbol mapper and makes the BICM system to resemble a serial concatenated convolutional code (SCCC) [4]. This suggests that iterative demapping/decoding of BICM can yield a turbo-like performance, as shown for 8-PSK [5], [6] and 16-QAM modulation [7]. For iteratively decoded BICM, the channel interleaver which is designed to provide

time diversity also serves as a code interleaver to provide *interleaving gain*, similar to the interleavers used in turbo encoders. Maximum likelihood (ML) decoding of BICM is infeasible because of the large number of states introduced by the interleaver. However, iterative demodulation and decoding serves as an effective, yet sub-optimal, decoding method for BICM that may approach the ML decoding performance.

In [6], [8], an error probability bounding technique is applied to 8-PSK-BICM based on the theory by Caire *et al.* [3]. This bounding technique is also applied to 16-QAM-BICM in [7]. These error probability bounds are derived under the assumption of error-free feedback (EFF), that is, there is no error in the iterative decoding process. Based on the EFF bound, the "optimal" mapper is found for both 8-PSK [6], [8] and 16-QAM [7] that maximizes the harmonic mean, a concept introduced in [3]. However, the error-free feedback assumption is an idealized assumption for iterative decoding of BICM; decoding errors always exist in reality.

BICM can be viewed as a special type of SCCC, and the ML error probability bound for SCCC in [4] can be extended to BICM. The "inner code" for BICM is the symbol mapper, which can be considered as a non-recursive single-state encoder. Unlike the inner code in a SCCC system that is recursive and outputs binary code bits, the symbol mapper is non-recursive and outputs complex symbols chosen from a modulation alphabet. Therefore, the error probability bounding techniques for SCCC cannot be directly applied to BICM. However, through appropriate analysis of the symbol mapper, we derive the ML error probability bound for BICM with a uniform interleaver assumption. Based on the ML decoding bound, a new symbol mapper, called the maximum squared Euclidean weight (MSEW) symbol mapper, is developed and applied to 8-PSK and 16-QAM. By using the density evolution approach [9]–[11] for BICM, we establish that there is no one mapper having the best performance over all SNR regions. In the high SNR region, our proposed MSEW mapper can provide the best performance. At low SNR region, we propose a "doping" technique, motivated by the doping concept of turbo codes [9], [12], to match the mapper to a given outer code to optimize the performance at a given channel SNR.

The remainder of this paper is organized as follows. Section II, briefly discusses the BICM system along with the necessary soft-input soft-output (SISO) modules needed for iterative decoding. Section III presents the performance analysis and design rules for the symbol mapper based on bit error rate (BER) union bounds for convolutionally coded and uniformly interleaved BICM systems. Density evolution techniques are briefly reviewed and the role of the MSEW mapper is discussed in Section IV. Simulation results are

presented in Section V to confirm our analytical results. Section VI wraps up with a summary of conclusions.

II. BICM SYSTEM AND ITERATIVE DECODING

A general bit-interleaved coded modulation (BICM) system consists of the interleaved concatenation of an encoder and a symbol mapper. The outer encoder encodes the information sequence $\{u_k\}$ into a coded sequence. The coded binary sequence is then bit interleaved by the interleaver π to become the interleaved sequence $\{c_k\}$, which is applied to the symbol mapper. The output of the symbol mapper is the modulating symbol sequence $\{x_k\}$, comprised of complex symbols chosen from an M-ary signal constellation such as M-PSK or M-QAM.

For M-ary modulation, where $M = 2^m$, a symbol mapper generates baseband complex symbols based on m-bit binary input sequences. The mapper can be represented by the one-to-one mapping $f : \{0,1\}^m \rightarrow \mathcal{X}, \quad x = f(\mathbf{c})$, where the input binary sequence is $\mathbf{c} = (c_0, \cdots, c_{m-1})$, $c_j \in \{0,1\}$, and x is chosen from the set $\mathcal{X} = \{x\}$ consisting of $M = 2^m$ complex signal points. For a length-N binary input sequence $\mathbf{c}$, the mapper produces a length-N/m output symbol sequence $\mathbf{x}$, where each symbol x_k represents the corresponding m-bit sub-sequence $(c_{(k-1)m}, c_{(k-1)m-1}, \cdots, c_{km-1})$ in $\mathbf{c}$. We may denote $\mathbf{x} = f(\mathbf{c})$.

A. SISO Modules and Iterative Decoding of BICM

For convolutional coded BICM, the structure of the corresponding iterative decoder is similar to the iterative decoder for SCCC in [13]. The SISO module corresponding to the outer convolutional encoder can be based on the MAP, Max-Log-MAP, SOVA and other such algorithms. The SISO module corresponding to the symbol mapper can use the MAP algorithm, since the symbol mapper is a single-state finite-state machine. This SISO module is called the SISO demapper.

The SISO demapper can be implemented in the log-likelihood ratio (LLR) domain. For each signal point $x = f(\mathbf{c})$, there is an m-bit sequence $\mathbf{c} = (c_0, \cdots, c_{m-1})$ with associated *a priori* information $L(c_0), \cdots, L(c_{m-1})$, where $L(c_j) \triangleq \log \frac{P(c_j=1)}{P(c_j=0)}$, $j = 0, \cdots, m-1$.

The channel can be modeled as $y_k = a_k x_k + n_k$, where n_k is the complex zero-mean Gaussian noise with variance $N_0/2$ in each dimension, and a_k is the complex time-variant fading gain. For coherent receivers with perfect channel knowledge, the conditional pdf of the observation y is $p(y|x,a) = \frac{1}{\pi N_0} \exp(-\frac{E_s}{N_0} \parallel y - ax \parallel^2)$. The log-probability becomes

$$
\begin{aligned}
\log p(y|x) &= \log \frac{1}{\pi N_0} - \frac{E_s}{N_0} \parallel y \parallel^2 - \parallel a \parallel^2 \frac{E_s}{N_0} \parallel x \parallel^2 \\
&\quad +2 \parallel a \parallel \frac{E_s}{N_0}(y_I x_I + y_Q x_Q),
\end{aligned} \tag{1}
$$

where $y = y_I + jy_Q$, $x = x_I + jx_Q$, and E_s/N_0 is the modulated symbol energy-to-noise ratio. The receiver's estimate of x is represented as $\hat{x}$ and $\hat{\mathbf{c}} = (\hat{c}_0, \cdots, \hat{c}_{m-1})$ denotes its corresponding binary representation where $\hat{x} = f(\hat{\mathbf{c}})$. The soft output is

$$
\begin{aligned}
L(\hat{c}_j) &\triangleq \log \frac{P(c_j=1|y)}{P(c_j=0|y)} \\
&= \log \frac{\displaystyle\sum_{x:c_j=1} \exp\left[\log p(y|x,a) + \log P(x)\right]}{\displaystyle\sum_{x:c_j=0} \exp\left[\log p(y|x,a) + \log P(x)\right]}. \tag{2}
\end{aligned}
$$

If all coordinates c_j in $\mathbf{c}$ are independent, then

$$
\log P(x) = \log P(\mathbf{c}) = \log P(c_0)P(c_1)\cdots P(c_{m-1}). \tag{3}
$$

Multiplying both the numerator and denominator in (2) by the term $\prod_{j=0}^{m-1} 1/P(c_j = 0)$ and using the relation $\log \frac{P(c_j)}{P(c_j=0)} = c_j L(c_j)$, yields

$$
L(\hat{c}_j) = \log \frac{\displaystyle\sum_{x:c_j=1} \exp\left[\lambda(y,x)\right]}{\displaystyle\sum_{x:c_j=0} \exp\left[\lambda(y,x)\right]}, \tag{4}
$$

where

$$
\begin{aligned}
\lambda(y,x) &\triangleq - \parallel a \parallel^2 \frac{E_s}{N_0} \parallel x \parallel^2 +2 \parallel a \parallel \frac{E_s}{N_0}(y_I x_I + y_Q x_Q) \\
&\quad + \sum_{j=0}^{m-1} c_j L(c_j). \tag{5}
\end{aligned}
$$

The actual output of the SISO demapper is the extrinsic information, given by $L_e(\hat{c}_j) = L(\hat{c}_j) - L(c_j)$.

III. CONVOLUTIONALLY CODED BICM

In this section, BER union bounds on maximum likelihood (ML) demapping/decoding of convolutionally encoded BICM are derived. The bounds are used to study the performance of BICM with different symbol mapping approaches. A design criterion for the symbol mapper is then introduced that can be used to optimize the asymptotic performance at high bit energy-to-noise ratios.

A. Transfer Functions of Mapper

To analyze BICM as a concatenated coding system, the symbol mapper is treated as a coding entity that is characterized by its input Hamming distance and output Euclidean distance. The symbol mapper is a single-state *encoder* that generates encoded symbols corresponding to its input binary sequence. For example, the QPSK Gray symbol mapper shown in Fig. 1 has a one-state trellis with four parallel transitions, where each state transition is characterized by a 2-bit input sequence and an output signal point. Throughout our development we assume normalized constellations, i.e., $E_s = 1$. For QPSK there are two squared Euclidean distances (SED) in the signal constellation, denoted as $\alpha_1 = 4$ and $\alpha_2 = 8 = 2\alpha_1$, where $\alpha_1 = \|x_{00} - x_{10}\|^2$ and $\alpha_2 = \|x_{00} - x_{11}\|^2$. The transfer function approach can be used to enumerate the possible state transitions and their distance properties. For this purpose, we introduce two dummy variables L and H, whose

exponents indicate the input Hamming distance and output SED, respectively, from a reference symbol. For example, if x_{00} is the reference symbol, the state transition with input sequence (11) can be denoted as $L^2 H^{2\alpha_1}$, indicating that the input Hamming distance from (00) is 2, and the output SED from x_{00} is $2\alpha_1$, shown in Fig. 1.
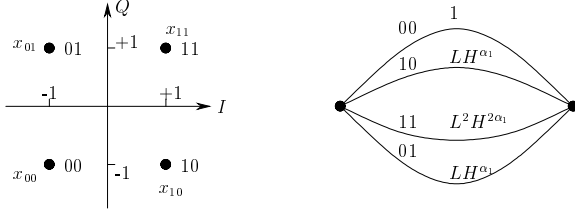

Fig. 1.   Gray symbol mapping for QPSK.

Assigning x_{00} as the reference symbol, the transfer function of the QPSK Gray mapper is

$$T(L,H) = 1 + 2LH^{\alpha_1} + L^2 H^{2\alpha_1} = (1 + LH^{\alpha_1})^2 . \quad (6)$$

For QPSK, the transfer function is invariant to the choice of reference symbol. Such symbol mappers are called *uniform* symbol mappers.

For 8-PSK mappers, the relations between the input Hamming weights and output SEDs are more complicated. The 8-PSK Gray mapper is shown in Fig. 2-(a). The list of all possible SEDs between the 8-PSK symbols is as follows:

$$
\begin{aligned}
\alpha_1 &= \parallel x_{001} - x_{000} \parallel^2 = 2 - \sqrt{2}, \\
\alpha_2 &= \parallel x_{011} - x_{000} \parallel^2 = 2, \\
\alpha_3 &= \parallel x_{010} - x_{000} \parallel^2 = 2 + \sqrt{2}, \\
\alpha_4 &= \parallel x_{110} - x_{000} \parallel^2 = 4. \quad (7)
\end{aligned}
$$

The corresponding mapper transfer function is

$$
\begin{aligned}
T_{\text{Gray}}(L,H) &= 1 + L(2H^{\alpha_1} + H^{\alpha_3}) + L^2(2H^{\alpha_2} + H^{\alpha_4}) \\
&\quad + L^3 H^{\alpha_3} . \quad (8)
\end{aligned}
$$

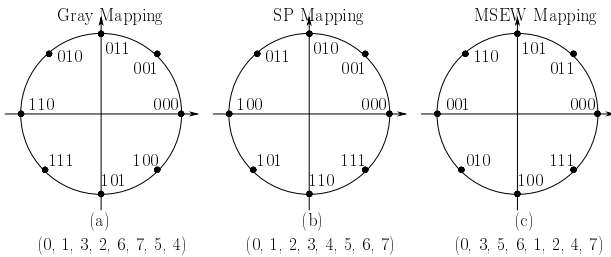Note that the 8-PSK Gray mapper is a uniform mapper.

Fig. 2.   8-PSK Symbol Mappers.

For conciseness, we use the eight-element vector $(p_0, \cdots, p_7)$ to represent an 8-PSK mapper, where p_n is the decimal number of the binary representation at the 8-PSK constellation point $e^{jn\pi/4}$. For example, the Gray mapping shown in Fig. 2-(a), can be represented as $(0, 1, 3, 2, 6, 7, 5, 4)$.

The 8-PSK set-partition (SP) mapper $(0, 1, 2, 3, 4, 5, 6, 7)$ is shown in Fig. 2-(b). The SP mapper is a mapper used for trellis coded modulation and is obtained through the principle of mapping by set partitioning [1]. The SP mapper is non-uniform and, thus, the transfer function depends on the choice of reference symbol. For example, choosing the reference symbol x_{000} gives

$$
\begin{aligned}
T(L,H)|_{x_{000}} &= 1 + L(H^{\alpha_1} + H^{\alpha_2} + H^{\alpha_4}) \\
&\quad + L^2(H^{\alpha_2} + 2H^{\alpha_3}) + L^3 H^{\alpha_1}. \quad (9)
\end{aligned}
$$

while choosing the reference symbol x_{001} gives

$$
\begin{aligned}
T(L,H)|_{x_{001}} &= 1 + L(H^{\alpha_1} + H^{\alpha_2} + H^{\alpha_4}) \\
&\quad + L^2(H^{\alpha_1} + H^{\alpha_2} + H^{\alpha_3}) + L^3 H^{\alpha_4}. \quad (10)
\end{aligned}
$$

If all points in the signal constellation are transmitted with equal probability, then averaging the transfer function over all possible reference symbols yields

$$
\begin{aligned}
T_{\text{SP}}(L,H) &= 1 + L(H^{\alpha_1} + H^{\alpha_2} + H^{\alpha_4}) \\
&\quad + L^2(\frac{1}{2}H^{\alpha_1} + H^{\alpha_2} + \frac{3}{2}H^{\alpha_3}) \\
&\quad + \frac{1}{2}L^3(H^{\alpha_1} + H^{\alpha_3}). \quad (11)
\end{aligned}
$$

The average transfer function is independent of the choice of reference symbol.

Finally, another 8-PSK symbol mapping, called MSEW mapper in our later discussion, is also shown in Fig. 2-(c). The MSEW mapper is non-uniform as well, but its transfer function averaged over all possible equally likely reference symbols is

$$
\begin{aligned}
T_{\text{N}}(L,H) &= 1 + L(H^{\alpha_2} + H^{\alpha_3} + H^{\alpha_4}) \\
&\quad + L^2(\frac{3}{2}H^{\alpha_1} + H^{\alpha_2} + \frac{1}{2}H^{\alpha_3}) \\
&\quad + L^3(\frac{1}{2}H^{\alpha_1} + \frac{1}{2}H^{\alpha_3}). \quad (12)
\end{aligned}
$$

B. Performance of BICM

By treating BICM as a serial concatenated convolutional code (SCCC), the analysis of SCCC in [4], [14] can be applied directly to BICM. With a *uniform interleaver* (which maps to any possible permutation with equal probability), the union bound on the bit error probability of BICM with ML decoding is

$$
\begin{aligned}
P_b &\leq \sum_{w=1}^{k} \frac{w}{k} \sum_{l=0}^{N} \frac{A_{w,l}^{(c)}}{\binom{N}{l}} A^{(i)}(l,H)|_{H=e^{-\frac{1}{4}mR\gamma_b}} \\
&= \sum_{l=0}^{N} \frac{1}{\binom{N}{l}} \left[\frac{1}{k} \sum_{w=1}^{k} w A_{w,l}^{(c)} \right] A^{(i)}(l,H)|_{H=e^{-\frac{1}{4}mR\gamma_b}}, \quad (13)
\end{aligned}
$$

where $\gamma_b = E_b/N_0$ is the bit energy-to-noise ratio, $R = k/N$ is the code rate with code bit length N, information bit length k, and m is the number of bits per modulated symbol. In (13), $A_{w,l}^{(c)}$ is the number of codewords in the outer code having input/output Hamming weights w and l, respectively, $A^{(i)}(l,h)$ is the number of modulating symbol sequences generated by the symbol mapper with input Hamming weight l and output squared Euclidean weight (SEW) h, and $A^{(i)}(l,H)$ is the

conditional weight enumerating function (CWEF) defined as $A^{(i)}(l, H) = \sum_h A^{(i)}(l, h) H^h$, which can be obtained from the input-output weight enumeration function (IOWEF) of the mapper.

To analyze the BER performance of BICM we use an approach similar to that used in the analysis of PCCC and SCCC [4], [15]. In the decoding of BICM, errors occur when the decoded symbol sequence is not the transmitted sequence. Such error events are the concatenation of length-1 error events in the mapper trellis. For an error symbol sequence of length-N/m, the number of error events can range from 1 to N/m. Denote $n_M^{(i)}$ as the maximum number of error events. The CWEF can be further represented as [4], [15]

$$A^{(i)}(l, H) = \sum_{j=1}^{n_M^{(i)}} \binom{N/m}{j} A^{(i)}(l, H, j), \tag{14}$$

where $A^{(i)}(l, H, j)$ is the weight enumerating function of sequences that concatenate j error events with input weight l. A similar approach can be taken for the outer convolutional code. After some manipulation, the BER upper bound becomes

$$
\begin{aligned}
P_b \quad < \quad & \sum_{w=1}^{k} \sum_{l=0}^{N} \sum_{n^{(c)}=1}^{n_M^{(c)}} \sum_{n^{(i)}=1}^{n_M^{(i)}} N^{n^{(c)}+n^{(i)}-l-1} \frac{l^l l! R^{n^{(c)}-1}}{n^{(c)}! n^{(i)}! m^{n^{(i)}}} \\
& \times A_{w,l,n^{(c)}}^{(c)} A^{(i)}(l, H, n^{(i)})|_{H=e^{-\frac{1}{4} m R \gamma_b}}, \tag{15}
\end{aligned}
$$

where $n^{(c)}$ and $n^{(i)}$ are the number of error events in the convolutional encoder trellis and symbol mapper trellis, respectively.

The exponent $n^{(c)} + n^{(i)} - l - 1$ of N in (15) determines the *interleaving gain*, where for a given SEW h the coefficients of the exponent depend on the input weight w, code weight l, and block length N, among other parameters. Define the maximum exponent as [4] $\alpha(h) \triangleq \max_{w,l}\{n^{(c)}+n^{(i)}-l-1\}$, which is the dominant coefficient among all coefficients for a sequence with SEW h. Define $q \triangleq \lceil d_f^{(c)}/m \rceil$, where $d_f^{(c)}$ is the free Hamming distance of the convolutional code. The minimum SEW of BICM signal sequences is, therefore, qh_i, where h_i is the minimum SEW of the symbol mapper. For 8-PSK, $h_i = \alpha_1 = 2 - \sqrt{2}$ from (7).

At high SNR, the performance of the BICM system is dominated by the sequence having minimum SEW $h = qh_i$. When $h = qh_i$, the number of error events in the symbol mapper trellis is $n^{(i)} = q$, the code weight is $l = d_f^{(c)}$, and the maximum number of error events of the convolutional code is $n_M^{(c)} \leq \lfloor \frac{d_f^{(c)}}{d_f^{(c)}} \rfloor = 1$. The exponent becomes

$$\alpha(qh_i) \leq q - d_f^{(c)}. \tag{16}$$

For $M = 2^m$-ary modulation, where $m > 1$, $q < d_f^{(c)}$ meaning that $\alpha(qh_i) < 0$. Therefore, an *interleaving gain* can be achieved at high SNR for the minimum SEW sequence, and the gain is determined by the exponent $q - d_f^{(c)}$. For example, an 8-PSK BICM system with a convolutional code of free distance $d_f = 5$ has $\alpha(qh_i) = -3$. Note that when $m = 1$, i.e., binary modulation such an interleaving gain does not exist.

When the block length $N \to \infty$, the asymptotic BER is dominated by the largest $\alpha(h)$ term in (15). Define $\alpha_M \triangleq \max_h\{\alpha(h)\} = \max_{w,l,h}\{n^{(c)}+n^{(i)}-l-1\}$ as the dominant contribution to the BER performance. Since $n^{(i)} = l$, the maximum number of error events of the symbol mapper equals the Hamming weight of the code sequence l. Hence, the maximum exponent is $\alpha_M = \max_{w,l,h}\{n^{(c)}-1\} = n_M^{(c)}-1 \geq 0$, which shows that an *interleaving gain* does not exist at high SNR. This conclusion is consistent with the results for SCCC with non-recursive inner codes [4]. Because the symbol mapper can be considered as a memoryless non-recursive encoder, no interleaving gain can be achieved for large code lengths N.

At high SNR where the BER performance is dominated by the minimum SEW sequence (qh_i sequence), the BICM system has *interleaving gain* in the sense that the BER contribution from the minimum SEW sequence decreases as code length N increases since $\alpha(qh_i) < 0$. However, when $N \to \infty$, the BER is not dominated by the minimum SEW sequence and the BICM system cannot provide interleaving gain because $\alpha_M \geq 0$. This means that there exists some block size N_M, such that for $N \leq N_M$ the BER performance can be improved by increasing N, i.e., there is interleaving gain. But for $N > N_M$ the performance cannot be improved by increasing N and there is no interleaving gain. Simulation results show that N_M is usually very large for practical systems. For 8-PSK BICM with a rate-2/3 convolutional code having free distance $d_f = 5$, $N_M > 5000$. This means that 8-PSK BICM can obtain an "interleaving gain" by increasing block length N, as long as $N < 5000$.

C. 8-PSK MSEW Mapper

In the analysis of the ML decoding bound of BICM in (13), we assume a uniform interleaver that permutes any Hamming weight l sequence to any of its permutations with equal probability $1/\binom{N}{l}$. Considering the outer code, the most probable error event is the error sequence with minimum Hamming weight, i.e., having the free distance d_f for convolutional codes. Given the minimum weight error sequence there is a very small chance that the d_f error bits after they are interleaved will be mapped to the same symbol since $N \gg d_f$. With high probability the d_f bits will be scattered throughout the permuted length-N sequence, such that the d_f bits will provide d_f Hamming weight-1 inputs that are mapped onto d_f different symbols. Consequently, the SEW of the Hamming weight-1 symbols is critical to the asymptotic BER performance. A low asymptotic BER can be achieved by maximizing the corresponding SEW of Hamming weight-1 input sequences to the symbol mapper.

The input-output weight enumeration function of the mapper can be represented as $A^{(i)}(L, H) = T^N(L, H)$, where N is the number of modulating symbols per block, and $T(L, H)$ is the transfer function of the mapper. For symbols mappers having m bits per symbol, the transfer function can be represented as

$$
\begin{aligned}
T(L, H) \quad = \quad & 1 + T(1, H)L + T(2, H)L^2 \\
& + \cdots + T(m, H)L^m, \tag{17}
\end{aligned}
$$

where $T(l, H)$ is the coefficient of the term L^l in $T(L, H)$. Given a codeword with Hamming weight d_f, when the interleaver permutes the codeword so that all d_f bits are at least m bits apart, the CWEF at $l = d_f$ becomes $A^{(i)}(l, H) = [T(1, H)]^{d_f}$. Based on the ML decoding bound of (13), we need to minimize

$$A^{(i)}(d_f, H)\big|_{H=e^{-\frac{1}{4}mR\gamma_b}} = T^{d_f}(1, H)\big|_{H=e^{-\frac{1}{4}mR\gamma_b}}. \quad (18)$$

The 8-PSK Gray mapping in (8) has $T(1, H) = 2H^{\alpha_1} + H^{\alpha_3}$, which indicates two Hamming weight-1 symbols with SEW α_1 and one Hamming weight-1 symbol with SEW α_3. The minimum SEW for Hamming weight-1 inputs is α_1. For the 8-PSK SP symbol mapping in (11), $T(1, H) = H^{\alpha_1} + H^{\alpha_2} + H^{\alpha_4}$, and there is only one Hamming weight-1 symbol with SEW α_1. Because $\alpha_1 < \alpha_2 < \alpha_3 < \alpha_4$, the RHS of (18) for SP mapping is smaller than that for Gray mapping. This indicates that the asymptotic BER with SP mapping should be lower than that with Gray mapping, since the dominant term in the error probability upper bound has an negative exponential dependency on the minimum SEW term, H^{α_1}. The MSEW symbol mapping shown in Fig. 2-(c) has a minimum SEW of α_2 for Hamming weight-1 input sequences. The asymptotic BER with this mapping should be lower than that achieved with either the Gray or SP mappings. The MSEW symbol mapper maximizes the minimum SEW for Hamming weight-1 input sequences, hence its acronym MSEW (maximum squared Euclidean weight).

The MSEW mapper is not unique. We have conducted an exhaustive search for all 8-PSK MSEW mappers, and have obtained the 14 distinct mappers listed in Tab. I. Equivalent mappers that are obtained through constellation rotation or mirror-mapping have been excluded. The MSEW mappers can provide optimal asymptotic performance because the minimum SEW for Hamming weight-1 symbols is maximized. Among the MSEW mappers, different mappers may have different performance due to differences in their transfer functions. The performance can be further optimized by minimizing the multiplicity of symbol pairs with minimum Euclidean distance, i.e., minimize the coefficient of the term of LH^{α_m}, where α_m is the minimum SEW. Such MSEW mappers are called *optimal* MSEW mappers. For 8-PSK, the *optimal* MSEW mappers are the first 12 mappers listed in Tab. I. The last two mappers are not optimal MSEW mappers.

D. 16-QAM MSEW Mapper

Our methodology can be readily applied to other signal constellations, and here we consider 16-QAM. We use a 16-element vector to represent the 16-QAM mappers, labelling the constellation symbols with the decimal representation of 4-bit binary numbers from upper-left to bottom-right. For example, the Gray mapper shown in Fig. 3-(a) can be represented as $(15, 11, 3, 7, 14, 10, 2, 6, 12, 8, 0, 4, 13, 9, 1, 5)$, where $15 = (1111)_2$ is the label for the upper-left symbol, $11 = (1011)_2$ is the label for the second symbol in the first row, and so on. The SP symbol mappings for 16-QAM are shown in Fig. 3-(b), with representation $(8, 13, 12, 9, 15, 10, 11, 14, 4, 1, 0, 5, 3, 6, 7, 2)$. The possible squared Euclidean distances (SEDs) between 16-QAM

TABLE I
8-PSK MSEW MAPPERS

8-PSK Mappers	Transfer function terms of L
$(0, 3, 4, 7, 1, 2, 5, 6)$	$H^{\alpha_2} + H^{\alpha_3} + H^{\alpha_4}$
$(0, 3, 4, 7, 2, 1, 6, 5)$	$H^{\alpha_2} + H^{\alpha_3} + H^{\alpha_4}$
$(0, 3, 5, 2, 4, 7, 1, 6)$	$H^{\alpha_2} + H^{\alpha_3} + H^{\alpha_4}$
$(0, 3, 5, 6, 1, 2, 4, 7)$	$H^{\alpha_2} + H^{\alpha_3} + H^{\alpha_4}$
$(0, 3, 6, 1, 4, 7, 2, 5)$	$H^{\alpha_2} + H^{\alpha_3} + H^{\alpha_4}$
$(0, 3, 6, 5, 2, 1, 4, 7)$	$H^{\alpha_2} + H^{\alpha_3} + H^{\alpha_4}$
$(0, 3, 4, 1, 7, 2, 5, 6)$	$H^{\alpha_2} + \frac{3}{2}H^{\alpha_3} + \frac{1}{2}H^{\alpha_4}$
$(0, 3, 4, 2, 7, 1, 6, 5)$	$H^{\alpha_2} + \frac{3}{2}H^{\alpha_3} + \frac{1}{2}H^{\alpha_4}$
$(0, 3, 5, 2, 7, 4, 1, 6)$	$H^{\alpha_2} + \frac{3}{2}H^{\alpha_3} + \frac{1}{2}H^{\alpha_4}$
$(0, 3, 5, 6, 1, 4, 2, 7)$	$H^{\alpha_2} + \frac{3}{2}H^{\alpha_3} + \frac{1}{2}H^{\alpha_4}$
$(0, 3, 6, 1, 7, 4, 2, 5)$	$H^{\alpha_2} + \frac{3}{2}H^{\alpha_3} + \frac{1}{2}H^{\alpha_4}$
$(0, 3, 6, 5, 2, 4, 1, 7)$	$H^{\alpha_2} + \frac{3}{2}H^{\alpha_3} + \frac{1}{2}H^{\alpha_4}$
$(0, 3, 4, 1, 6, 5, 2, 7)$	$2H^{\alpha_2} + H^{\alpha_3}$
$(0, 3, 4, 2, 5, 6, 1, 7)$	$2H^{\alpha_2} + H^{\alpha_3}$

symbols are α, 2α, 4α, 5α, 8α, 9α, 10α, 13α, and 18α, where α is the minimum SED of 16-QAM. From the Gray and SP symbol mapping, the minimum SEW corresponding to Hamming weight-1 input sequences is α. Like the case of 8-PSK, these symbol mappings are not optimal for BICM. In [7], a different mapper called a modified SP symbol mapper (shown in Fig. 3-(c)) with $(15, 11, 3, 7, 14, 10, 2, 6, 12, 8, 0, 4, 13, 9, 1, 5)$ is applied to 16-QAM BICM systems to provide better performance than the SP mapper. However, this symbol mapper is still not optimal in the sense that the minimum SEW between Hamming weight-1 symbols is not maximized; in this case the minimum SEW is equal to 2α. A MSEW symbol mapper shown in Fig. 3-(d) has been identified through computer search and provides a minimum SEW of 5α for Hamming weight-1 input sequences.

TABLE II
SOME 16-QAM MSEW MAPPERS

16-QAM Mappers	Transfer function terms of L
$(2, 1, 7, 4, 8, 11, 13, 14, 5, 6, 0, 3, 15, 12, 10, 9)$	$2H^{5\alpha} + H^{8\alpha} + H^{13\alpha}$
$(1, 2, 7, 4, 8, 11, 14, 13, 6, 5, 0, 3, 15, 12, 9, 10)$	$2H^{5\alpha} + H^{8\alpha} + H^{13\alpha}$
$(2, 1, 11, 8, 4, 7, 13, 14, 9, 10, 0, 3, 15, 12, 6, 5)$	$2H^{5\alpha} + H^{8\alpha} + H^{13\alpha}$
$(1, 2, 11, 8, 4, 7, 14, 13, 10, 9, 0, 3, 15, 12, 5, 6)$	$2H^{5\alpha} + H^{8\alpha} + H^{13\alpha}$
$(1, 2, 12, 15, 4, 7, 9, 10, 13, 14, 0, 3, 8, 11, 5, 6)$	$2H^{5\alpha} + H^{8\alpha} + H^{13\alpha}$
$(1, 2, 12, 15, 8, 11, 5, 6, 13, 14, 0, 3, 4, 7, 9, 10)$	$2H^{5\alpha} + H^{8\alpha} + H^{13\alpha}$
$(2, 1, 12, 15, 4, 7, 10, 9, 14, 13, 0, 3, 8, 11, 6, 5)$	$2H^{5\alpha} + H^{8\alpha} + H^{13\alpha}$
$(2, 1, 12, 15, 8, 11, 6, 5, 14, 13, 0, 3, 4, 7, 10, 9)$	$2H^{5\alpha} + H^{8\alpha} + H^{13\alpha}$
$(4, 1, 7, 2, 8, 13, 11, 14, 3, 6, 0, 5, 15, 10, 12, 9)$	$2H^{5\alpha} + H^{8\alpha} + H^{13\alpha}$
$(1, 4, 7, 2, 8, 13, 14, 11, 6, 3, 0, 5, 15, 10, 9, 12)$	$2H^{5\alpha} + H^{8\alpha} + H^{13\alpha}$
$(8, 1, 11, 2, 15, 6, 12, 5, 3, 10, 0, 9, 4, 13, 7, 14)$	$\frac{5}{2}H^{5\alpha} + H^{8\alpha} + \frac{1}{2}H^{13\alpha}$
$(8, 2, 11, 1, 15, 5, 12, 6, 3, 9, 0, 10, 4, 14, 7, 13)$	$\frac{5}{2}H^{5\alpha} + H^{8\alpha} + \frac{1}{2}H^{13\alpha}$
$(8, 1, 13, 4, 15, 6, 10, 3, 5, 12, 0, 9, 2, 11, 7, 14)$	$\frac{5}{2}H^{5\alpha} + H^{8\alpha} + \frac{1}{2}H^{13\alpha}$
$(8, 4, 13, 1, 15, 3, 10, 6, 5, 9, 0, 12, 2, 14, 7, 11)$	$\frac{5}{2}H^{5\alpha} + H^{8\alpha} + \frac{1}{2}H^{13\alpha}$
$(8, 2, 14, 4, 15, 5, 9, 3, 6, 12, 0, 10, 1, 11, 7, 13)$	$\frac{5}{2}H^{5\alpha} + H^{8\alpha} + \frac{1}{2}H^{13\alpha}$
$(8, 4, 14, 2, 15, 3, 9, 5, 6, 10, 0, 12, 1, 13, 7, 11)$	$\frac{5}{2}H^{5\alpha} + H^{8\alpha} + \frac{1}{2}H^{13\alpha}$
$(1, 8, 11, 2, 15, 6, 5, 12, 10, 3, 0, 9, 4, 13, 14, 7)$	$\frac{5}{2}H^{5\alpha} + H^{8\alpha} + \frac{1}{2}H^{13\alpha}$
$(15, 8, 6, 1, 4, 3, 13, 10, 9, 14, 0, 7, 2, 5, 11, 12)$	$3H^{5\alpha} + H^{13\alpha}$
$(15, 8, 6, 1, 2, 5, 11, 12, 9, 14, 0, 7, 4, 3, 13, 10)$	$3H^{5\alpha} + H^{13\alpha}$
$(15, 8, 5, 2, 4, 3, 14, 9, 10, 13, 0, 7, 1, 6, 11, 12)$	$3H^{5\alpha} + H^{13\alpha}$
$(15, 8, 5, 2, 1, 6, 11, 12, 10, 13, 0, 7, 4, 3, 14, 9)$	$3H^{5\alpha} + H^{13\alpha}$
$(15, 8, 3, 4, 2, 5, 14, 9, 12, 11, 0, 7, 1, 6, 13, 10)$	$3H^{5\alpha} + H^{13\alpha}$

An exhaustive computer search for 16-QAM MSEW mappers yielded 84 distinct MSEW mappers, where those with rotational symmetries are excluded. Some of the 84 MSEW
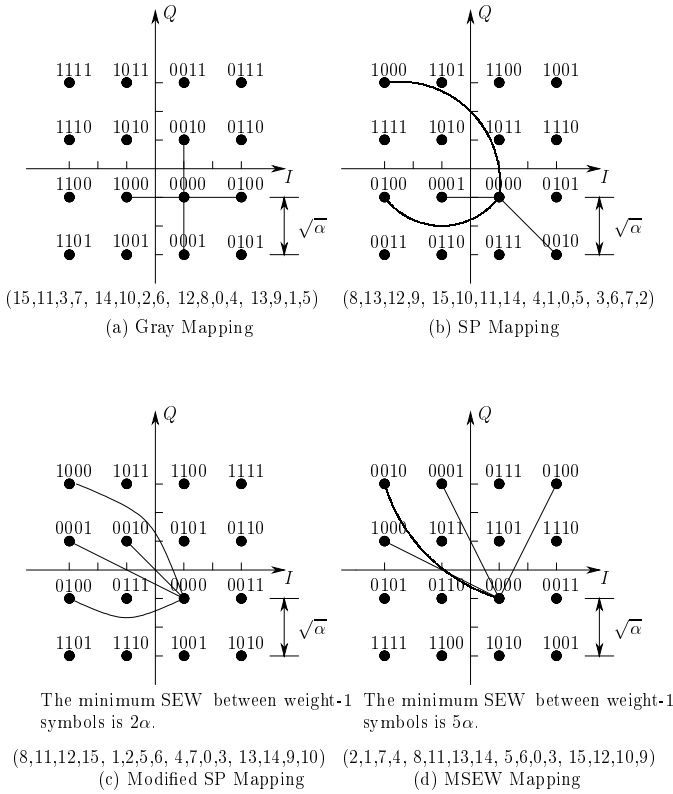
(a) Gray Mapping $(15,11,3,7, 14,10,2,6, 12,8,0,4, 13,9,1,5)$

(b) SP Mapping $(8,13,12,9, 15,10,11,14, 4,1,0,5, 3,6,7,2)$

The minimum SEW between weight-1 symbols is 2α.

(c) Modified SP Mapping $(8,11,12,15, 1,2,5,6, 4,7,0,3, 13,14,9,10)$

The minimum SEW between weight-1 symbols is 5α.

(d) MSEW Mapping $(2,1,7,4, 8,11,13,14, 5,6,0,3, 15,12,10,9)$

Fig. 3.   16-QAM Mappers: Gray, SP, MSP, MSEW.

mappers are listed in Tab. II, together with their transfer function term with input Hamming weight-1. All 84 mappers can be classified into three types according to the multiplicity of the transfer function term with minimum SEW, i.e., the coefficient of $H^{5\alpha}$. Among the 84 MSEW mappers, the minimal coefficient of $H^{5\alpha}$ is 2 and, therefore, the mappers with the term $2LH^{5\alpha}$ in their transfer functions are *optimal* MSEW mappers. There are total of 48 distinct *optimal* MSEW mappers for 16-QAM.

When the constellation size is not very large, such as the 8-PSK 16-QAM examples considered in this paper, an exhaustive computer search can be used to find all MSEW mappers. However, when the constellation size is large such as 64-QAM, an exhaustive search is time consuming although it is feasible if we just need to find one optimal MSEW mapper rather than all of them. However, unlike the Gray or SP mappers, there is no *simple* heuristic approach to generalize the design procedure of MSEW mappers. By using edge profile concepts, it might be possible to obtain general procedure to construct an MSEW mapper for an arbitrary constellation, such as the symmetric-ultracomposite labelling structure [16]. However, the problem is still open.

IV. DENSITY EVOLUTION ANALYSIS OF BICM

To investigate the convergence properties of BICM and to illustrate the effects of the MSEW mapper on BICM performance, density evolution techniques can be used to track the extrinsic information exchange in the iterative demodulation and decoding process of BICM. We use the mutual information

measurement for the extrinsic information transfer (EXIT) chart as proposed by ten Brink [11], [17]. This approach has been applied to the SCCC [18] and BICM in [19].

A. EXIT Characteristics of Mappers

Following the similar approach for EXIT chart in [19], we can obtain the EXIT characteristics for the various mappers using a Monte Carlo approach. We consider several 8-PSK and 16-QAM BICM mappers in our analysis. Mapper A is an arbitary 8-PSK mapper $(0,5,6,3,2,7,4,1)$, and the 8-PSK anti-gray mapper is $(0,7,1,6,2,5,3,4)$, where the representation conventions for 8-PSK is identical to those of the 8-PSK mappers shown in Fig. 2. For the 16-QAM anti-gray mapper, the (-3,-1,+1,+3) amplitudes have mapping $(00, 11, 01, 10)$ on either in-phase and quadrature components [18], resulting in the mapping $(2, 14, 6, 10, 1, 13, 5, 9, 3, 15, 7, 11, 0, 12, 4, 8)$.
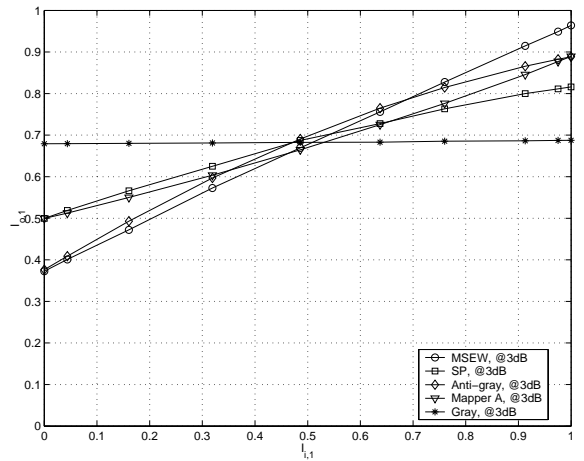

Fig. 4.   EXIT Chart of 8-PSK Mappers on AWGN Channel.

Fig. 4 shows the EXIT characteristics of different 8-PSK mappers. The function $T_m(\cdot)$ is the EXIT characteristics of the SISO demapper, given by $I_{o,1} = T_m(I_{i,1})$, where $I_{i,1}$ and $I_{o,1}$ are the mutual information between the extrinsic information and the transmitted codewords corresponding to the input and output of the demapper, respectively. The Gray mapper's EXIT function is a constant line. In the absences of *a priori* information, i.e., $I_{i,1} = 0$, the Gray mapper has the largest $I_{o,1}$ value among all mappers. This suggests that for the first iteration, Gray mapping provides the best performance. However, with iterations BICM with Gray mapping achieves no performance gain because the mutual information $I_{o,1}$ cannot be improved. As shown in Fig. 4, the EXIT functions of the non-Gray mappers have lower value of $I_{o,1}$ than the Gray mapper at $I_{i,1} = 0$, but they have higher value of $I_{o,1}$ at $I_{i,1} = 1$. A higher value of $I_{o,1}$ implies a lower error probability. When $I_{o,1} = 1$, error-free transmission is achieved.

Fig. 4 indicates the mapper with the largest value of $I_{o,1}$ at $I_{i,1} = 1$ is the MSEW mapper, whose EXIT function has the steepest slope. In our previous discussion, the MSEW mapper provides the best performance at high SNR. That is, at high

SNR, the bit error probability of BICM with ML decoding can be minimized with the MSEW mapper. This indirectly implies that the MSEW mapper can provide the maximum mutual information $I_{o,1}$ at $I_{i,1} = 1$. Although there is no analytical proof of this, our investigation concludes that the MSEW can provide the best mutual information $I_{o,1}$ at $I_{i,1} = 1$.

Note that the MSEW mappers for 8-PSK or 16-QAM are not unique. The EXIT functions of the different MSEW mappers with the same multiplicity have slight differences at one end where $I_{i,1} = 0$. This is due to the difference of their SEW properties. However, at the high end where $I_{i,1} = 1$, the value of their EXIT functions are virtually identical. We use the 8-PSK MSEW mapper shown in Fig. 2-(c) and the 16-QAM MSEW mapper shown in Fig. 3-(d).

B. EXIT Chart of BICM

The MSEW mapper can provide a large value of $I_{o,1}$ at $I_{i,1} = 1$. However, this does not guarantee that BICM can reach its potential performance through iterative decoding and demodulation. The iterative decoding process depends not only on the EXIT characteristics of the mapper, but also on the EXIT characteristics of the outer code. It is desirable that the outer code should provide a "tunnel" at the lowest possible channel SNR so that the maximum $I_{o,1}$ can be achieved through iterations at lowest channel SNR.

There are two major properties of the EXIT characteristics of the outer code. First, the EXIT function of the outer code does not depend on the channel SNR. Secondly, the EXIT characteristics of the outer code can always achieve $I_{o,2} = 1$ when $I_{i,2} = 1$, where the error-free transmission is possible. Further, when $I_{i,2} = 0$ the output mutual information $I_{o,2} = 0$. Because the mapper EXIT function $T_m(\cdot)$ usually cannot reach $(1,1)$ at the upper-left corner, the two EXIT functions will intersect near the $(1,1)$ point. The intersection point is the end of the iterating "tunnel", and determines the error floor of BICM at a given channel SNR. When the channel SNR is large, the mapper EXIT function $T_m(\cdot)$ is shifted upwards and it can achieve higher $I_{o,1}$, resulting in a smaller error probability. For the outer code, it is desirable to provide a high value of $I_{o,2}$ for a modest value of $I_{i,2}$. Usually a convolutional code with a larger free distance will reach a larger value of $I_{o,2}$ at the high end of $I_{i,2}$. Therefore, a convolutional code with longer constraint length can usually provide a lower error floor than one with a shorter constraint length.

The MSEW mapper provides the largest value of $I_{o,1}$ at the high end of $I_{i,1}$, which means that the error floor of BICM with the MSEW mapper should be the lowest. However, because the lower end has a small value of $I_{o,1}$, the MSEW EXIT function may intersect to the outer code EXIT function and close the iterative "tunnel" after a few iterations. One instance is shown in Fig. 5, where BICM with MSEW mapper fails because the iteration converges at the intersecting point. But the other mapper with a higher value of $I_{o,1}$ at the low end of $I_{i,1}$ can start the iteration process and reach its corresponding $I_{o,1}$ at the high end of the mapper EXIT function. Hence, this mapper can outperform the MSEW mapper, even though its potential error floor is worse than the MSEW mapper. This

indicates that for a given outer code, the MSEW mapper is not always the best mapper; it is just the mapper with best potential asymptotic performance.
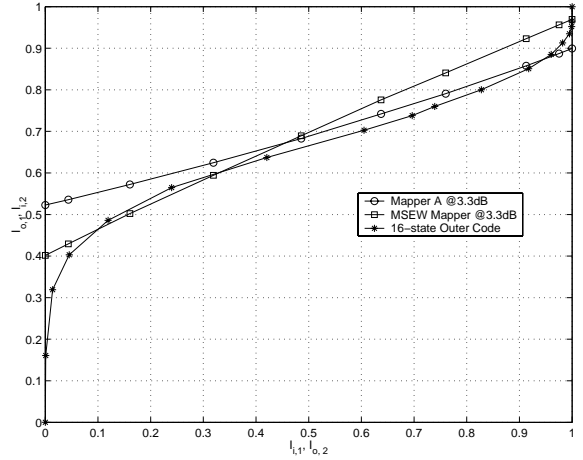

Fig. 5. EXIT Chart of BICM.

To achieve the best performance with BICM, the mapper and the outer code should be considered together. For a given outer code the best mapper is not necessary the MSEW mapper, but rather the mapper for which EXIT function "matches" the outer code EXIT function. For a given mapper, the optimal outer code should provide an iteration "tunnel" at lowest possible channel SNR and make the intersection point of their EXIT functions as close as possible to the $(1,1)$ point in the $I_{i,1} \times I_{o,1}$ space. Based on the EXIT chart of the BICM, we can summarize the design criteria for BICM as follows.

1) (Minimizing error floor) In order to give the lowest asymptotic error probability, the mapper with maximum value of output mutual information $I_{o,1}$ at the high end of input mutual information $I_{i,1}$ should be used. This mapper is the MSEW mapper.
2) (Minimizing threshold channel SNR) In order to achieve the best iterative performance, the EXIT function of the mapper should match the EXIT function of the outer code at the lowest possible channel SNR. This mapper may or may not be the MSEW mapper.

C. MSEW Mapper Doping with Gray Mapper

Compromise is needed at low channel SNR when the iterative decoder for BICM with the MSEW mapper converges at early stages. Usually, we select another mapper with a higher output mutual information $I_{o,1}$ at the lower end of its EXIT function $T_m(\cdot)$ to *match* the outer code. For different outer codes, the selection of the matching mappers may be different which leads to some inconvenience in the mapper design.

An easier approach for changing the slope of the MSEW EXIT function is to combine or "dope" the MSEW mapper with other mappers. This idea is motivated by the "doping" concept that was introduced to achieve better binary turbo codes in [9], [12]. Since different mappers have different slopes and their EXIT functions are almost straight lines,

combining two mappers will surely change the slope of the mapper. By combining we mean that some of the symbol mappings are changed in the transmitted block. In one block there are two mapping schemes, one is the original mapper, and the other one is a "doping" mapper. Since the MSEW mapper has the largest slope of its EXIT function, and the Gray mapper's EXIT function is a constant line, we can use the MSEW mapper and dope with some Gray mappings. The doping rate $\eta \overset{\triangle}{=} \frac{N_d}{N}$ is defined as the ratio of the number of the doping symbol mappings, N_d, and the block length, N. By controlling the doping rate, the EXIT of the doped mapper exhibits different slopes and lies between the Gray mapper and the MSEW mapper. Fig. 6 shows the EXIT functions of the doped 8-PSK mappers with different doping rates.
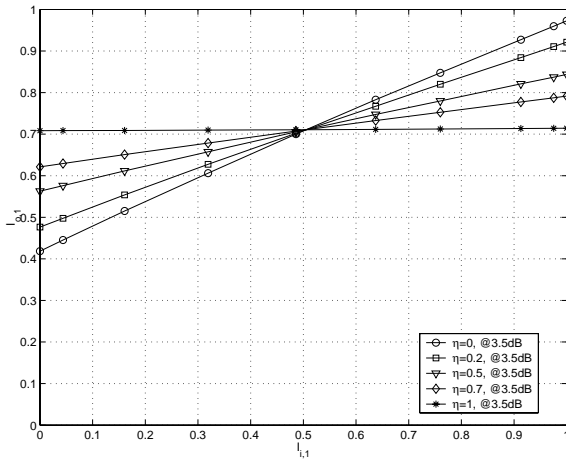

Fig. 6. EXIT Chart for doped 8-PSK Mappers.

The doping technique provides an easy approach to generate mappers with almost arbitrary EXIT function slopes. This approach is very effective for matching the mapper to the outer code. Simulations can show that different mappers, including the doped mapper, that have the same EXIT function slope will provide similar performance.

V. SIMULATION RESULTS AND DISCUSSION

In our simulation, a 16-state rate $1/2$ convolutional code with generators $(46, 72)$ [20] [1] is used as the outer code. To achieve rate $2/3$ or rate $3/4$, puncturing is used with puncture patterns [21]

$$\text{Rate } 2/3: \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{Rate } 3/4: \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}. \quad (19)$$

A. 8-PSK-BICM Performance on AWGN Channels

The BER performance of convolutionally coded 8-PSK BICM with block length $N = 4000$ information bits and different symbol mappings is shown in Fig. 7 for an AWGN channel. From Fig. 7, the MSEW symbol mapping with doping rate $\eta = 0.2$ outperforms both the SP and the Mapper A symbol mappings at high E_b/N_o, which confirms our analysis with

[1]This code is exactly the same code as the one with generators $(23, 35)$ in [21]

EXIT charts. At low E_b/N_0, SP mapping performs better than the doped MSEW mapper with doping rate 0.2. Increasing the doping rate 0.2 to 0.5 can improve the low SNR performance of the MSEW mapper, but sacrifices the error floor. Although not shown in Fig. 7, we note that a Gray symbol mapper will outperform both the SP and MSEW mappers for non-iteratively decoded BICM. However, for iteratively decoded BICM no performance gain will be realized through further iterations with a Gray symbol mapper, and the other mappers will perform better.
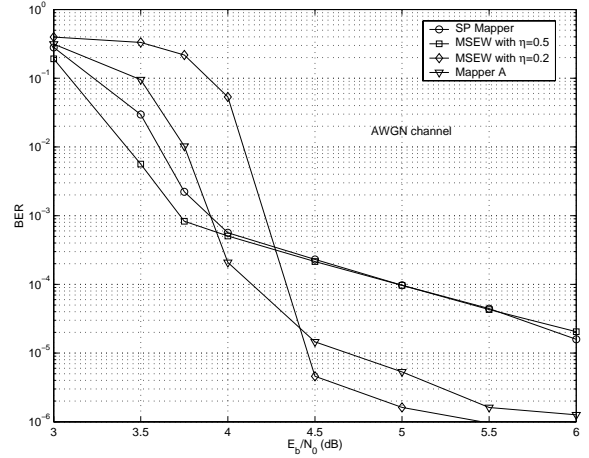

Fig. 7. BER performance of rate-2/3, 16-state, convolutionally coded 8-PSK BICM with different symbol mappings on an AWGN channel; N=4000, 10 iterations.

B. 16-QAM-BICM Performance on Rayleigh Fading Channel

Simulation results for the 16-QAM BICM system with rate $3/4$ coding and 10 iterations are shown in Fig. 8. Among the four different mappers, MSP, SP, anti-Gray, and MSEW mapper, the MSEW mapper provides the best performance at high SNR. This confirms our analysis that MSEW mapper provides the smallest error floor at high SNR. At low SNR the anti-Gray mapper has the best performance because its mutual information has highest value at the lower end of its EXIT function. But its performance at high SNR is similar to that of the SP mapper and worse than that of the MSEW and MSP mappers. Again, the low SNR performance of the MSEW and MSP mappers can be improved by using doping with some sacrifice of the error floor at high SNR.

C. Effects of outer code

Different outer codes may require different *matching* mappers to achieve the desired performance. Fig. 9 shows simulation results for the rate $1/2$ 16-QAM BICM with 10 iterations. Of the three mappers, the MSEW mapper is still the best mapper at high SNR. The performance of the MSP mapper is very close to that of the MSEW mapper, and it has better performance at low SNR than the MSEW mapper. The anti-Gray mapper provides good performance at low SNR, and similar performance can be obtained by doping the MSEW mapper with doping rate 0.3, as shown in Fig. 9. Comparing Figs. 9 and 8 shows the diminished error floor of a code with larger free distance, d_f.
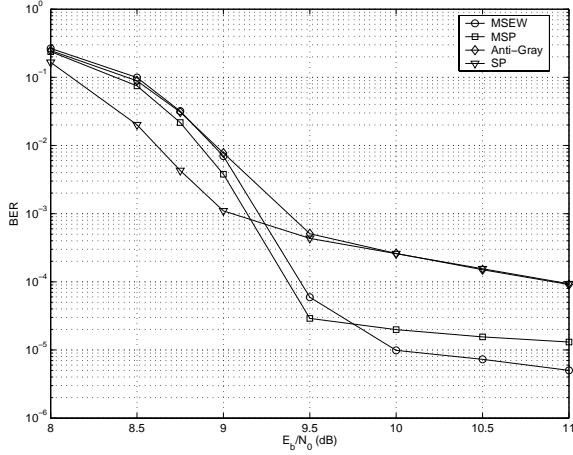
Fig. 8. BER performance of rate-3/4, 16-state, convolutionally coded 16-QAM BICM with different symbol mappings on Rayleigh fading channel; N=4000, 10 iterations.
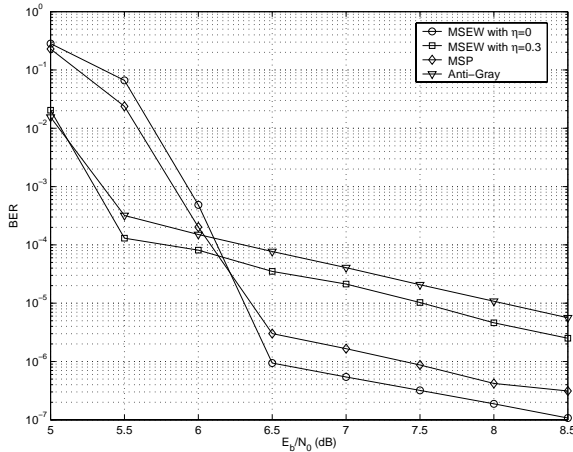

Fig. 10. BER performance of rate-2/3, 16-state, convolutionally coded 8-PSK on correlated fading channels; $N = 4000$, 10 iterations.

E. Comparison with Turbo-coded BICM

This paper considers BICM with convolutional coding. More sophisticated codes, like turbo codes or LDPC codes can be used instead of the convolutional code. A good discussion of this is provided in [22]. When a turbo code or LDPC code is used, the Gray mapper provides the best performance. Iterative decoding for turbo or LDPC codes can provide a large performance gain. However, there is little gain in using iterative demapping and decoding. With iterative demapping/decoding, convolutional coded BICM provides comparable performance to that of turbo or LDPC coded modulation over both AWGN and fading channels for moderate block sizes considered in this paper, e.g., $n = 4000$ [22]. However, as $N \rightarrow \infty$ the turbo or LDPC coded systems will be better because the convolutional coded BICM with iterative demapping/decoding does not have any interleaving gain as $N \rightarrow \infty$. Convolutional coded BICM is superior to turbo or LDPC coded modulation in terms of implementation complexity. This is due to the simple trellis structure of the BICM symbol mapper and the relatively simple structure for soft-input soft-output decoding of the convolutional code.

Fig. 9. BER performance of rate-1/2, 16-state, convolutionally coded 16-QAM BICM with different symbol mappings on Rayleigh fading channel; N=4000, 10 iterations.

VI. CONCLUSIONS

Analysis and design of iteratively decoded BICM is presented in this paper. The ML bit error probability bound for BICM is obtained and analysis of the error probability bound leads a new symbol mapper called the MSEW mapper. The iterative decoding and demodulation process is investigated using EXIT chart to determine the role of the mapper. Our major conclusions are as follows:

- At high SNR, where the BICM performance is dominated by its minimum SEW, an *interleaving gain* can be achieved. However, due to the non-recursive property of the mapper, there is no interleaving gain when the block length $N \rightarrow \infty$. However, for a small block length, N, increasing N can achieve better performance at high SNR. When N is "large" enough, no further interleaving gain will be achieved. Usually this limiting block length is very large and it depends on the channel SNR.

D. Correlated Fading Channel

As discussed earlier, the interleaver between the outer code and the symbol mapper serves as a code interleaver in the serial concatenated code scheme. On fading channels, the interleaver also serves as a channel interleaver provided that the interleaver size is large enough. This is because the soft output of the demapper depends on one received symbol only at one time, and it does not depend on the previous and later symbols in the sequence. At the output of the SISO demapper, the soft outputs are then de-interleaved so that input of the outer decoder becomes uncorrelated and the channel resembles a fully interleaved fading channel. Applying a channel interleaver after the symbol mapper with BICM is unnecessary and does not further improve the performance, provided of course that the block length is sufficiently large (as is the case in our example). Fig. 10 shows the performance of the BICM on a correlated Rayleigh fading channel.
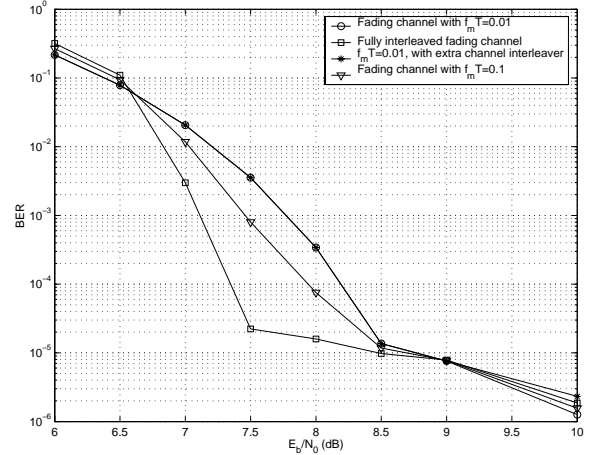
- The MSEW (*maximum squared Euclidean weight*) mapper is defined as the mapper which has the maximum SED for any Hamming weight-1 symbol pairs. The optimal MSEW mapper is defined as the MSEW mapper to have the minimum number of symbol pairs which have the minimum SED with Hamming distance-1. There are 12 and 48 optimal MSEW mappers for 8-PSK and 16-QAM, respectively. The optimal MSEW mapper can achieve better performance at high SNR than other mappers.

- The EXIT function of the MSEW mapper has the maximum mutual information value when the input mutual information $I_{i,1} = 1$. This generally indicates that the EXIT function of the MSEW mapper has steepest slope, and thus can provide the best potential performance for BICM.

- The MSEW mapper is not always the best mapper. For a given outer code, the optimal mapper is the mapper which *matches* the EXIT function of the outer code in the EXIT chart. The MSEW mapper can be doped with the Gray mapper to provide a good *match* to a given outer code. The doping technique is a practical and flexible scheme to offer good performance given a outer code and a channel SNR.

DISCLAIMER

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U. S. Government.

REFERENCES

[1] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. on Information Theory*, vol. IT-28, no. 1, pp. 55–67, Jan. 1982.
[2] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. on Communications*, vol. 40, no. 5, pp. 873–884, May 1992.
[3] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. on Information Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
[4] S. Benedetto, G. Montorsi, D. Divsalar, and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design and iterative decoding," JPL, TDA Progress Report 42-126, August 1996.
[5] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Communications Letters*, vol. 1, no. 6, pp. 169–171, Nov. 1997.
[6] X. Li, A. Chindapol, and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding and 8PSK signaling," *IEEE Trans. on Communications*, vol. 50, no. 8, pp. 1250–1257, Aug. 2002.
[7] A. Chindapol and J. A. Ritcey, "Design, analysis, and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels," *IEEE J. on Selected Areas in Communications*, vol. 19, no. 5, pp. 944–957, May 2001.
[8] X. Li and J. A. Ritcey, "Trellis-coded modulation with bit interleaving and iterative decoding," *IEEE J. on Selected Areas in Communications*, vol. 17, no. 4, pp. 715–724, Apr. 1999.
[9] D. Divsalar, S. Dolinar, and F. Pollara, "Iterative turbo decoder analysis based on density evolution," *IEEE J. on Selected Areas in Communications*, vol. 19, no. 5, pp. 891–907, May 2001.
[10] H. E. Gamal and J. Hammons, "Analyzing the Turbo decoder using the Gaussian approximation," *IEEE Trans. on Information Theory*, vol. 47, no. 2, pp. 671–686, Feb. 2001.
[11] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. on Communications*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.
[12] ——, "Code doping for triggering iterative decoding convergence," in *IEEE Int. Symp. on Inform. Theory*, Washington, DC, June 2001, p. 235.
[13] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "A soft-input soft-output maximum a posteriori (MAP) module to decode parallel and serial concatenated codes," JPL, TDA Progress Report 42-127, November 1996.
[14] S. Benedetto and E. Biglieri, *Principles of Digital Transmission with Wireless Applications*. Kluwer, 1999.
[15] S. Benedetto and G. Montorsi, "Design of parallel concatenated convolutional codes," *IEEE Trans. on Communications*, vol. 44, no. 5, pp. 591–600, May 1996.
[16] R. D. Wesel, X. Liu, J. M. Cioffi, and C. Komninakis, "Constellation labeling for linear encoder," *IEEE Trans. on Information Theory*, vol. 47, no. 6, pp. 2417–2432, Sept. 2001.
[17] S. ten Brink, "Convergence of iterative decoding," *Electronics Letters*, vol. 35, no. 13, pp. 1117–1118, June 1999.
[18] M. Tüchler, S. ten Brink, and J. Hagenauer, "Measures for tracing convergence of iterative decoding algorithms," in *Proc. 4th International ITG Conference on Source and Channel Coding*, Berlin,Germany, Jan. 2002.
[19] S. ten Brink, "Designing iterative decoding schemes with the extrinsic information transfer chart," *AEÜ International Journal of Electronics and Communications*, vol. 54, no. 6, pp. 389–398, Dec. 2000.
[20] S. Lin and J. Daniel J. Costello, *Error Control Coding, Fundamentals and Applications*. Prentice-Hall,Inc., 1983.
[21] J. G. Proakis, *Digital Communications*, 4th ed. McGraw-Hill, 2000.
[22] J. Tan, "Iterative decoding and multicarrier modulation for wireless communications," Ph.D. dissertation, Georgia Institute of Technology, Atlanta, GA, August 2003.

Jun Tan received his B.S.E.E and M.S.E.E degree from University of Electronic Science and Technology of China, Chengdu, China in 1988 and 1991, and his Ph.D in Electrical Engineering in 2003 from Georgia Institute of Technology, Atlanta, Georgia, USA.

From 1991 to 1996 he was a Lecturer and an Assistant Professor in computer engineering at University of Electronic Science and Technology of China. During 1996 to 1998 he was with Tennessee Technology University, Cookeville, TN, as a research associate. From 1998 to 2003, he was a Research Assistant and Teaching Assistant at Wireless Systems Laboratory, Georgia Institute of Technology, Atlanta. His current research interests lie in the area of communication theory, applied to modulation and coding for wireless systems. Specific topics include turbo/iterative coding, multicarrier modulations, space-time coding, spread spectrum, and wireless system design.

Gordon L. Stüber received the B.A.Sc. and Ph.D. degrees in Electrical Engineering from the University of Waterloo, Ontario, Canada, in 1982 and 1986 respectively. Since 1986, he has been with the School of Electrical and Computer Engineering, Georgia Institute of Technology, where he is currently the Joseph M. Pettit Professor in Communications.

Dr. Stüber's research interests are in wireless communications and communication signal processing. He is author of the textbook Principles of Mobile Communication, Kluwer Academic Publishers, 1996, 2/e 2001. He was co-recipient of the Jack Neubauer Memorial Award in 1997 recognizing the best systems paper published in the IEEE Transactions on Vehicular Technology. He became an IEEE Fellow in 1999 for contributions to mobile radio and spread spectrum communications. He received the IEEE Vehicular Technology Society James R. Evans Avant Garde Award in 2003 recognizing his contributions to theoretical research in wireless communications.

Dr. Stüber served as Technical Program Chair for the 1996 IEEE Vehicular Technology Conference (VTC'96), Technical Program Chair for the 1998 IEEE International Conference on Communications (ICC'98), General Chair of the Fifth IEEE Workshop on Multimedia, Multiaccess and Teletraffic for Wireless Communications (MMT'2000), General Chair of the 2002 IEEE Communication Theory Workshop, and and General Chair of the Fifth International Symposium on Wireless Personal Multimedia Communications (WPMC'2002). He is a past Editor for Spread Spectrum with the IEEE Transactions on Communications (1993-1998), and a past member of the IEEE Communications Society Awards Committee (1999-2002). He is currently an elected member of the IEEE Vehicular Technology Society Board of Governors (2001-2003).