# Code Design for Iterative Decoding of Multilevel Codes

Yi Wang and Alister G. Burr, *Member, IEEE*

***Abstract*—The code design problem for multilevel coded modulation with iterative decoding (MLCM-ID) has been left unsolved over a decade. In this paper, we define the code design criterion for MLCM-ID based on a novel concept—parametrically-mapped EXIT (PM-EXIT) function. We derive the EXIT functions for MLCM-ID, and present the mathematical work by which the three area theorems of PM-EXIT function are proved. This gives theoretical support of the curve-fitting techniques used in MLCM-ID, and also provides firm proof that the PM-EXIT function design rule allows in principle capacity-achieving code assignment. The simulation results perfectly match the theoretical analysis, and also confirm that the design rule proposed fully exploits the flexibility of MLCM-ID in the choice of codes. Both in theory and simulations, our code design is verified.**

***Index Terms*—Multilevel codes, iterative decoding, EXIT functions.**

## I. Introduction

THE fundamental principle of a bandwidth-efficient multilevel coded modulation (MLCM) scheme is to use different rate binary component codes for the protection of each bit level. It has been demonstrated [1] that the channel capacity can be achieved by multilevel coded modulation and overall maximum-likelihood decoding (MLD). However, MLCM-MLD is unfeasible due to the prohibitive computational costs. A good tradeoff between capacity-achieving performance and decoding complexity is multistage decoding [2]–[4], which is in principle capable of achieving capacity if the code rates selected for each level approach the respective equivalent channel capacity [1]. An alternative [5], [6] to MLCM-MSD is MLCM-ID which employs iterative decoding to reduce the error multiplicity in a probabilistic sense. In comparison to MLCM-MSD, MLCM-ID allows simpler codes to be used in achieving the constellation constrained capacity (CCC), i.e., the maximum capacity of a system with a specific constellation of the transmitted symbols, and also more flexible code choice at each level. We refer to [5], [7] for the detailed discussions.

Manuscript received October 30, 2014; revised April 4, 2015; accepted May 15, 2015. Date of publication June 1, 2015; date of current version July 13, 2015. The work described in this paper was supported by the European Commission Framework 7 Programme under grant agreement 318177 (DIWINE). The associate editor coordinating the review of this paper and approving it for publication was L. L. Szczecinski.

The authors are with the Department of Electronics, University of York, York YO10 5DD, U.K. (e-mail: yi.wang@york.ac.uk; alister.burr@york.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2015.2438067

### A. Problem Statements

For conventional MLCM, Wachsmann *et al.* have shown some feasible code design rules in their landmark paper [1]; they are, the capacity design rule, balanced distance rule, coding exponent rule, cut-off rate rule and equal error probability rule. All these rate design rules restrict the maximum rate for each level. If the code rate assigned at each level exceeds their corresponding maximum rate, it is in principle impossible to achieve reliable transmission over an AWGN channel. For MLCM-ID, the authors have shown [5] that one merit of MLCM-ID over MLCM is that it allows a wider choice of codes, which increases the flexibility in the design of code. However Isaka and Imai [5] also pointed out that code design for MLCM-ID is not easy. The principal constraints are:

1) The convergence behaviour of iterative decoding has to be evaluated in multidimensional space [5].
2) The code design rule (CDR) should allow a wider class[1] of code choice at each level.
3) The overall code rates should approach CCC as closely as possible. We expect that the code design rule allows in principle capacity lossless design[2] for MLCM-ID.
4) The allocated code rates must be limited to the rate bounding region of the constellation mapping, as described in Condition 2 of Theorem 2 in [1].
5) The rate and EXIT property of the code must be jointly considered to achieve full convergence.

Constraint 1 mentioned in [5] imposes the main technical difficulty in the design of codes for MLCM-ID. Constraint 2 follows from the fact that MLCM-ID allows more flexible code choice at each level compared to the non-iterative case [5], and hence the code design rule should be subject to this constraint. The CDR should be also subject to constraint 3 to ensure that the overall rate approaches the capacity. Furthermore, the CDR should be subject to constraint 4 where the rates assigned at each level must be limited to the theoretical rate region given in [1]. Constraint 5 follows from the basic principle of the EXIT chart curve-fitting technique [8].

As far as the authors know, the code design problem for MLCM-ID is still open, and the only approach so far available is that given in [9] which is based on a set of three dimensional (3D) EXIT charts [10] (which is a special case of the code design proposed here), giving only a single solution to the code design problem. Note that a multi-dimensional EXIT

[1] A wider range of selection of code rates and types.

[2] The CDR should allow capacity to be achieved.

chart analysis has previously been proposed for the convergence analysis of multiple concatenated turbo codes [11], [12] which uses multiple parallel concatenated component codes. Exploiting this concept, we specifically design a different 3D EXIT chart for MLCM-ID. The design approach in [9] does not exploit the full potential of flexible code selection in MLCM-ID. Therefore, a general and efficient code design rule for MLCM-ID has not been solved.

### B. Contributions

The aim of this paper is to introduce an easy-to-use code design rule for MLCM-ID, which fully exploits the flexibility in the choice of codes, allows theoretically the overall rate achieving CCC, and also has full convergence guaranteed. The contributions of this paper are summarized as:

1) We introduce the novel concept of the parametrically-mapped EXIT (PM-EXIT) function which resolves a set of complex multidimensional EXIT functions into separate 2D EXIT functions.
2) We propose a code design theorem and define a code design rule for MLCM-ID. The proposed code design rule relaxes the five aforementioned constraints, and now the complex code design problem reduces to a simple parametric function selection problem.
3) We derive analytically a general form of the multi-variate bit level extrinsic information transfer (MV-BL-EXIT) function for multilevel codes based on the *a priori* binary erasure channel (BEC).
4) We prove three area theorems based on the analytical results, which gives theoretical support for our code design rule, and also prove that the PM-EXIT function rule allows in principle capacity lossless multilevel code design.
5) We verify the code design rule and theorems presented, further by means of simulations.
6) We show the possible parametric function selections in the design of the PM-EXIT functions.

### C. Notation Definitions

We first define some notations used throughout this paper: $\mathbb{C}$ and $\mathbb{R}$ denote the set of complex numbers and real numbers, respectively. $\boldsymbol{a}_{\setminus i} \triangleq [a_0, \cdots, a_{i-1}, a_{i+1}, \cdots, a_{m-1}]$ for a length-$(m-1)$ vector; $\dot{\mathbb{F}}_2 \triangleq \{0, 1\}$, $\dot{\mathbb{F}}_2 \in \mathbb{Z}$; $\mathbb{F}_2$ denotes a binary finite field. The bold italic $\boldsymbol{A}$ represents the vector-valued random variables. The lowercase $a$ or $\boldsymbol{a}$ represents the scalar or vector realization of random variable $A$ or $\boldsymbol{A}$. $f'(\cdot)$ represents the first derivative of function $f(\cdot)$.

Section II briefly describes the MLCM-ID structure and the operation of the demodulator/decoder. In Section III, we derive the MV-BL-EXIT functions for multilevel codes based on the *a priori* BECs. Section IV states and proves, with the aid of the 3D EXIT chart, the code design theorem on the conditions for convergence of MLCM-ID. Area theorems are proved in Section V, which confirms that in principle CCC can be achieved based on MLCM-ID and the code design rule proposed. Section VI first investigates the influence of the mapping in MLCM-ID and then deals with the simulation examples and analysis.

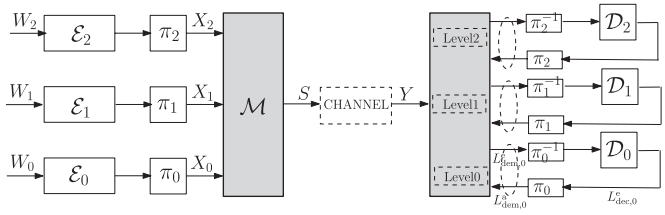



Fig. 1. System model of MLCM-ID.

## II. System Model

In this section, the system structure and two iterative decoding modes of MLCM-ID are described. At the transmitter, information bits of the $i^{\text{th}}$ level enter the corresponding encoder $\mathcal{E}_i$, $i = 0, 1 \cdots, m-1$, and these coded bits are then interleaved by a random interleaver $\pi_i$, where $m = \log_2(M)$ and $M$ is the number of constellation points. For simplicity, a 3-level MLCM-ID structure is illustrated in Fig. 1. The interleaved bits $X_i$ form binary address vectors $\boldsymbol{X} = [X_0, X_1, \ldots, X_{m-1}]$, $x_i \in \mathbb{F}_2$, $\boldsymbol{x} \in \mathbb{F}_2^m$ which are mapped to the complex symbol $S = \mathcal{M}(\boldsymbol{X})$ from a $2^m$-ary signal constellation $\zeta$, based on a bijective mapping function $\mathcal{M} : \boldsymbol{X} \to S$. The symbol $S$ is transmitted over the complex AWGN channel. We denote $\mathbb{X}_{\setminus i}$ as the powerset of the index set $\mathcal{I}$ of $\boldsymbol{X}_{\setminus i}$. At the receiver, given the matched filter output $Y$, the demodulator calculates the *extrinsic* L-value for the $i^{\text{th}}$ level which is then fed into the corresponding decoder $\mathcal{D}_i$. If no *a priori* knowledge of the other bits is available, the posterior probability for $X_i = x_i$, at the $i^{\text{th}}$ level is given by

$$\Pr(X_i = x_i|Y) = \sum_{s \in \zeta_{x_i}^i} \Pr(S = s|Y) = \frac{1}{2^{m-1}} \sum_{s \in \zeta_{x_i}^i} \frac{\Pr(Y|S = s)}{\Pr(Y)} \quad (1)$$

where $\zeta_{x_i}^i$ denotes the subset of symbol set $\zeta$ whose bit labels have the value $X_i = x_i$ at the $i^{\text{th}}$ position; e.g., if $x_i = 1$, then $\zeta_{x_i}^i = \{S = \mathcal{M}(\boldsymbol{X}) \mid \ \forall \boldsymbol{x} \in \mathbb{F}_2^m, \ x_i = 1\}$. We also denote $s^\ell \in \mathbb{F}_2$, $\ell = 0, 1, \cdots, m-1$, the bit value at the $\ell^{\text{th}}$ level of the constellation point $s$. Then, in the case where the *a priori* knowledge of other bits is available:

$$\Pr(X_i = x_i|Y) = \sum_{s \in \zeta_{x_i}^i} \frac{P(Y|S = s) \prod_{\ell=0, \ell \neq i}^{m-1} \Pr_{\mathrm{a}}(X_\ell = s^\ell)}{P(Y)} \quad (2)$$

The decoding modes available are iteration-aided parallel independent decoding (IA-PID) and multistage decoding (IA-MSD). For the first mode, the demodulator calculates the extrinsic probability for each level simultaneously. Then, with the aid of the output L-value, $L^{\mathrm{e}}_{\mathrm{dem},i} = \log_2(P_{\mathrm{e}}(X_i = 1)/P_{\mathrm{e}}(X_i = 0))$, each decoder computes the new metrics $L^{\mathrm{e}}_{\mathrm{dec},i}$ independently and simultaneously, which in turn serve as the *a priori* L-value for the demodulator. Here, $L^{\mathrm{a}}_{\cdot,\cdot}$ and $L^{\mathrm{e}}_{\cdot,\cdot}$ represent the *a priori* and *extrinsic* L-value of the corresponding decoder/demodulator at

a particular level. An alternative is multistage decoding. We assume that multistage decoding proceeds from the lowest level to the highest level. At the first stage of the first iteration, the demodulator first calculates the extrinsic probability of the code bit at level 0 using equation (1), where the $m-1$ *a priori* L-value, $[L^{\mathrm{a}}_{\mathrm{dem},1}, \cdots, L^{\mathrm{a}}_{\mathrm{dem},m-1}]$ are zeros at the first stage. The output $L^{\mathrm{e}}_{\mathrm{dem},0}$ is deinterleaved and passes through the associated decoder 0 to produce $L^{\mathrm{e}}_{\mathrm{dec},0}$. The updated $L^{\mathrm{a}}_{\mathrm{dem},0}$ along with $[L^{\mathrm{a}}_{\mathrm{dem},2} \cdots, L^{\mathrm{a}}_{\mathrm{dem},m-1}]$ update $L^{\mathrm{e}}_{\mathrm{dem},1}$. This process is repeated and all levels are activated in turn for the second and subsequent iterations.

## III. Multi-Variate Bit-Level Extrinsic Information and Combined Information

The EXIT chart is a powerful technique which is applied for the convergence analysis of iterative systems. The EXIT function is commonly obtained as shown in [10], where the *a priori* channel was assumed to be the Binary-Input AWGN (BiAWGN) channel. EXIT functions were also investigated based on the other *a priori* channel models, such as BEC and binary symmetric channel (BSC). For example, by exploiting the concept of information combining [13], the authors in [13], [14] derived the EXIT functions for single parity check codes and repetition codes based on the *a priori* BEC and BSC, respectively. In [15], based on the BEC, the authors proved the area properties of the EXIT functions for several classes of codes,[3] giving theoretical support for the curve fitting technique.

In this section, we derive the EXIT function for a particular bit level of multilevel codes based on the *a priori* BEC, where all level's code bits are coupled by the constraint of complex constellation mapping, rather than the single parity check or repetition constraints. Due to the constellation mapping constraint, the extrinsic information for the $i^{\mathrm{th}}$ code bit depends on the channel observation, and the $m-1$ *a priori* information $\boldsymbol{I}^{\mathrm{a}}_{\mathrm{dem},\setminus i} = [I^{\mathrm{a}}_{\mathrm{dem},0}, \cdots, I^{\mathrm{a}}_{\mathrm{dem},i-1}, I^{\mathrm{a}}_{\mathrm{dem},i+1}, \cdots, I^{\mathrm{a}}_{\mathrm{dem},m-1}]$, where $I^{\mathrm{a}}_{\mathrm{dem},i}$ denotes the *a priori* information of the $i^{\mathrm{th}}$ level of the demodulator.

Let all *a priori* channels for coded bits $\boldsymbol{X}_{\setminus i}$ be independent BECs, with erasure probabilities $\boldsymbol{\delta}_{\setminus i}$. $X_i$, $i \in \{1, 2, \cdots, m-1\}$ is either perfectly received or erased. We define $V_j$ as a discrete random variable with Bernoulli distribution,

$$V_j = \begin{cases} 1 & \text{if } X_j \text{ is perfectly received: probability } 1-\delta_j \\ 0 & \text{if } X_j \text{ is erased: probability } \delta_j \end{cases}$$

The probability function of $V_j$ is $p(v_j) = P(V_j = v_j) = \delta_j^{1-v_j}(1-\delta_j)^{v_j}, j = 0, \cdots, m-1, j \neq i$. Each random variable in $\boldsymbol{V}_{\setminus i} \in \dot{\mathbb{F}}_2^{m-1}$ is assumed to be independent due to the independence of the *a priori* BECs, and their multivariate distribution is:

$$p_{\boldsymbol{V}_{\setminus i}}(\boldsymbol{v}_{\setminus i}) = P(\boldsymbol{V}_{\setminus i} = \boldsymbol{v}_{\setminus i}) = \prod_{j=0,\cdots,m-1, j\neq i} \delta_j^{1-v_j}(1-\delta_j)^{v_j} \quad (3)$$

where $\dot{\mathbb{F}}_2^{m-1}$ is a set including all $2^{m-1}$ possible realizations for vector-valued random variable $\boldsymbol{V}_{\setminus i}$. Each element in $\dot{\mathbb{F}}_2^{m-1}$ can be mapped to the corresponding element of the power-set $\mathbb{X}_{\setminus i}$ under a mapping function $F(\cdot)$ such that $F(\boldsymbol{v}_{\setminus i}) \in \mathbb{X}_{\setminus i}$ (note that $F(\cdot)$ is both one-to-one and onto). For example, given $m = 3$ and $i = 0$, $F(v_1 = 1, v_2 = 0) := F(V_1 = v_1, V_2 = v_2)_{|v_1=1, v_2=0} \rightarrow X_1$.

The capacity of the BEC depends linearly on the erasure probability [16], $I^{\mathrm{a}}_{\mathrm{dem},j} = 1 - \delta_j$. Hence the joint probability function in (3) becomes:

$$\begin{aligned} p_{\boldsymbol{V}_{\setminus i}}(\boldsymbol{v}_{\setminus i}) &= P(\boldsymbol{V}_{\setminus i} = \boldsymbol{v}_{\setminus i}) \\ &= \prod_{j=0, j\neq i}^{m-1} \left(1 - I^{\mathrm{a}}_{\mathrm{dem},j}\right)^{1-v_j} \left(I^{\mathrm{a}}_{\mathrm{dem},j}\right)^{v_j} \end{aligned} \quad (4)$$

The extrinsic information $I^{\mathrm{e}}_{\mathrm{dem},i}$ for the $i^{\mathrm{th}}$ level of multilevel codes is the expected value of the conditional mutual information $I(Y; X_i|F(\boldsymbol{V}_{\setminus i}))$ with respect to the joint probability function of $\boldsymbol{V}_{\setminus i}$

$$I^{\mathrm{e}}_{\mathrm{dem},i} = E_{\{\boldsymbol{V}_{\setminus i}\}}\left[I\left(Y; X_i|F(\boldsymbol{V}_{\setminus i})\right)\right] \quad (5)$$

$$= \sum_{\forall \boldsymbol{v}_{\setminus i} \in \dot{\mathbb{F}}_2^{m-1}} I\left(Y; X_i|F(\boldsymbol{v}_{\setminus i})\right) p_{\boldsymbol{V}_{\setminus i}}(\boldsymbol{v}_{\setminus i}) \quad (6)$$

where $I(Y; X_i|F(\boldsymbol{V}_{\setminus i}))$ is given in (7) and (8), shown at the bottom of the page. Here $\mathbb{K}$ is a set including all possible

[3] Experimental evidence shows that area properties holding for the BEC also apply to the Bi-AWGN channel [15].

$$I\left(Y; X_i|F(\boldsymbol{v}_{\setminus i})\right) = E_{\{X_i, F(\boldsymbol{v}_{\setminus i}), Y\}}\left[\log_2 \frac{P\left(Y, X_i|F(\boldsymbol{v}_{\setminus i})\right)}{P\left(Y|F(\boldsymbol{v}_{\setminus i})\right) P\left(X_i|F(\boldsymbol{v}_{\setminus i})\right)}\right]$$

$$= \frac{1}{2^{\lambda+1}} \sum_{\forall x_i \in \mathbb{F}_2} \sum_{F(\boldsymbol{v}_{\setminus i}) \in \mathbb{K}} \int_{\mathbb{C}} P\left(Y|F(\boldsymbol{v}_{\setminus i}), X_i = x_i\right) \log_2 \frac{P\left(Y|F(\boldsymbol{v}_{\setminus i}), X_i = x_i\right)}{\frac{1}{2}\sum\limits_{x_i \in \mathbb{F}_2} P\left(Y|F(\boldsymbol{v}_{\setminus i}), X_i = x_i\right)} \mathrm{d}Y \quad (7)$$

$$P\left(Y|F(\boldsymbol{v}_{\setminus i}), X_i\right) = \frac{1}{2^{m-1-\lambda}} \sum_{s \in \zeta^i_{X_i, F(\boldsymbol{v}_{\setminus i})}} P(Y|S = s) \quad (8)$$

realizations of code bits $F(\boldsymbol{v}_{\setminus i})$, and $\lambda = \log_2 |\mathbb{K}|$; $\zeta^i_{x_i,F(\boldsymbol{v}_{\setminus i})}$ denotes the subset of symbols $s \in \zeta$ whose bit labels have the value $x_i$ and $F(\boldsymbol{v}_{\setminus i})$ in the corresponding positions.

*Definition 1:* For an $m$-level multilevel code, the MV-BL-EXIT function $f^{\mathrm{e}}_{\mathrm{MBL},i}$ of the $i$th level at a fixed SNR is defined as:

$$f^{\mathrm{e}}_{\mathrm{MBL},i} : \boldsymbol{I}^{\mathrm{a}}_{\mathrm{dem},\setminus i} \rightarrow I^{\mathrm{e}}_{\mathrm{dem},i}$$

and mathematically can be expressed as:

$$f^{\mathrm{e}}_{\mathrm{MBL},i}\left(\boldsymbol{I}^{\mathrm{a}}_{\mathrm{dem},\setminus i}\right) = \sum_{\forall \boldsymbol{v}_{\setminus i} \in \dot{\mathbb{F}}_2^{m-1}} I\left(Y; X_i | F(\boldsymbol{v}_{\setminus i})\right) p_{\boldsymbol{V}_{\setminus i}}(\boldsymbol{v}_{\setminus i}) \quad (9)$$

based on the *a priori* BEC.

Equation (9) is a multinomial with $m-1$ variables $\boldsymbol{I}^{\mathrm{a}}_{\mathrm{dem},\setminus i}$. When all elements of $\boldsymbol{I}^{\mathrm{a}}_{\mathrm{dem},\setminus i}$ are replaced by a single *a priori* information variable $I^{\mathrm{apri}}$, thus, $I^{\mathrm{a}}_{\mathrm{dem},j} = I^{\mathrm{apri}}$ for $j = 0, \cdots, m-1, j \neq i$, we get the BL-EXIT function.

*Definition 2:* For an $m$-level multilevel code, the BL-EXIT function $f^{\mathrm{e}}_{\mathrm{BL},i}$ of the $i^{\mathrm{th}}$ level at a fixed SNR is defined as:

$$f^{\mathrm{e}}_{\mathrm{BL},i} : I^{\mathrm{apri}} \rightarrow I^{\mathrm{e}}_{\mathrm{dem},i}$$

and mathematically can be expressed as:

$$f^{\mathrm{e}}_{\mathrm{BL},i}(I^{\mathrm{apri}}) = \sum_{\forall \boldsymbol{v}_{\setminus i} \in \dot{\mathbb{F}}_2^{m-1}} I\left(Y; X_i | F(\boldsymbol{v}_{\setminus i})\right) \times (1 - I^{\mathrm{apri}})^{\sum_{j=0, j\neq i}^{m-1} (1-v_j)} I^{\mathrm{apri}}{}^{\sum_{j=0, j\neq i}^{m-1} v_j} \quad (10)$$

based on the *a priori* BEC.

Equation (10) is a degree $m-1$ monomial with variable $I^{\mathrm{apri}}$. $I(Y; X_i|F(\boldsymbol{v}_i))$ includes the physical channel information and is a fixed value when SNR, constellation, and mapping are known. Note that the convergence behaviour is typically analyzed based on the known SNR, constellation, and mapping. $I(Y; X_i|F(\boldsymbol{v}_i))$ can be evaluated using numerical integration. For example, by using Monte Carlo integration over equation (10), (7), and (8), we plot the BL-EXIT functions for three levels of 8PSK with set partitioning (SP) mapping at SNR $= 7$ dB, which is illustrated in Fig. 2.

As stated above, the convergence analysis for MLCM-ID is not easy since the extrinsic information of a particular level is determined by the $m-1$ *a priori* information signals, thus $I^{\mathrm{e}}_{\mathrm{dem},i} = f^{\mathrm{e}}_{\mathrm{MBL},i}(\boldsymbol{I}^{\mathrm{a}}_{\mathrm{dem},\setminus i})$. We hope to simplify the convergence analysis and hence define a single information variable having equivalent effect as the $m-1$ *a priori* information variables $\boldsymbol{I}^{\mathrm{a}}_{\mathrm{dem},\setminus i}$ on the output extrinsic information $I^{\mathrm{e}}_{\mathrm{dem},i}$. In other words, when all elements in $\boldsymbol{I}^{\mathrm{a}}_{\mathrm{dem},\setminus i}$ are equal to this single information value (which is called the combined information), the same $I^{\mathrm{e}}_{\mathrm{dem},i}$ can be obtained. We give the detailed definition of the combined information as follows:

*Definition 3:* For an $m$-level multilevel code, the combined information $I^{\mathrm{comb}}_i \in [0,1]$ of the $m-1$ *a priori* information $\boldsymbol{I}^{\mathrm{a}}_{\mathrm{dem},\setminus i}$ is defined as:

$$I^{\mathrm{comb}}_i = {f^{\mathrm{e}}_{\mathrm{BL},i}}^{-1}\left(f^{\mathrm{e}}_{\mathrm{MBL},i}\left(\boldsymbol{I}^{\mathrm{a}}_{\mathrm{dem},\setminus i}\right)\right) \quad (11)$$

*Definition 4:* $\forall z_1 \in [0,1]$, $\forall z_2 \in [0,1]$, if $z_1 > z_2$, we have $f(z_1) > f(z_2)$ and $\frac{\mathrm{d}f(z)}{\mathrm{d}z}$ exists for $z \in [0,1]$, function $f(\cdot)$ is called the monotonically increasing continuous (MIC) function.

We have proved, in Theorem 1 of a companion paper [17], that $f^{\mathrm{e}}_{\mathrm{BL},i}$ is an MIC function (by which the existence of ${f^{\mathrm{e}}_{\mathrm{BL},i}}^{-1}$ is implied), and there exists one and only one $\boldsymbol{I}^{\mathrm{comb}}_i \in [0,1]$ given the $(m-1)$ *a priori* information $\boldsymbol{I}^{\mathrm{a}}_{\mathrm{dem},\setminus i}$.

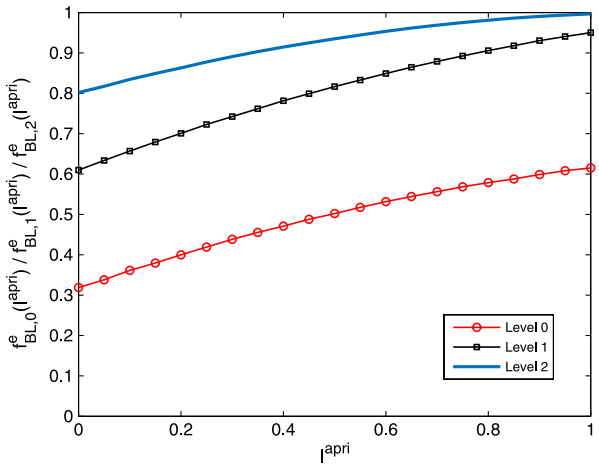



Fig. 2. Monte Carlo Integration Results for BL-EXIT functions based on its definition in (10), and equations (7) and (8). 8PSK; Set Partitioning Mapping; SNR $= 7$ dB.

## IV. Code Design Theory

One specific code design criterion for MLCM-ID has been shown in [9], stating that code of each level should be selected such that its swapped EXIT function closely matches the corresponding BL-EXIT function. Although this criterion gives only one code assignment possibility, it has shown its effectiveness in solving the multilevel code design problem using a two-dimensional EXIT chart. In this paper we explore a general code design rule which allows a wide class of code design and hence fully exploits the inherent flexibility of MLCM-ID in the choice of codes. Of course, we still hope to allocate a code at each level with the aid of the 2D curve fitting technique which has shown its robustness in the design of codes for iterative systems. Hence our aim is:

*Aim: To establish a 2D EXIT function for each level of the multilevel demodulator, with the two dimensions being the a priori and extrinsic information of the decoder at the corresponding level.*

### A. Parametrically-Mapped EXIT Function

Things become easy if the *a priori* information of all levels is associated with one variable. Based on this assumption, we found that a simple mathematical model would be very useful to solve this problem. The model is the parametric equation.

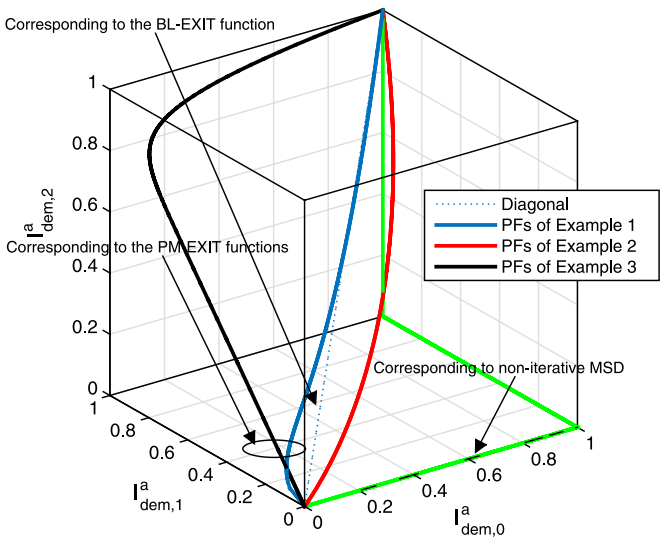



Fig. 3. Examples of 3D trajectories based on parametric equation in (12) for a 3-level multilevel code.

Hence we express the *a priori* information of each level of the multilevel demodulator in the form of:

$$\begin{aligned} I^{a}_{dem,0} &= g_0(\check{t}), \\ &\vdots \\ I^{a}_{dem,m-1} &= g_{m-1}(\check{t}); \qquad \check{t} \in [0,1] \end{aligned} \tag{12}$$

where the parametric functions (PFs) $g_i(\check{t})$ for $i = 0, 1, \cdots, m-1$, must fulfill two conditions:

1) $g_i(\check{t})$ is an MIC function;
2) $g_i(\check{t}) \in [0, 1]$.

Note that the two conditions potentially imply that $g_i(0) = 0$ and $g_i(1) = 1$.

To elaborate a little further, we plot trajectories based on (12) in an $m$-dimensional orthonormal basis $\mathcal{O}^m$, where each dimension represents $I^{a}_{dem,i}$, $i = 0, 1, \cdots, m-1$. Each of trajectories is defined as the locus of a point $\mathcal{Q}$ whose coordinates satisfy the conditions of (12). Fig. 3 shows three trajectories (blue, red and black) based on the PFs of Examples 1, 2 and 3 in Section VI-B, respectively. The three *a priori* information variables form a unit cube and point $\mathcal{Q}(I^{a}_{dem,0}, I^{a}_{dem,1}, I^{a}_{dem,2})$ moves from the origin (0,0,0) and finally converges at (1,1,1) following the trajectory defined by the selected a group of PFs. When $\check{t}$ increases from 0 to 1, the *a priori* information $I^{a}_{dem,i}$ at each level achieves full convergence. Note the green discontinuous trajectory, which gives a geometrical explanation of the principle for conventional MSD. Thus, $I^{a}_{dem,0}$ needs to reach 1 after the first stage of decoding, and given this, $I^{a}_{dem,1}$ needs to reach 1 after the second stage of decoding, and given the first two stages are perfectly decoded, $I^{a}_{dem,2}$ is capable of reaching (1,1,1) which guarantees reliable decoding. Hence $\mathcal{Q}$ moves to (1,1,1) following the green trajectory. Note that "$I^{a}_{dem,0}$ reaches 1" means that bits are perfectly decoded at level 0. If the decoding of any previous stage is not perfect, the green trajectory will not be able to reach (1,1,1) and the decoding performance is degraded therefore. However, if point $\mathcal{Q}$ moves following the continuous trajectories, e.g., the red one, this problem can be easily avoided. Any trajectory defined by PFs of (12) corresponds to one possible way of code choice in MLCM-ID, and the flexibility can be easily seen since there are infinite number of continuous trajectories moving from (0, 0, 0) to (1, 1, 1).

The extrinsic information of the multilevel demodulator at the $i^{\text{th}}$ level depends on $\boldsymbol{I}^{a}_{dem,\setminus i}$—the *a priori* information of all levels except $I^{a}_{dem,i}$. $I^{a}_{dem,i}$ is the extrinsic information of the decoder at the $i^{\text{th}}$ level. Hence, given an $m$-dimensional trajectory defined by (12) in space $\mathcal{O}^m$, its $m-1$ coordinates determine $I^{e}_{dem,i}$, and the remaining coordinate is related to the output of the decoder at the $i^{\text{th}}$ level. Therefore the $m$ coordinates $(I^{a}_{dem,0}, \cdots, I^{a}_{dem,m-1})$ are associated with both the extrinsic information of the demodulator and the decoder, at the $i^{\text{th}}$ level. Based on parametric equation in (12), we have:

$$I^{e}_{dem,i} = f^{e}_{MBL,i}\left(\boldsymbol{g}_{\setminus i}(\check{t})\right) \tag{13}$$

$$g_i(\check{t}) = f_{dec,i}\left(I^{a}_{dec,i}\right) \tag{14}$$

$f_{dec,i}$ is the transfer function of the decoder at the $i^{\text{th}}$ level. A 2D EXIT function has been defined in (13) for the $i^{\text{th}}$ level of a multilevel demodulator. In order to achieve the aim mentioned in the beginning of this section, the argument in (13) should be replaced by the extrinsic information of the decoder of the $i^{\text{th}}$ level. Hence, let $t = g_i(\check{t})$, $t \in [0, 1]$, equations (13) and (14) become:

$$\dot{I}^{e}_{dem,i} = f^{e}_{MBL,i}\underbrace{\left(\boldsymbol{g}_{\setminus i}\left(g_i^{-1}(t)\right)\right)}_{\boldsymbol{I}^{a}_{dem,\setminus i}} \tag{15}$$

$$t = f_{dec,i}\left(I^{a}_{dec,i}\right) \tag{16}$$

*Note that $t$ represents the extrinsic information output from the decoder of the $i^{\text{th}}$ level.* Based on equation (15), we give the definition of the PM-EXIT function:

*Definition 5:* The PM-EXIT function $f^{e}_{PM,i}$ for the $i^{\text{th}}$ level of an $m$-level multilevel demodulator represents the relation between the input $t$ and output $\dot{I}^{e}_{dem,i}$, denoted by:

$$\begin{aligned} f^{e}_{PM,i} &: t \longmapsto \dot{I}^{e}_{dem,i}; \\ f^{e}_{PM,i} &:= f^{e}_{MBL,i}\left(\boldsymbol{g}_{\setminus i}\left(g_i^{-1}(t)\right)\right) \end{aligned}$$

We refer to $\boldsymbol{g}_{\setminus i}(g_i^{-1}(t))$ and $\dot{I}^{e}_{dem,i}$ as the parametrically mapped *a priori* information vector for the demodulator and the Parametrically Mapped Extrinsic Information (PMEI) of the $i^{\text{th}}$ level, respectively. It is clear that when $g_0(\check{t}) = g_1(\check{t}) = \cdots = g_{m-1}(\check{t})$ which corresponds to the diagonal line in $\mathcal{O}^m$, then $f^{e}_{MBL,i}(\boldsymbol{g}_{\setminus i}(g_i^{-1}(t))) = f^{e}_{BL}(t)$. Hence the BL-EXIT function is a special case of the PM-EXIT function. Any trajectory in $\mathcal{O}^m$ defined by parametric equation (12) corresponds to a group of $m$ PM-EXIT functions for $m$ levels. This also potentially implies there are any number of possible PM-EXIT functions because of the wide range of choice of parametric equations resulting in infinite possible trajectories in $\mathcal{O}^m$ based on (12).

### B. 3D EXIT Chart and Convergence Analysis Based on PM-EXIT Function

To obtain the PM-EXIT function of the $i^{\text{th}}$ level, we manually produce the *a priori* L-value $\in \mathbb{R}$ for each the $j^{\text{th}}$ level based on a particular *a priori* binary channel model (e.g. BiAWGN), with the channel parameter chosen to provide the required parametrically mapped *a priori* information value $g_j(g_i^{-1}(t))$ for $j = 0, 1, \cdots, m-1, j \neq i$. The mutual information of the output L-value at the $i^{\text{th}}$ level is the PMEI. Similarly we can also obtain the EXIT function for the decoder, but without the parametrically mapped operation. In the following we show how to analyze the convergence behaviour of MLCM-ID based on the PM-EXIT function and the decoder EXIT function.

In a BICM-ID [18], [19] system, the convergence behaviour can be observed through a two-dimensional staircase trajectory which in fact reflects information exchange between the demodulator and decoder. We also expect to analyse the convergence characteristic of MLCM-ID by observing the information exchange between the PM-EXIT function and decoder EXIT function at each level. At real iterative detection environment, the decoder updates its extrinsic information at each level, such that PMEI of the $i^{\text{th}}$ level is determined by $(m-1)$ variables $\boldsymbol{I}^{\text{e}}_{\text{dec},\backslash i}$, expressed as:

$$\dot{I}^{\text{e}}_{\text{dem},i} = f^{\text{e}}_{\text{MBL},i}\left(g_0\left(g_i^{-1}\left(I^{\text{e}}_{\text{dec},0}\right)\right), \cdots, g_{i-1}\left(g_i^{-1}\left(I^{\text{e}}_{\text{dec},i-1}\right)\right), g_{i+1}\left(g_i^{-1}\left(I^{\text{e}}_{\text{dec},i+1}\right)\right), \cdots, g_{m-1}\left(g_i^{-1}\left(I^{\text{e}}_{\text{dec},m-1}\right)\right)\right) \quad (17)$$

Obviously there are $(m-1)$ variables determining PMEI, and it is not easy to construct an EXIT space. However, returning to the concept of combined information, as aforementioned in Section III and equation (17), we can construct a three-dimensional EXIT space, which is beneficial to the convergence analysis of MLCM-ID. First we define the parametrically-mapped combined information (PMCI) as follows:

*Definition 6:* For an $m$-level multilevel code, the PMCI of the $m-1$ extrinsic information $\boldsymbol{I}^{\text{e}}_{\text{dec},\backslash i}$ is defined in (18), shown at the bottom of the page, and the PMCI function is defined as:

$$f_{\text{PMCI},i} : \boldsymbol{I}^{\text{e}}_{\text{dec},\backslash i} \rightarrow \dot{I}^{\text{comb}}_{i}$$

Note that $f^{\text{e}}_{\text{PM},i}$ is also an MIC function which is in fact a straightforward extension of $f^{\text{e}}_{\text{BL},i}$, hence equation (18) is well defined. Now, we are able to construct a 3D EXIT chart for each level, with three dimensions $I^{\text{e}}_{\text{dec},i}$ ($x$-axis), $\dot{I}^{\text{comb}}_{i}$ ($y$-axis) and $\dot{I}^{\text{e}}_{\text{dem},i}$ ($z$-axis), respectively. Fig. 4 shows an example of the 3D EXIT chart for level 0 of a 3-level MLCM-ID with the modified set partitioning (MSP) [19] mapping. 3D EXIT charts for other levels have a similar form to level 0. As observed in Fig. 4, the PM-EXIT function and the decoder EXIT function are plotted in 3 dimensions and hence form the EXIT surfaces. The PM-EXIT surface of the $i^{\text{th}}$ level, shown as a grid of dotted lines, is determined by the input $t = \dot{I}^{\text{comb}}_{i}$ using (15) which gives $\dot{I}^{\text{e}}_{\text{dem},i}$, and hence is independent of $I^{\text{e}}_{\text{dec},i}$ ($x$-axis). The decoder EXIT surface, shown as a grid of solid lines, is determined by $\dot{I}^{\text{e}}_{\text{dem},i}$ which gives $I^{\text{e}}_{\text{dec},i}$ and is independent of $\dot{I}^{\text{comb}}_{i}$ ($y$-axis). The PM-EXIT function of the $i^{\text{th}}$ level is the projection of the intersection curve between the diagonal plane (vertical plane marked with a grid of dash-dotted lines) and the PM-EXIT surface onto the $(x, z)$ plane.

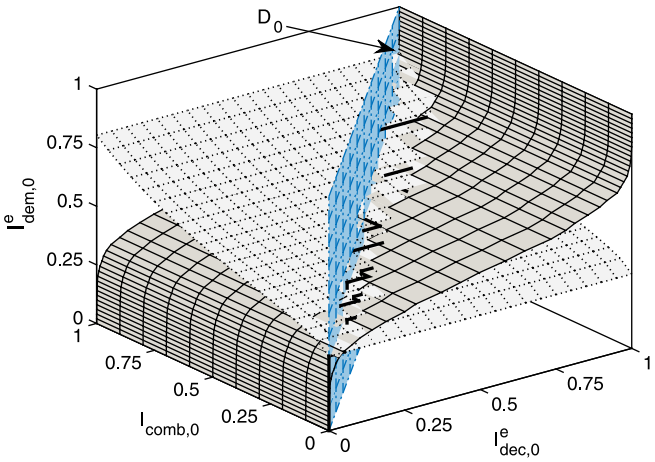



Fig. 4. 3D-EXIT chart for level 0 of MLCM-ID at $E_{\text{s}}/N_0 = 7$ dB; MSP; 8PSK; AWGN channel; The PM-EXIT surfaces are based on an arbitrary selection of parametric functions: $f_0(t) = t^{\frac{3}{5}}, f_1(t) = t^{\frac{19}{50}}$ and $f_2(t) = t^2$.

### C. Code Design Theorem

*Theorem 1:* In a $2^m$-ary digital modulation scheme, full convergence can be achieved by multilevel encoding with code rates $\mathcal{R}^i$, $i = 0, \cdots, (m-1)$, and iteration-aided multistage decoding or iteration-aided parallel independent decoding, if the condition below is fulfilled:

At each level, a code is allocated to ensure there is an open tunnel between its EXIT function and the corresponding PM-EXIT function, where each level's PM-EXIT function $f^{\text{e}}_{\text{PM},i}$, $i = 0, 1, \cdots, m-1$, is generated on the basis of a group of selected parametric equations in (12).

*Proof:* Let $f_{\text{INT},i}(\cdot)$ denote a function whose curve is the projection of the intersection curve between the PM-EXIT surface and the decoder surface of the $i^{\text{th}}$ level onto the $(x, y)$ plane of the 3D EXIT chart. Then, we have:

$$I^{\text{e},(l)}_{\text{dec},i} = f_{\text{INT},i}\left(\dot{I}^{\text{comb}}_{i,(l-1)}\right) \quad (19)$$

$$\dot{I}^{\text{comb}}_{i,(l)} = f_{\text{PMCI},i}\left(\mathbf{I}^{\text{e},(l)}_{\text{dec},\backslash i}\right) \quad (20)$$

where $(l)$ denotes the $l^{\text{th}}$ iteration, e.g, $\dot{I}^{\text{comb}}_{i,(l)}$ denotes the obtained PMCI of the $i^{\text{th}}$ level at the $l^{\text{th}}$ iteration. In initial stage,

$$\dot{I}^{\text{comb}}_{i} = {f^{\text{e}}_{\text{PM},i}}^{-1}\left(f^{\text{e}}_{\text{MBL},i}\left(g_0\left(g_i^{-1}\left(I^{\text{e}}_{\text{dec},0}\right)\right), \cdots, g_{i-1}\left(g_i^{-1}\left(I^{\text{e}}_{\text{dec},i-1}\right)\right), g_{i+1}\left(g_i^{-1}\left(I^{\text{e}}_{\text{dec},i+1}\right)\right), \cdots, g_{m-1}\left(g_i^{-1}\left(I^{\text{e}}_{\text{dec},m-1}\right)\right)\right)\right) \quad (18)$$

$\forall i$, $\dot{I}^{\text{comb}}_{i,(0)} = 0$, $I^{\text{e},(0)}_{\text{dec},i} = 0$. We first analyze parallel independent decoding. Given $I^{\text{comb}}_{i,(0)}$, equation (19) produces $I^{\text{e},(1)}_{\text{dec},i}$ which must be larger than $I^{\text{e},(0)}_{\text{dec},i} = 0$, provided this condition is fulfilled. Note that $I^{\text{e},(1)}_{\text{dec},i}$ increases simultaneously for all levels. Due to the increase of $I^{\text{e},(1)}_{\text{dec},i}$ at all levels, we obtain $I^{\text{comb}}_{i,(1)}$ based on equation (20). Since $\forall i$, $I^{\text{e},(1)}_{\text{dec},i} > I^{\text{e},(0)}_{\text{dec},i}$ and $f_{\text{PMCI},i}$ is the MIC function of $\boldsymbol{I}^{\text{e}}_{\text{dec},\setminus i}$, we have $\dot{I}^{\text{comb}}_{i,(1)} > \dot{I}^{\text{comb}}_{i,(0)}$.

At the second iteration where $l = 2$, we can obtain $I^{\text{e},(2)}_{\text{dec},i}$ from equation (19) because we know $\dot{I}^{\text{comb}}_{i,(1)}$. Provided this condition is fulfilled, $f_{\text{INT},i}(\cdot)$ is an MIC function, and since $\dot{I}^{\text{comb}}_{i,(1)} > \dot{I}^{\text{comb}}_{i,(0)}$, we have $I^{\text{e},(2)}_{\text{dec},i} > I^{\text{e},(1)}_{\text{dec},i}$ for all levels. This further results in $\dot{I}^{\text{comb}}_{i,(2)} > \dot{I}^{\text{comb}}_{i,(1)}$ based on equation (20). It is clear that $I^{\text{e},(l)}_{\text{dec},i}$ and $\dot{I}^{\text{comb}}_{i,(l)}$ are iteratively exchanged among $m$ pairs of equations (19) and (20), (one pair for each level) and provided this condition is fulfilled they will both increase in each iteration until $I^{\text{e},(l)}_{\text{dec},i} = 1$ and $\dot{I}^{\text{comb}}_{i,(l)} = 1$, where the target point $(1, 1, I(Y; X_i|\boldsymbol{X}_{\setminus i}))$ (denoted as $\text{D}_i$) is reached (see Fig. 4), and full convergence is achieved at each level. Codes ensuring full convergence for IA-PID must guarantee full convergence for IA-MSD, since the higher levels receive soft information corresponding to the decision of previous stage decoder at each iteration in IA-MSD, where $\dot{I}^{\text{comb}}_{i,(l)}|_{\text{IA-MSD}} \geq \dot{I}^{\text{comb}}_{i,(l)}|_{\text{IA-PID}}$, $I^{\text{e},(l)}_{\text{dec},i}|_{\text{IA-MSD}} \geq I^{\text{e},(l)}_{\text{dec},i}|_{\text{IA-PID}}$.

As shown in Section IV-A, there is a trajectory in space $\mathcal{O}^m$ which is associated with functions $f^{\text{e}}_{\text{PM},i}$ and $f_{\text{dec},i}$ of each level, since variable $t$ in (15) and (16) identifies $m$ coordinates of this trajectory. Once the point $\text{D}_i$ is reached for $i = 0, 1, \cdots, m-1$, the code assigned at each level ensures full convergence, hence $t = I^{\text{e}}_{\text{dec},i} = 1$ in (15) and (16), which further results in $g_0(1) = \cdots, g_{m-1}(1) = 1$, where point $\mathcal{Q}$ reaches $(1, \cdots, 1)$ based on this trajectory. In terms of the convergence characteristic of the trajectory defined in (12), when codes assigned on the basis of this trajectory achieve full convergence at each level, they also ensure full convergence between the respective EXIT surfaces and the BL-EXIT surfaces constructed by the diagonal trajectory in $\mathcal{O}^m$, since point $\mathcal{Q}$ is capable of reaching $(1, \cdots, 1)$ following diagonal trajectory with the shortest path. This means that code assigned at each level guarantees full convergence of MLCM-ID. □

Provided the condition of Theorem 1 is fulfilled, code rates are limited to the theoretical rate region [1] and satisfy Shannon's noisy-channel coding theorem which are regarded as the necessary conditions to ensure full convergence. This is proved in the next section.

## V. Area Theory

The EXIT chart is widely known as a very efficient and powerful tool in the design of iterative system, e.g., BICM-ID, Turbo codes and concatenated system, because it reduces the design problem into simple curve fitting problem. In this section, we prove three area theorems for the proposed PM-EXIT functions. This gives theoretical support for both Theorem 1 and the code design rule presented at the end of this section.

### A. Area Theorem of BL-EXIT Function

Let $\mathcal{A}^{(i)}_{\text{BL}}, i = 0, \cdots, (m-1)$ be the area under the BL-EXIT function of the $i^{\text{th}}$ level. According to *Definition 2*, the area $\mathcal{A}^{(i)}_{\text{BL}}$ is the result from the integration of $f^{\text{e}}_{\text{BL},i}$ in the domain [0, 1]:

$$\mathcal{A}^{(i)}_{\text{BL}} = \int_0^1 f^{\text{e}}_{\text{BL},i}(t)\text{d}t$$

and mathematically is expressed as (21), shown at the bottom of the page.

*Theorem 2A:* In a $2^m$-ary digital modulation scheme, the sum of the areas $\mathcal{A}^{(i)}_{\text{BL}}, i = 0, \cdots, (m-1)$, is equal to the CCC, thus:

$$I(Y; X_0, \cdots, X_{m-1}) = \sum_{i=0}^{m-1} \mathcal{A}^{(i)}_{\text{BL}} \tag{23}$$

*Proof:* See Appendix A. □

### B. Area Theorem of PM-EXIT Function

*Theorem 2B:* In a $2^m$-ary digital modulation scheme, if $\mathcal{A}^{(i)}_{\text{PM}}$ is the area under the PM-EXIT function of the $i^{\text{th}}$ level, the sum of the areas $\mathcal{A}^{(i)}_{\text{PM}}$ for $i = 0, 1, \cdots, (m-1)$ is equal to the CCC, thus:

$$I(Y; X_0, \cdots, X_{m-1}) = \sum_{i=0}^{m-1} \mathcal{A}^{(i)}_{\text{PM}} \tag{24}$$

*Proof:* According to the *Definition 5*, $\mathcal{A}^{(i)}_{\text{PM}}$ can be expressed as:

$$\begin{aligned}\mathcal{A}^{(i)}_{\text{PM}} &= \int_0^1 f^{\text{e}}_{\text{PM},i}(t)\text{d}t \\ &= \int_0^1 f^{\text{e}}_{\text{MBL},i}\left(\boldsymbol{g}_{\setminus i}\left(g_i^{-1}(t)\right)\right)\text{d}t\end{aligned}$$

$$\mathcal{A}^{i}_{\text{BL}} = \int_0^1 \sum_{\forall \boldsymbol{v}_{\setminus i} \in \dot{\mathbb{F}}_2^{m-1}} I\left(Y; X_i|F(\boldsymbol{v}_{\setminus i})\right)(1-t)^{\sum_{j=0, j\neq i}^{m-1}(1-v_j)} t^{\sum_{j=0, j\neq i}^{m-1} v_j}\,\text{d}t \tag{21}$$

$$\mathcal{A}^{i}_{\text{PM}} = \int_0^1 \sum_{\forall \boldsymbol{v}_{\setminus i} \in \dot{\mathbb{F}}_2^{m-1}} I\left(Y; X_i|F(\boldsymbol{v}_{\setminus i})\right) \prod_{j=0, j\neq i}^{m-1} \left(1 - g_j\left(g_i^{-1}(t)\right)\right)^{(1-v_j)} g_j\left(g_i^{-1}(t)\right)^{v_j}\text{d}t \tag{22}$$

Based on equation (9), $\mathcal{A}_{\mathrm{PM}}^{(i)}$ can be written in the form of (22), shown at the bottom of the previous page. The details of the proof are given in Appendix B. □

Theorem 2B makes an important information theoretic point about PM-EXIT functions. Thus the sum of area under each level PM-EXIT function is always equal to the constellation constrained capacity, and is independent of the selection of the parametric functions. This directly proves that Theorem 1 ensures the overall code rate is less than the capacity $\sum_{i=0}^{m-1}\mathcal{R}^{(i)} \le I(Y; X_0, X_1, \cdots, X_{m-1})$, and hence preserves Shannon's noisy-channel coding theorem.

*Theorem 2C:* The area $\mathcal{A}_{\mathrm{PM}}^{(i)}$ satisfies the following condition: $\sum_{i\in\mathbb{J}}\mathcal{A}_{\mathrm{PM}}^{(i)} \le I(Y; \{X_i|i\in\mathbb{J}\}|\{X_j|j\in\overline{\mathbb{J}}\})$, for all possible sets $\mathbb{J}\subset\{0,\cdots,m-1\}$ of indices, where $\overline{\mathbb{J}}$ is the complementary set of $\mathbb{J}$.

*Proof:* For simplicity, let us consider a three-level multilevel code. As Theorem 2B stated, the maximum sum-area $\sum_{i=0}^{m-1}\mathcal{A}_{\mathrm{PM}}^{(i)}$ is equal to CCC,

$$\mathcal{A}_{\mathrm{PM}}^{(0)} + \mathcal{A}_{\mathrm{PM}}^{(1)} + \mathcal{A}_{\mathrm{PM}}^{(2)} = I(Y; X_0X_1X_2)$$

We know $g_j(g_i^{-1}(t=0)) = 0$, $g_j(g_i^{-1}(t=1)) = 1$, and $f_{\mathrm{PM},i}^{\mathrm{e}}$ is an MIC function, area $\mathcal{A}_{\mathrm{PM}}^{(i)}$ is bounded as

$$f_{\mathrm{PM},i}^{\mathrm{e}}(0)\cdot 1 \le \mathcal{A}_{\mathrm{PM}}^{(i)} \le f_{\mathrm{PM},i}^{\mathrm{e}}(1)\cdot 1$$

Here, $f_{\mathrm{PM},i}^{\mathrm{e}}(0) = I(Y; X_i)$ and $f_{\mathrm{PM},i}^{\mathrm{e}}(1) = I(Y; X_i|\boldsymbol{X}_{\setminus i})$, and hence:

$$I(Y; X_i) \le \mathcal{A}_{\mathrm{PM}}^{(i)} \le I(Y; X_i|\boldsymbol{X}_{\setminus i}) \tag{25}$$

The maximum rate at the $i^{\mathrm{th}}$ level is given by $I(Y; X_i|\boldsymbol{X}_{\setminus i})$. For $m=3$, this means that

$$\mathcal{A}_{\mathrm{PM}}^{(0)} \le I(Y; X_0|X_1X_2)$$
$$\mathcal{A}_{\mathrm{PM}}^{(1)} \le I(Y; X_1|X_0X_2)$$
$$\mathcal{A}_{\mathrm{PM}}^{(2)} \le I(Y; X_2|X_0X_1)$$

Applying (25), we have

$$I(Y; X_0X_1X_2) - \mathcal{A}_{\mathrm{PM}}^{(i)} \le I(Y; X_0X_1X_2) - I(Y; X_i) \tag{26}$$

$$\sum_{j=0, j\neq i}^{m-1}\mathcal{A}_{\mathrm{PM}}^{(j)} \le I(Y; \boldsymbol{X}_{\setminus i}|X_i) \tag{27}$$

where the left and right sides of (27) follow from the Theorem 2B and chain rule of mutual information, respectively. Using the same arguments, the results for a 3-level multilevel code can be readily extended to arbitrary $m$-level multilevel codes. This proves Theorem 2C. □

Theorem 2C also proves that Theorem 1 ensures that code rates assigned at each level are limited to the rate bounding region. A set of possible rate bounds for 8PSK MSP mapping are shown in Fig. 5. We describe the code design method as follows:

*Code Design Rule (PM-EXIT Function Rule):* For a $2^m$-ary digital modulation scheme the code at the $i^{\mathrm{th}}$, $i=0,1,\cdots,(m-1)$ protection level of an MLCM-ID scheme should be chosen such that its EXIT function closely matches[4] the corresponding PM-EXIT function.

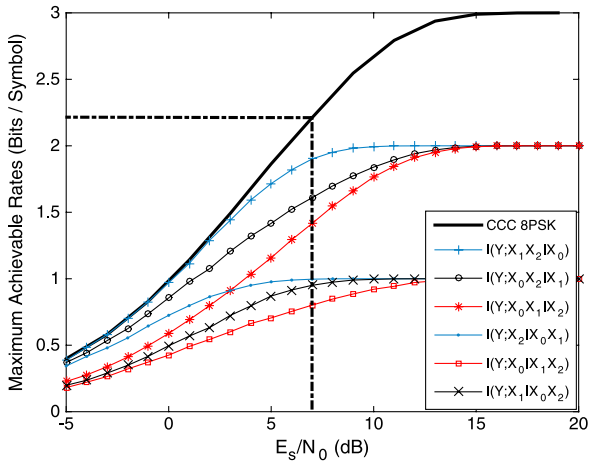



Fig. 5. Maximum achievable rates versus $E_{\mathrm{s}}/N_0$ for 8PSK; MSP; AWGN channel.

The area theorems of the PM-EXIT function give theoretical support for the curve-fitting techniques used in MLCM-ID, and confirm that the PM-EXIT function rule allows: 1. in principle capacity-lossless code design; 2. codes to be limited to the rate region; 3. 2D curve-fitting solution. The code design process thus amounts to the two processes:

1) choosing a group of MIC parametric functions defined in (12), and obtain the PM-EXIT function for each level.
2) selecting codes such that their EXIT functions closely match the corresponding PM-EXIT functions.

Fig. 7 shows a flow chart to describe the procedures of the PM-EXIT function rule.

## VI. Mapping Influence, Examples and Discussions

### A. Influence of the Mapping in MLCM-ID

The index assignment of the mapping affects the BER performance and the flexibility of code selection at each level. We refer to the points $(1, I(Y; X_i|\boldsymbol{X}_{\setminus i}))$ and $(1, I(Y; X_i))$ of the PM-EXIT function as the full iteration gain (FIG) point and no iteration gain (NIG) point, respectively. The best BER performance for each individual level is determined by the corresponding FIG point. A better overall BER performance can be achieved if the FIG point at each level approaches (1,1). The flexibility in the choice of codes depends on the relative ordinate value of the FIG and NIG points. A mapping having a larger value of $|I(Y; X_i|\boldsymbol{X}_{\setminus i}) - I(Y; X_i)|$ at the $i^{\mathrm{th}}$ level, $i=0,1,\cdots,(m-1)$, provides a wider region of variation and this typically means that we are able to assign codes for this level with more flexibility.

[4]"Match" means that the swapped decoder EXIT function should keep an open tunnel with the corresponding PM-EXIT function. This guarantees that the decoder extrinsic information approaches 1 after sufficient number of iterations. "Closely" means that the area of the open tunnel should be as small as possible since it represents the capacity loss. Typically this area is within 5% of the total area of the EXIT chart.

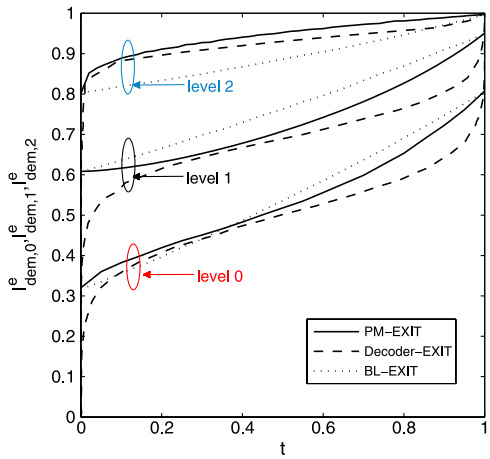



Fig. 6. The BL-EXIT and PM-EXIT functions for Example 1.

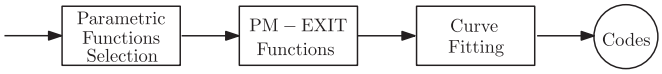



Fig. 7. A flow chart illustrating the procedures of the PM-EXIT function rule.

### B. Examples and Discussion

We first exemplify the practical design procedure by using some appropriate parametric functions and the overall BER performance is then evaluated. The throughput of MLCM-ID may be expressed as $\eta = \sum_{i=0}^{m-1} \mathcal{R}^{(i)}$ bits per channel use. $E_b/N_0|_{dB} = E_s/N_0|_{dB} - 10\log_{10}\eta$, where $N_0$ is the noise power spectral density; $E_b$ and $E_s$ denote the transmit energy per information bit and per channel use, respectively. We take convolutional codes as examples to verify our code design rule. The convolutional decoders in all simulations are based on the *maximum a posteriori* criterion and the BCJR algorithm [20]. The first three examples are based on 8PSK with MSP mapping, but we also consider 16QAM with SP mapping in Example 4. Trajectories of the parametric functions of Examples 1–3 were shown in Fig. 3.

*Example 1:* Assume only power functions are employed. We arbitrarily choose powers for three functions, such as $g_0(\check{t}) = \check{t}^{\frac{3}{5}}$, $g_1(\check{t}) = \check{t}^{\frac{19}{50}}$ and $g_2(\check{t}) = \check{t}^2$, to make a more convincing verification of the PM-EXIT function rule.[5] Fig. 6 shows the corresponding PM-EXIT curves at $E_s/N_0 = 7$ dB. (In the sequel, all the PM-EXIT curves and decoder EXIT curves are obtained by using the *a priori* BiAWGN channel). Dotted lines represent the corresponding BL-EXIT curves. It is obvious that the shapes of the BL-EXIT functions at all levels are distorted, which forms the PM-EXIT functions. It is observed that in this case, the PM-EXIT curve has a convex shape for level 2, but a concave shape for level 1. For level 0, the PM-EXIT curve is at first convex, and then concave.

We evaluate the area $\mathcal{A}^i_{PM}$ via numerical integration based on *Definition 5*, and the result complies with Theorem 2B and 2C. Codes are assigned at each level in terms of the PM-EXIT function rule. The EXIT curves of the decoder are represented as the dashed lines in Fig. 6, with rates $\mathcal{R}^{(2)} = \frac{13}{14}$, $\mathcal{R}^{(1)} = \frac{9}{13}$ and $\mathcal{R}^{(0)} = \frac{1}{2}$ for each level (see Table I for code details). Fig. 8(a) shows the attainable BER performance using IA-PID at overall throughput $\mathcal{R} = 2.12$ bits/symbol. Since each decoder EXIT curve keeps a narrow open tunnel[6] with the corresponding PM-EXIT curves at $E_b/N_0 = 7 - 10\log_{10}(2.12) \approx 3.74$ dB, this ensures full convergence at each level after sufficient iterations. As shown in Fig. 8(a), a sharp "turbo cliff" reaches an overall BER of $4.6 \times 10^{-5}$ at $E_b/N_0 = 3.74$ dB with 20 iterations. This result is consistent with Theorem 1. An error floor arises after $E_b/N_0 = 3.74$ dB, because the convolutional code is employed at each level. In the case where capacity-achieving codes, e.g., sparse graph codes [21] are used, this can easily be avoided. In Example 3 we show the design of MLCM-ID using the turbo code (one family of sparse graph codes). However the removal of the error floor is not the principal objective of this paper, and we refer to, e.g., [22] for the possible solution.

Fig. 8(b) shows the BER performance of IA-MSD, with the same component codes as in Fig. 8(a). The overall BER performance of IA-MSD is the same as IA-PID at $E_b/N_0 = 3.74$ dB with 12 iterations and block length $1.82 \times 10^5$, since the same point $D_i$ for $i = 0, 1, \cdots, m-1$ is reached when full convergence is achieved. It is shown that CCC is closely approached for both IA-PID and IA-MSD, with the overall code rate 2.12 bits/symbol. Hence, based on the PM-EXIT function rule, we have designed codes for MLCM-ID allowing the overall rate only 0.09 bits/symbol (or 0.5 dB) from the capacity, at BER $= 4.6 \times 10^{-5}$. Note that IA-MSD requires less number of iterations to achieve the full convergence than IA-PID, since the higher levels receive soft information corresponding to the decisions of previous stage decoders at each iteration in IA-MSD. When the block length is reduced to 3640, there is still turbo cliff with sufficient number of iterations, but the performance is slightly worse since the error correction capability of the convolutional code is reduced when the block length decreases.

We have shown in the above example that MLCM-ID achieves low BER $= 4.6 \times 10^{-5}$ at a rate which is only 0.09 bits/symbol from CCC. This is attributed to the enhanced flexibility of MLCM-ID in the choice of codes, and the fact that the PM-EXIT function rule fully exploits this property. Hence, it becomes easy to achieve reliable transmission at a rate closer to CCC, when compared to e.g., bit-interleaved coded modulation with iterative decoding (BICM-ID). Analysis and design of BICM-ID basically relies on a simple 2-dimensional EXIT chart. In order to guarantee the full convergence, the swapped decoder EXIT function should match the EXIT function of the demodulator. We consider a BICM-ID system with 8PSK MSP mapping. The demodulator EXIT function over AWGN channel is fixed when we set again $E_s/N_0 = 7$ dB (the same as Example 1). We expect to find a code satisfying 2 conditions: 1. the code rate $\mathcal{R}_{BICM} = 0.71$; this allows the overall throughput approximately the same as Example 1. 2. the swapped decoder EXIT function matches the EXIT function of the demodulator.

[5]Note that we use in this example a group of arbitrarily selected parametric functions to confirm that for *any a group of parametric functions* satisfying the two conditions described in Section IV-A, the PM-EXIT function rule relaxes the 5 constraints mentioned in Section I and ensures full convergence

[6]In all examples, we try to assign codes such that the tunnel is narrow. Hence the simulation results are convincing in support of the area theorems and design rule.

TABLE I
PARAMETERS OF THE COMPONENT CODE USED FOR EACH EXAMPLE [a]

| Example | Level $i$ | Convolutional Codes | | | |
|---|---|---|---|---|---|
| | | Rate | Generator | Puncture Pattern | Inner Loop |
| **1** | 2 | $\frac{13}{14}$ | $[177;117]_8$ | $[10031;03577]_8^{13}$ | N/A |
| | 1 | $\frac{9}{13}$ | $[177;117]_8$ | $[735774;646657]_8^{18}$ | N/A |
| | 0 | $\frac{1}{2}$ | $[37;31]_8$ | N/A | N/A |
| **2** | 2 | $\frac{17}{20}$ | $[217;327]_8$ | $[367245;254721]_8^{17}$ | N/A |
| | 1 | $\frac{39}{50}$ | $[217;327]_8$ | $\begin{bmatrix}4\text{EAEABDD93}\\ 2\text{BABDED6BD}\end{bmatrix}_{16}^{39}$ | N/A |
| | 0 | $\frac{9}{20}$ | $[17,11;15]_8$ | $[735;435;577]_8^9$ | N/A |
| **3** | 2 | 0.76 | $[1;\frac{153}{177}]_8$ | $[1777777;1010001;1000041]_8^{19}$ | 6 |
| | 1 | 0.56 | $[1;\frac{23}{37}]_8$ | $[37777;20443;31241]_8^{14}$ | 6 |
| | 0 | 0.73 | $[1;\frac{47}{65}]_8$ | $\begin{bmatrix}1\text{FFFFFFFFFFFFFFFFF}\\ 1144108404448411401\\ 10008211000201\text{C}8001\end{bmatrix}_{16}^{73}$ | 6 |
| **4** | 3 | $\frac{13}{14}$ | $[177;117]_8$ | $[10031;03577]_8^{13}$ | N/A |
| | 2 | 0.8 | $[177;107]_8$ | $[231;273]_8^8$ | N/A |
| | 1 | 0.67 | $[217;327]_8$ | $\begin{bmatrix}54\text{EFFF6D8F8DBF3EF}\\ 67\text{EEFFAF7FBE6FAB7}\end{bmatrix}_{16}^{67}$ | N/A |
| | 0 (Turbo) | 0.61 | $[1;\frac{21}{65}]_8$ | $\begin{bmatrix}1\text{FFFFFFFFFFFFFFFF}\\ 0\text{E2421009214F491}\\ 0\text{E2421009214F491}\end{bmatrix}_{16}^{61}$ | 6 |

| Example | Level $i$ | Turbo Codes | | | |
|---|---|---|---|---|---|
| | | Rate | Generator | Puncture Pattern | Inner Loop |
| Example | Level $i$ | Turbo Codes/ Convolutional Codes | | | |
| | | Rate | Generator | Puncture Pattern | Inner Loop |

[a] $[\cdot]_8$ is the octal representation of the generator polynomial. $[\cdot]_8^d$ and $[\cdot]_{16}^d$ are the octal and hexadecimal representations of the puncture pattern, with the puncture period $d$. To read generator polynomials, the most significant bit (refer to the most left bit) in binary form corresponds to the first tap of the shift register. To read puncturing pattern, the most significant bit refers to the most left bit in binary form. The puncturing pattern should be read in binary form from left to right starting from the first non-zero binary number. These codes are obtained by exaustively searching a wide range of possible codes, subject to the PM-EXIT function rule.

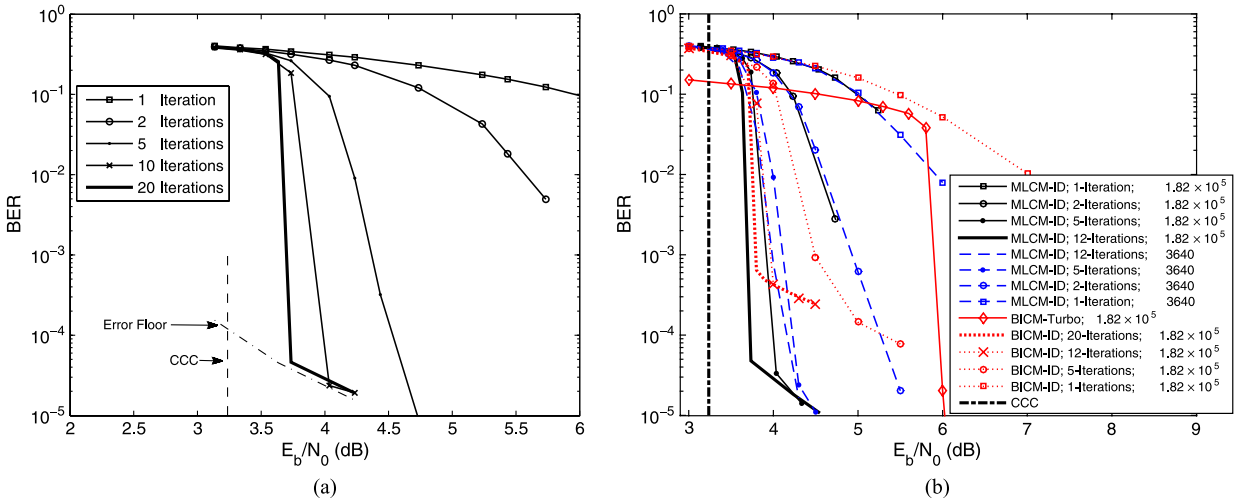
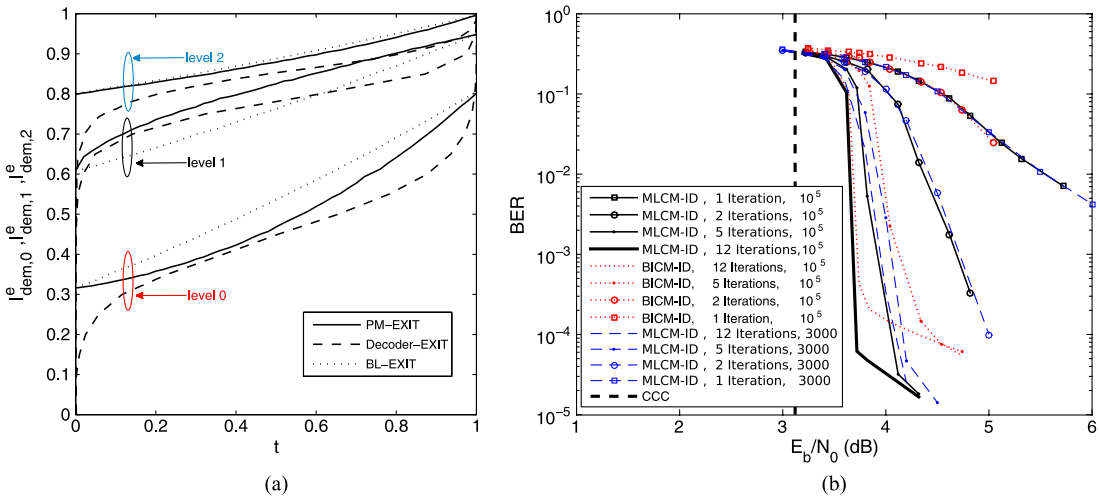



Fig. 8. (a) BER against $E_b/N_0$ of IA-PID, Block length $1.82 \times 10^5$. (b) BER against $E_b/N_0$ of IA-MSD for Example 1. Block length $1.82 \times 10^5$ or 3640; AWGN channel; 8PSK; MSP. Details of the component codes are summarized in Table I. (a) IA-PID. (b)IA-MSD.

After exhaustive search, we cannot find a code which strictly meet these two requirements. This is not surprising. The shapes of the two EXIT functions are quite different but we need them to match as closely as possible since the overall throughput is very close to CCC at $E_s/N_0 = 7$ dB. This is not an easy task, and becomes very difficult when higher throughput is required. Note that for BICM-ID, we cannot distort the demodulator EXIT function as we do the PM-EXIT function in MLCM-ID.




(a) (b)

Fig. 9. (a) The BL-EXIT and PM-EXIT functions for Example 2. (b) BER against $E_b/N_0$ of IA-MSD for Example 2. Block length $10^5$ or 3000; AWGN channel; 8PSK; MSP. Details of the component codes are summarized in Table I. (a) EXIT chart. (b) BER performance.

We pick one code (with memory-6) which so far gives the best match to the EXIT function of the demodulator, where the two EXIT curves intersect at a point slightly lower than the FIG point. Hence we anticipate that there will be a turbo cliff (with sufficient iterations) reaching an overall BER worse than that for MLCM-ID of Example 1. The BER simulations of BICM-ID in Fig. 8(b) are well consistent with our anticipation. The BER reaches $6 \times 10^{-4}$ at around 3.8 dB, and there is 3.5 dB loss at BER $= 10^{-5}$ compared to MLCM-ID. The block length of the simulated BICM-ID is still $1.82 \times 10^5$. Note that the overall number of trellis states for BICM-ID is larger than that for MLCM-ID, but this does not help much. Hence, MLCM-ID showed its ability to approach capacity, with the aid of the PM-EXIT function design rule. The shape of the PM-EXIT functions becomes quite flexible in terms of the parametric functions selected, and this has obtained the actual gain.

We also give comparisons between MLCM-ID and turbo coded BICM. Again, we use 8PSK with MSP mapping. The component convolutional code of the turbo code has memory-5 and rate $\mathcal{R}_{\text{T-BICM}} = 0.71$. The number of internal iteration of the turbo code is 7. There is a loss of approximately 1.5 dB over MLCM-ID at BER $= 10^{-5}$.

*Example 2:* Assuming a high rate code is required for level 1, we may choose a group of parametric functions with the following form, $g_1(\check{t}) = \check{t}^{\tau}$, $\tau > 1$, $g_0(\check{t}) = \check{t}$ and $g_2(\check{t}) = \check{t}$. We arbitrarily select $\tau = 1.8$ and the corresponding PM-EXIT functions are shown in Fig. 9(a). As expected, the PM-EXIT curve of level 1 is convex, which enables a higher rate code to be used at this level. Again, we design code for each level based on the proposed code design rule, with rates $\mathcal{R}^{(0)} = \frac{9}{20}$, $\mathcal{R}^{(1)} = \frac{39}{50}$ and $\mathcal{R}^{(2)} = \frac{17}{20}$ and their EXIT curves well match the corresponding PM-EXIT curves. Fig. 9(b) shows the corresponding BER performance while IA-MSD is employed. The turbo cliff reaches an overall BER $5 \times 10^{-5}$ at $E_b/N_0 = 3.81$ dB, with overall code rates 2.08 bits/symbol which is 0.13 bits/symbol (or 0.7 dB) from the capacity, using 12 iterations.

In this example, we also compare MLCM-ID with BICM-ID. We need to find a rate-0.69 code (which gives the same overall rate as MLCM-ID) for BICM-ID, where its swapped EXIT function matches the EXIT function of the demodulator at $E_s/N_0 = 7$ dB. The computer search returns a memory-6 convolutional code that meets the conditions. Since the overall rate is smaller than Example 1, it is relatively easier to design the code in this example when compared to the code design of BICM-ID in Example 1. This is consistent with our conclusion in Example 1. The BER performance is shown in Fig. 9(b). The overall BER of BICM-ID reaches approximately $3 \times 10^{-4}$ with block length $10^5$ at $E_b/N_0 = 3.9$ dB, which is still worse than MLCM-ID.

*Example 3:* Here we address the design of MLCM-ID using capacity-achieving codes, in this case turbo-codes. Since the swapped EXIT function of the turbo code shows a near-horizontal characteristic, we can easily assign a code rate close to the maximum achievable rate $I(Y, X_i|X_{\setminus i})$ of the $i^{\text{th}}$ level. For example, to assign a code at the 0th level with rate approaching $I(Y, X_0|X_1X_2)$, we choose the parametric functions: $g_0(\check{t}) = \check{t}^{20}$, $g_1(\check{t}) = \check{t}$, and $g_2(\check{t}) = \check{t}$. The corresponding PM-EXIT functions are shown in Fig. 10(a) at $E_s/N_0 = 7$ dB. The PM-EXIT curve of the 0th level has a convex shape which is near-horizontal and passes through the point $(1, I(Y, X_0|X_1X_2))$. The PM-EXIT functions of the other levels must be at concave (which is necessary following Theorem 2B) and hence the allowable rates for these two levels are decreased. Based on the PM-EXIT function design rule, we assign a turbo code with rate $\mathcal{R}^{(0)} = 0.73$ at level 0, which closely approaches the corresponding $I(Y, X_0|X_1X_2) = 0.79$, while the code rates assigned for the other two levels are $\mathcal{R}^{(1)} = 0.56$ and $\mathcal{R}^{(2)} = 0.76$. Note that, a small rate-loss arises since the area below the swapped EXIT function of the turbo code is slightly larger than the actual rate [23]. Since these codes are designed at $E_s/N_0 = 7$ dB and their EXIT functions closely match the corresponding PM-EXIT functions, we expect that there will be a sharp turbo cliff within $E_b/N_0 = 3.88$ dB. Fig. 10(b) clearly confirms our anticipation and again verifies the PM-EXIT function rule. The designed MLCM-ID performs within 0.9 dB of CCC at BER $= 10^{-6}$ after 5 iterations.

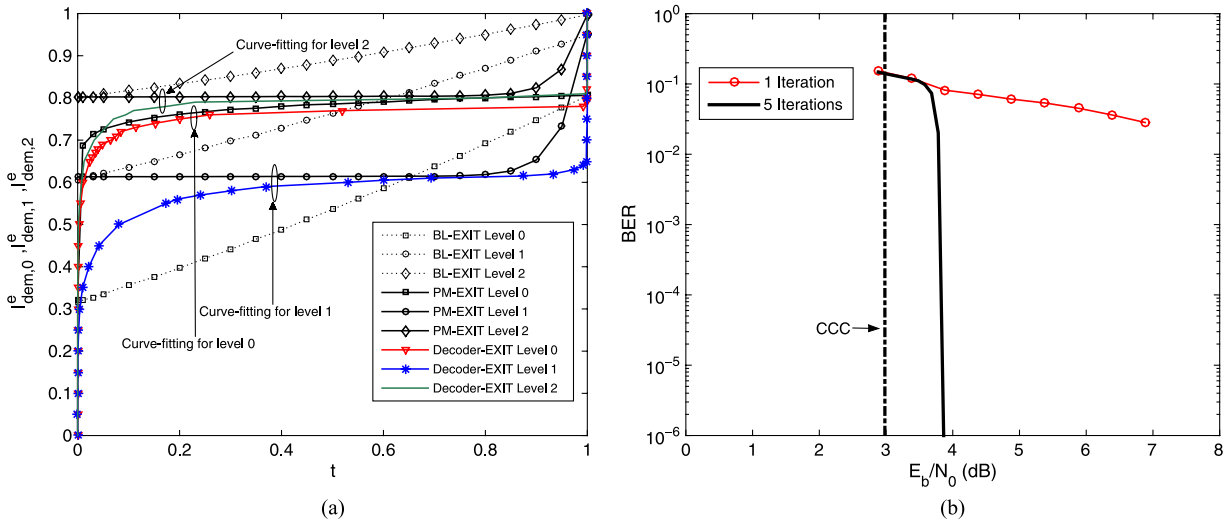



Fig. 10. (a) The BL-EXIT and PM-EXIT functions for Example 3. (b) BER against $E_b/N_0$ of IA-MSD for Example 3. Block length $10^5$; AWGN channel; 8PSK; MSP. Details of the component codes are summarized in Table I. (a) EXIT chart. (b) BER performance.

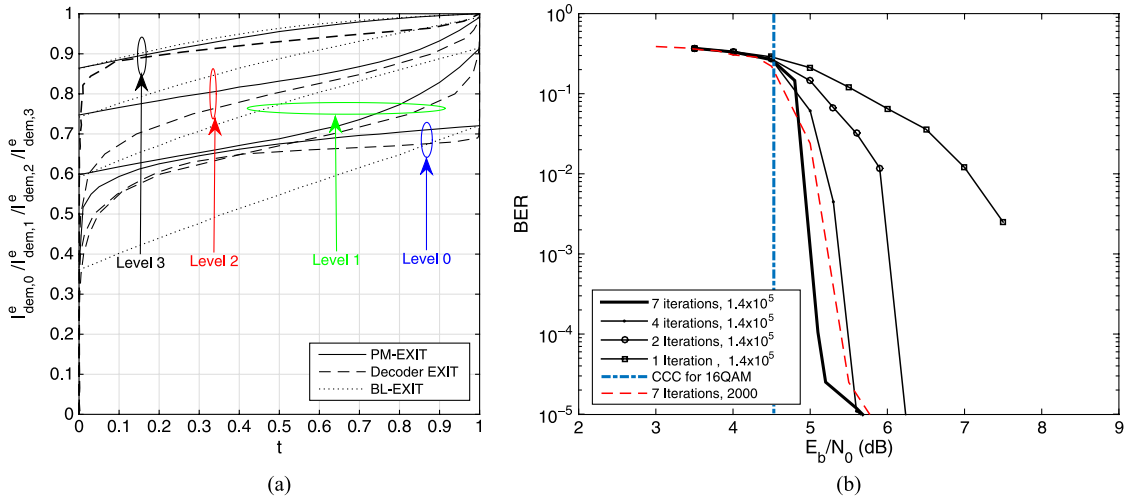



Fig. 11. (a) The BL-EXIT and PM-EXIT functions for Example 4. (b) BER against $E_b/N_0$ of IA-MSD for Example 4. Block length $1.4 \times 10^5$ or 2000; AWGN channel; 8PSK; MSP. Details of the component codes are summarized in Table I. (a) EXIT chart. (b) BER performance.

*Example 4:* In this example we apply the PM-EXIT function rule to a higher order MLCM-ID scheme, i.e., 16QAM (with set partitioning mapping). In order to further verify that the PM-EXIT function rule allows flexible choice of codes at each level, we expect to assign a high rate code for level 0. This typically becomes difficult for non-iterative MSD in terms of the capacity design rule presented in [1]. We choose a group of the parametric functions: $g_0(\check{t}) = \check{t}^5$, $g_1(\check{t}) = \check{t}^{1.1}$, $g_2(\check{t}) = \check{t}$ and $g_3(\check{t}) = \check{t}$, to make the PM-EXIT function of level 0 convex. Fig. 11(a) illustrates the BL-EXIT and PM-EXIT functions at $E_s/N_0 = 10$ dB. Since the PM-EXIT function of level 0 has a slope near horizontal, we employ a turbo code as the component code for this level. Again, we design codes for each level strictly following the PM-EXIT function rule. Thus, the EXIT function of the code should match the corresponding PM-EXIT function as closely as possible. The resultant code rates are $\mathcal{R}^{(0)}/\mathcal{R}^{(1)}/\mathcal{R}^{(2)}/\mathcal{R}^{(3)} = 0.61/0.67/0.80/\frac{13}{14}$, giving an overall rate $\mathcal{R} \approx 3$ bis/symbol (since $\mathcal{R}$ is slightly over 3 bits/symbol). The details of the selected codes are summarised in Table I. In [1, Fig. 11], the authors give simulation results for 16QAM based on the non-iterative MSD. The code rates designed for each level are $\mathcal{R}^{(0)}_{\text{MSD}}/\mathcal{R}^{(1)}_{\text{MSD}}/\mathcal{R}^{(2)}_{\text{MSD}}/\mathcal{R}^{(3)}_{\text{MSD}} = 0.29/0.75/0.96/1$. Note that the code rate for level 0 is limited to $I(Y; X_0) = 0.36$ bits/symbol in terms of the capacity rule, and hence the authors assigned a rate-0.29 turbo code for that level, where the rate is less than half of the code rate $\mathcal{R}^{(0)}$ of MLCM-ID. This clearly exemplifies what we mean by "flexible code design based on the PM-EXIT function rule in MLCM-ID." We should notice that level 3 is left uncoded. This increases the overall throughput whereas the overall performance is dominated by level 3. Hence the decoding performance of the previous stages must be sufficiently good to avoid the error propagation to the uncoded level which normally results in severe performance degradation. In order to avoid error propagation, the authors assume perfect decoding at the lower levels in that example. Thus, the correct transmitted bits at lower levels are directly fed into higher stages (without using the estimates of the previous stage). In example 4, we employ turbo code

only for level 0. The other three levels use convolutional codes. It is observed in Fig. 11(b) that a sharp turbo cliff reaching BER $= 2.5 \times 10^{-5}$ occurs at $E_b/N_0 = E_s/N_0 - 10\log_{10}(3) =$ 5.23 dB. The results are well consistent with the anticipation of Fig. 11(a), which verifies the PM-EXIT function rule for a 4-level multilevel code. Example 4 (one turbo code & three convolutional codes) achieves overall BER performance similar to Fig. 11 of [1] (with 3 turbo codes) at $10^{-5}$, but note that this is based on an unfair comparison since MSD does not use the estimated bits. It is reasonable to consider that the performance of Example 4 will be better than Fig. 11 of [1] if the estimates of the lower levels are used for decoding the higher levels. This is not surprising since MLCM-ID often performs as well as maximum likelihood decoding [5], but the conventional MSD for finite block length is suboptimal. It may be interesting to investigate more details between MSD and MLCM-ID in terms of other aspects, but we will not give more discussions here because it is beyond the scope of this paper. However, based on the aforementioned facts, the PM-EXIT function rule allows MLCM-ID to approach closer to CCC than MSD. This is indeed the actual gain of MLCM-ID over MSD.

Note that MSD needs capacity-achieving codes (e.g., turbo [24] and LDPC [25], [26]) to produce good performance. However here we do not emphasize the performance comparison between the traditional MLCM and MLCM-ID, but refer to [1], [5], [7] for the detailed analysis. In this paper we have given some simulation results which demonstrate the proposed code design rule, using only convolutional and turbo codes as examples. However, this code design rule is in principle feasible for other types of binary channel codes provided that the corresponding decoder is soft-in soft-out.

### C. Selection of the Parametric Functions

There are many possible parametric functions $g_i(\check{t})$, e.g., $g_i(\check{t}) = \alpha_1(\check{t})^{\tau_1} + \cdots + \alpha_n(\check{t})^{\tau_n}$, $\alpha_1 + \cdots + \alpha_n = 1$, $\alpha_1, \cdots, \alpha_n \geq 0$, $\tau_1, \cdots, \tau_n \geq 0$. Varied selection of parametric functions provides diverse code design, which indicates that a wide class of code choice for MLCM-ID is available on the basis of the PM-EXIT function rule. As mentioned above, since a group of parametric functions is capable of changing the shape of the PM-EXIT function at each level, this in fact directly affects the allowable code rate at the corresponding level. The PM-EXIT function rule potentially implies that when the area under the PM-EXIT function increases, we can assign a higher rate code at that level, and vice versa.

Taking power functions e.g., $g_i(\check{t}) = t^{\omega_i}$ into account, empirically we found that code rates are determined by the relative power of each parametric function. Typically the power function having the maximum power results in a convex shape for the PM-EXIT function, and hence allows higher code rate at this level. By contrast, the power function having the minimum power reduces the allowable code rate for this level.

The selection of the parametric function is based on the practical requirement of the code rate at each level. We show the selection approach based on a 3-level multilevel as follows:

1) If we need to increase the code rate for a level, e.g., level 1, powers should be selected such that $\omega_1 > \omega_0, \omega_1 > \omega_2$.
2) If we need to assign code for a level, e.g., level 0, with rate approaching the maximum achievable code rate $I(Y; X_0|X_1X_2)$, then $\omega_0 \gg \omega_1, \omega_0 \gg \omega_2$.
3) If higher code rates are required at two levels, e.g., level 1 and 2, then the parametric functions should be selected such that $\omega_2 > \omega_0, \omega_1 > \omega_0$.
4) If we need to increase rate of level 0, but decrease rate of level 1, then $\omega_0 > \omega_1, \omega_0 > \omega_2, \omega_1 < \omega_2$.

In example 2, we need a high rate code at level 1, which means that the parametric functions should be selected in terms of criterion 1. In example 3, we want to assign a code at level 0, with rate approaching the maximum achievable rate at that level. Hence, criterion 2 is used for guidance.

Using the same principle, we may choose parametric functions for any $m$-level multilevel code, $m > 3$, based on the practical code rate requirements. There are no parametric functions which can simultaneously increase or reduce the allowable code rate for all levels since the overall rate is equal to CCC (see Theorem 2B). This also shows the optimality of the PM-EXIT function design rule from the aspect of capacity, where any of a group of parametric functions gives the optimal code design in terms of capacity. Note that BER of the multilevel demodulator is independent of the parametric function selection, because the end point of the PM-EXIT function of each level is always the FIG point of that level, regardless of which parametric functions are selected.

Empirically, we also found the number of required iterations achieving full convergence is dominated by the overall rate selected. Typically more iterations are needed to achieve reliable transmission at higher code rate.

It is interesting to examine the gap between CCC and the overall code rate selected in terms of the PM-EXIT function rule. Following the area theorems developed in Section V, CCC is equal to $\sum_{i=0}^{m-1} \mathcal{A}_{\mathrm{PM}}^{(i)}$, where $\mathcal{A}_{\mathrm{PM}}^{(i)}$ denotes the area under the PM-EXIT function of the $i^{\mathrm{th}}$ level. This implies that the sum area of the open tunnels between each levels' PM-EXIT function and the swapped decoder EXIT function represents the overall capacity loss. Hence codes designed in terms of the PM-EXIT function rule are capable of approaching CCC. It would be interesting to do further research on the PM-EXIT function, e.g.,

- To design a PM-EXIT function which has a shape similar to the corresponding swapped EXIT function of the component code. This will lead to better matching between the two curves at each level, and it becomes easier to choose codes approaching capacity and reduce the capacity loss. This also potentially indicates another merit of the PM-EXIT function rule.
- To investigate code design for other types of parametric functions.

These are interesting topics for the future research and may lead to new techniques for multilevel codes. Essentially the concept of the PM-EXIT functions, and the idea of parametric representation may motivate more research in other communications areas.

## VII. Conclusion

We have first shown the analytical derivation for the EXIT characteristic of multilevel codes. Then, taking Theorem 1, and three area theorems into account, we have shown that the PM-EXIT function rule is a simple code design method that in principle allows capacity-achieving transmission, full exploitation of the inherent flexibilities in the choice of codes, and full convergence. Area properties of the PM-EXIT function were derived, giving the theoretical support for the PM-EXIT function rule. We have shown by these means that the PM-EXIT function rule proposed for MLCM-ID is easy to use. Simulation results verify our theorems and design rule, and show that MLCM-ID is capable of closely approaching CCC, with the aid of the PM-EXIT function rule.

## Appendix A
## The Proof of Theorem 2A

Following equation (21), the area $\mathcal{A}_{\mathrm{BL}}^{(i)}$ can be derived in (28)–(30), shown at the bottom of the page. Theintegral of (28) is a special case of the Beta function, or can be evaluated by Euler's integral of the first kind. For simplicity, we prove equation (23) by considering a 3-level multilevel code, $m = 3$. Then $\mathcal{A}_{\mathrm{BL}}^{(0)}$ can be expressed by (31), shown at the bottom of the page, and similarly, $\mathcal{A}_{\mathrm{BL}}^{(1)}$ and $\mathcal{A}_{\mathrm{BL}}^{(2)}$ can be obtained in a similar way. According to the chain rule of mutual information, $I(Y; X_0X_1X_2)$ have $m!$ expansions [see (32), shown at the bottom of the page].

Summing these expansions and dividing the results by $m!$, we obtain (33), shown at the bottom of the page. Using (31)

$$\mathcal{A}_{\mathrm{BL}}^{(i)} = \int_0^1 \sum_{\forall \boldsymbol{v}_{\setminus i} \in \dot{\mathbb{F}}_2^{m-1}} I\left(Y; X_i | F(\boldsymbol{v}_{\setminus i})\right) (1-t)^{\sum_{j=0, j\neq i}^{m-1} (1-v_j)} t^{\sum_{j=0, j\neq i}^{m-1} v_j} \mathrm{d}t \tag{28}$$

$$= \sum_{\forall \boldsymbol{v}_{\setminus i} \in \dot{\mathbb{F}}_2^{m-1}} I\left(Y; X_i | F(\boldsymbol{v}_{\setminus i})\right) \frac{\left(\sum_{j=0, j\neq i}^{m-1} (1-v_j)\right)! \left(\sum_{j=0, j\neq i}^{m-1} v_j\right)!}{m!} \tag{29}$$

$$= \sum_{\forall \boldsymbol{v}_{\setminus i} \in \dot{\mathbb{F}}_2^{m-1}} I\left(Y; X_i | F(\boldsymbol{v}_{\setminus i})\right) \frac{\left(m-1-\sum_{j=0, j\neq i}^{m-1} v_j\right)! \left(\sum_{j=0, j\neq i}^{m-1} v_j\right)!}{m(m-1)!}$$

$$= \sum_{\forall \boldsymbol{v}_{\setminus i} \in \dot{\mathbb{F}}_2^{m-1}} I\left(Y; X_i | F(\boldsymbol{v}_{\setminus i})\right) \frac{1}{m\binom{m-1}{\sum_{j=0, j\neq i}^{m-1} v_j}} \tag{30}$$

$$\mathcal{A}_{\mathrm{BL}}^{(0)} = \sum_{\forall (v_1, v_2) \in \dot{\mathbb{F}}_2^{m-1}} I\left(Y; X_0 | F(\boldsymbol{v}_{\setminus 0})\right) \frac{1}{m\binom{m-1}{v_1+v_2}} = \frac{1}{3}\left[ I(Y; X_0) + \frac{1}{2}\sum_{j=1,2} I(Y; X_0|X_j) + I(Y; X_0|X_1X_2) \right] \tag{31}$$

$$\begin{aligned} I(Y; X_0X_1X_2) &= I(Y; X_0) + I(Y; X_1|X_0) + I(Y; X_2|X_0X_1) \\ &= I(Y; X_0) + I(Y; X_2|X_0) + I(Y; X_1|X_0X_2) \\ &\vdots \\ &= I(Y; X_2) + I(Y; X_1|X_2) + I(Y; X_0|X_1X_2) \end{aligned} \tag{32}$$

$$I(Y; X_0X_1X_2) = \frac{1}{3}\left[ \sum_{i=0}^{2} I(Y; X_i) + \frac{1}{2}\left( \sum_{j=1,2} I(Y; X_0|X_j) + \sum_{j=0,2} I(Y; X_1|X_j) + \sum_{j=0,1} I(Y; X_2|X_j) \right) + \sum_{i=0}^{2} I(Y; X_i|\boldsymbol{X}_{\setminus i}) \right] \tag{33}$$

and (33), we obtain

$$I(R; X_0X_1X_2) = \mathcal{A}_{\mathrm{BL}}^{(0)} + \mathcal{A}_{\mathrm{BL}}^{(1)} + \mathcal{A}_{\mathrm{BL}}^{(2)} \tag{34}$$

These results can be readily extended to any integer values of $m$, and Theorem 2A is thus proved.

## Appendix B
## The Proof of Theorem 2B

Following the partial proof in Theorem 2B, equation (22) can be modified to (35), via some mathematical manipulation, shown at the bottom of the page. For simplicity, we assume $m=3$. However, the result can be immediately extended to any modulation order $m$. We define some notations to simplify the tedious mutual information expression: $I(Y; X_0) = a_0$, $I(Y; X_0|X_1) = b_0$, $I(Y; X_0|X_2) = c_0$, $I(Y; X_0|X_1X_2) = d_0$, $I(Y; X_1) = a_1$, $I(Y; X_1|X_0) = b_1$, $I(Y; X_1|X_2) = c_1$, $I(Y; X_1|X_0X_2) = d_1$, $I(Y; X_2) = a_2$, $I(Y; X_2|X_0) = b_2$, $I(Y; X_2|X_1) = c_2$ and $I(Y; X_2|X_0X_1) = d_2$. After some mathematical operations based on equation (35), the sum of the area $\mathcal{A}_{\mathrm{PM}}^{(i)}$, $i = 0,1,2$ can be derived in the form of (36), shown at the bottom of the page, which can also be further manipulated to (37), shown at the bottom of the page.

From the chain rule of mutual information (32), we can get:

$$b_0 - a_0 = b_1 - a_1; \quad c_1 - a_1 = c_2 - a_2; \quad b_2 - a_2 = c_0 - a_0 \tag{38}$$

$$a_0 + d_0 - b_0 - c_0 = a_1 + d_1 - b_1 - c_1 = a_2 + d_2 - b_2 - c_2 \tag{39}$$

Based on (38) and (39), we are able to further derive equation (37) which is shown in (40), shown at the bottom of the page. Nowthe final result in (40) proves Theorem 2B.

In summary, we have proven that for *any group of parametric function* $g_i(t)$, $i = 0, 1, \cdots, m-1$, the sum-area under all PM-EXIT functions is equal to the CCC. It is enough to prove Theorem 2B based on $m = 3$ since the proof for any higher modulation orders $m > 3$ is simply a repetition of the process of the current proof, but it is more tedious. The proof is representative and Theorem 2B is thus proved.

$$\mathcal{A}_{\mathrm{PM}}^{(i)} = \int_0^1 \sum_{\forall \boldsymbol{v}_{\setminus i} \in \dot{\mathbb{F}}_2^{m-1}} I\left(Y; X_i | F(\boldsymbol{v}_{\setminus i})\right) \prod_{j=0, j\neq i}^{m-1} \left(1 - g_j(\check{t})\right)^{(1-v_j)} \left(g_j(\check{t})\right)^{v_j} \mathrm{d}g_i(\check{t}) \tag{35}$$

$$\begin{aligned}
\mathcal{A}_{\mathrm{PM}}^{(0)} + \mathcal{A}_{\mathrm{PM}}^{(1)} + \mathcal{A}_{\mathrm{PM}}^{(2)} =& \int_0^1 a_0 + (b_0 - a_0)g_1(\check{t}) + (c_0 - a_0)g_2(\check{t}) + (a_0 + d_0 - b_0 - c_0)g_1(\check{t})g_2(\check{t})\mathrm{d}g_0(\check{t}) \\
&+ \int_0^1 a_1 + (b_1 - a_1)g_0(\check{t}) + (c_1 - a_1)g_2(\check{t}) + (a_1 + d_1 - b_1 - c_1)g_0(\check{t})g_2(\check{t})\mathrm{d}g_1(\check{t}) \\
&+ \int_0^1 a_2 + (b_2 - a_2)g_0(\check{t}) + (c_2 - a_2)g_1(\check{t}) + (a_2 + d_2 - b_2 - c_2)g_0(\check{t})g_1(\check{t})\mathrm{d}g_2(\check{t}) \qquad (36) \\
\mathcal{A}_{\mathrm{PM}}^{(0)} + \mathcal{A}_{\mathrm{PM}}^{(1)} + \mathcal{A}_{\mathrm{PM}}^{(2)} =& a_0g_0(\check{t})\,|_0^1 + \int_0^1 (b_0 - a_0)g_1(\check{t})\mathrm{d}g_0(\check{t}) + \int_0^1 (c_0 - a_0)g_2(\check{t})\mathrm{d}g_0(\check{t}) \\
&+ \int_0^1 (a_0 + d_0 - b_0 - c_0)g_1(\check{t})g_2(\check{t})\mathrm{d}g_0(\check{t}) \\
&+ a_1g_1(\check{t})\,|_0^1 + \int_0^1 (b_1 - a_1)g_0(\check{t})\mathrm{d}g_1(\check{t}) + \int_0^1 (c_1 - a_1)g_2(\check{t})\mathrm{d}g_1(\check{t}) \\
&+ \int_0^1 (a_1 + d_1 - b_1 - c_1)g_0(\check{t})g_2(\check{t})\mathrm{d}g_1(\check{t}) \\
&+ a_2g_2(\check{t})\,|_0^1 + \int_0^1 (b_2 - a_2)g_0(\check{t})\mathrm{d}g_2(\check{t}) + \int_0^1 (c_2 - a_2)g_1(\check{t})\mathrm{d}g_2(\check{t}) \\
&+ \int _0^1 (a_2 + d_2 - b_2 - c_2)g_0(\check{t})g_1(\check{t})\mathrm{d}g_2(\check{t}) \qquad (37)
\end{aligned}$$

$$\begin{aligned}
&\mathcal{A}_{\mathrm{PM}}^{(0)} + \mathcal{A}_{\mathrm{PM}}^{(1)} + \mathcal{A}_{\mathrm{PM}}^{(2)} \\
&=(a_0 + a_1 + a_2) + (b_0 - a_0)\left(g_1(\check{t})g_0(\check{t})\right)\,|_0^1 + (c_1 - a_1)\left(g_2(\check{t})g_1(\check{t})\right)\,|_0^1 \\
&\quad + (b_2 - a_2)\left(g_0(\check{t})g_2(\check{t})\right)\,|_0^1 + (a_0 + d_0 - b_0 - c_0)\left(g_0(\check{t})g_1(\check{t})g_2(\check{t})\right)\,|_0^1 \\
&=(a_0 + a_1 + a_2) + (b_0 - a_0) + (c_1 - a_1) + (b_2 - a_2) + (a_0 + d_0 - b_0 - c_0) \\
&=c_1 + b_2 + a_0 + d_0 - c_0 = a_2 + c_1 + d_0 = I(Y; X_2) + I(Y; X_1|X_2) + I(Y; X_0|X_1X_2) \\
&=I(Y; X_0X_1X_2)
\end{aligned} \tag{40}$$

## REFERENCES


[1] U. Wachsmann, R. Fischer, and J. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.

[2] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 3, pp. 371–377, May 1977.

[3] T. Takata, S. Ujita, T. Kasami, and S. Lin, "Multistage decoding of multilevel block M-PSK modulation codes and its performance analysis," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1204–1218, Jul. 1993.

[4] A. Burr and T. Lunn, "Block-coded modulation optimized for finite error rate on the white Gaussian noise channel," *IEEE Trans. Inf. Theory*, vol. 43, no. 1, pp. 373–385, Jan. 1997.

[5] M. Isaka and H. Imai, "On the iterative decoding of multilevel codes," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 935–943, May 2001.

[6] R. Tee, S. Ng, and L. Hanzo, "Three-dimensional EXIT chart analysis of iteraltive detection aided coded modulation schemes," in *Proc. IEEE VTC Spring*, Melbourne, Australia, May 2006, vol. 5, pp. 2494–2498.

[7] T. Woerz and J. Hagenauer, "Iterative decoding for multilevel codes using reliability information," in *Proc. IEEE Global Telecommun. Conf.*, Orlando, FL, USA, Dec. 1992, pp. 1779–1784.

[8] S. ten Brink, "Exploiting the chain rule of mutual information for the design of iterative decoding schemes," in *Proc. 39th Annu. Allerton Conf. Commun, Control, Comput.*, Monticello, IL, USA, Oct. 2001, pp. 293–300.

[9] Y. Wang and A. G. Burr, "Code assignment of multilevel coded modulation with iterative decoding," in *Proc. Eur. Wireless Conf.—Sustainable Wireless Technol.*, Vienna, Austria, Apr. 2011, pp. 1–8.

[10] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[11] F. Brännström, L. Rasmussen, and A. Grant, "Convergence analysis and optimal scheduling for multiple concatenated codes," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3354–3364, Sep. 2005.

[12] S. ten Brink, "Convergence of multidimensional iterative decoding schemes," in *Conf. Rec. 35th Asilomar Signals, Syst. Comput.*, Pacific Grove, CA, USA, vol. 1, Nov. 2001, pp. 270–274.

[13] I. Land and J. Huber, "Information combining," in *Foundations and Trends in Communications and Information Theory*, vol. 3. Boston, MA, USA: Now Publisher, 2006.

[14] I. Land, P. Hoeher, and J. Huber, "Analytical derivation of EXIT charts for simple block codes and for LDPC codes using information combining," in *Proc. EUROSIPCO*, Sep. 2004, pp. 1561–1564.

[15] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: Model and erasure channel properties," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2657–2673, Nov. 2004.

[16] T. Cover and J. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1991.

[17] Y. Wang and A. Burr, "Analytical derivation of multi-variate bit-level EXIT functions for multilevel codes," *IEEE Commun. Lett.*, vol. 17, no. 2, pp. 249–252, Feb. 2013.

[18] A. Alvarado, L. Szczecinski, E. Agrell, and A. Svensson, "On BICM-ID with multiple interleavers," *IEEE Commun. Lett.*, vol. 14, no. 9, pp. 785–787, Sep. 2010.

[19] N. Tran and H. Nguyen, "Signal mappings of 8-ary constellations for bit interleaved coded modulation with iterative decoding," *IEEE Trans. Broadcast.*, vol. 52, no. 1, pp. 92–99, Mar. 2006.

[20] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 2, pp. 284–287, Mar. 1974.

[21] D. Mackay, *Information Theory, Inference, and Learning Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[22] K. Fukawa, S. Ormsub, A. Tölli, K. Anwar, and T. Matsumoto, "EXIT-constrained BICM-ID design using extended mapping," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, p. 40, Feb. 2012.

[23] F. Schreckenbach, "Iterative decoding of bit-interleaved coded modulation," Ph.D. dissertation, Institute for Communications Engineering, Munich Univ. Technol., Munich, Germany, 2007.

[24] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes. 1," in *Proc. IEEE ICC*, Geneva, Switzerland, May 1993, vol. 2, pp. 1064–1070.

[25] R. Gallager, "Low-density parity-check codes," *IRE Trans. Inf. Theory*, vol. 8, no. 1, pp. 21–28, Jan. 1962.

[26] D. MacKay and R. Neal, "Near Shannon limit performance of low density parity check codes," *Electron. Lett.*, vol. 33, no. 6, pp. 457–458, Mar. 1997.


**Yi Wang** received the B.Eng. degree (1st class) in communications engineering from Xidian University, University of Electronics Science and Technology of Xian, in 2005, and the M.Sc. (with distinction) and Ph.D. in electronics and communications engineering from University of York, in 2008 and 2012, respectively. He is currently a Research Associate at the Communications and Signal Processing Group, University of York, doing research and software development for European Commission and Ministry of Defence. His research interests are in information theory, lattice codes and network coding, data analysis, statistical signal processing, and coded modulation.

**Alister G. Burr** (M'91) was born in London, U.K, in 1957. He received the B.Sc. degree in electronic engineering from the University of Southampton, U.K., in 1979, and the Ph.D. from the University of Bristol, in 1984. Between 1975 and 1985, he worked at Thorn-EMI Central Research Laboratories in London. In 1985, he joined the Department of Electronics at the University of York, U.K, where he has been Professor of Communications since 2000. His research interests are in wireless communication systems, especially modulation and coding, and cooperative systems including especially physical layer network coding. He has published more than 200 papers in refereed international conferences and journals, and is the author of *Modulation and Coding for Wireless Communications* (Prentice-Hall/PHEI). In 1999, he was awarded a Senior Research Fellowship by the U.K. Royal Society, and in 2002, he received the J. Langham Thompson Premium from the Institution of Electrical Engineers. He has also held a Visiting Professorship at Vienna University of Technology, and given numerous invited presentations, including at the First International Conference on Turbocodes and Related Topics. He is currently Chair, Working Group 2, of the European COST IC1004 Programme Cooperative Radio Communications for Green Smart Environments.