

Received October 6, 2020, accepted October 18, 2020, date of publication November 16, 2020, date of current version December 8, 2020. Digital Object Identifier 10.1109/ACCESS.2020.3038082

LDPC-Coded Modulation Performance Analysis and System Design

CORNELIUS TOMAS HEALY^{®1}, ANAS AL RAWI^{®2}, (Member, IEEE), AND CHARALAMPOS C. TSIMENIDIS^{®1}, (Senior Member, IEEE)

¹School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K.

²Research and Technology, British Telecom (BT), Ipswich IP5 3RE, U.K.

Corresponding author: Cornelius Tomas Healy (cornelius.healy@newcastle.ac.uk)

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) through the Cooperative Backhaul Aided Next-Generation Digital Subscriber Loops under Project EP/N004299/1.

ABSTRACT In this work, the low-density parity-check (LDPC)-coded modulation scheme which has been selected for inclusion in the next generation of International Telecommunication Union (ITU) broadband standard is investigated. The multi-level mapping of this scheme offers excellent performance along with a straightforward modulation which is separable into simple one-dimensional schemes. A subset of modulated bit positions are protected by the LDPC code, while the remaining positions benefit from improved Euclidean distance through assignment to distinct constellation regions. To date, there has been a lack of analytical treatment of the scheme. This is problematic given the number of tunable scheme and code parameters, necessitating time-consuming Monte Carlo simulation. This motivates the novel analytical work of this article, which derives overall error-rate and threshold performances through considering separately the effects of the channel on the coded and uncoded bits. First, new closed-form expressions are derived for the hard-decision performance of the bit mapping of the multi-level modulation scheme, for both coded and uncoded bits. Extrinsic information transfer (EXIT) analysis is then applied to the coded bits in the scheme. The derived theoretical performance of the uncoded bit positions is used in combination with the derived EXIT threshold to provide for the first time for this modulation scheme a method to evaluate analytically whether a designed system will offer capacity-approaching performance. Following this, an approximation to the capacity curve for the coded modulation scheme is produced, again based on the derived analytical performance. Finally, through the relationship between mutual information and error rate performance for error control codes, the derived analytical expressions are used to produce a semi-analytical finite-length performance predictor, whereby error rate results for a given code on the binary input additive white Gaussian noise (AWGN) channel can be mapped to the coded modulation scheme under consideration for any coded system parameters.

INDEX TERMS LDPC codes, coded modulation, EXIT charts.

I. INTRODUCTION

Low-density parity-check (LDPC) codes [1] along with turbo codes have comprised the modern graph-based iteratively decoded class of codes capable of performance close to channel capacity, and have as such seen adoption in a broad number of systems and international standards [2]–[5], both as options and, increasingly, as the default coding scheme for data channels. Important early work on LDPC codes included generalisation of the code family and improved performance

The associate editor coordinating the review of this manuscript and approving it for publication was Zilong Liu¹⁰.

through code design [6], along with the development of classes of codes with simplified encoding [7], [8] and decoding [9], and work on performance analysis [10], [11].

Coded modulation offers improved spectral efficiencies and increased data rate when the channel conditions support the use of higher-order constellations, and for LDPCcoded systems the bit-interleaved coded modulation (BICM) scheme [12] has been widely used. Another coded modulation scheme which has seen use in contemporary and past generations of digital subscriber line (DSL) standards [13], among other applications, is trellis-coded modulation (TCM) [14], whereby successive symbols generated by the trellis of a convolutional code are used to select constellation symbols. For the next DSL standard from the ITU-T [15], an alternative to both TCM and BICM has been selected, known as LDPC-coded modulation (LCM) [16], [17], which has some similarities to TCM and offers the benefit of system rate matching across constellation orders with a fixed LDPC code graph. LCM may be viewed as a multi-level modulation scheme [18] whereby a subset of bits are mapped to the constellation symbol uncoded but afforded some protection through regional mapping in conjunction with the remaining coded bits. This scheme (similarly to BICM) offers differing levels of protection to different bit positions in the bit-tosymbol mapping, known as unequal error protection (UEP) [19] and the effect of this can be modelled through the use of equivalent bit channels [20]-[22]. These bit channels allow straightforward application of methods developed for the analysis of LDPC codes on binary-input continuous-output channels, as demonstrated in the authors prior work [23].

While the spectral efficiency and error rate performance of the LCM scheme have been demonstrated in the literature to be excellent, resulting in its inclusion in the upcoming MGFAST DSL standard, the results presented to date rely on time-consuming Monte Carlo simulation for a selected LDPC code and its parameters, and for each modulation order and set of LCM parameters. The lack of analytical treatment for the scheme introduces significant difficulties in overall system design taking account of the effect of varying the code and scheme parameters. In particular, the effect of varying the proportion of coded and uncoded bits per symbol is not well described by the results in the literature. Motivated by this gap in the literature, in this work the overall performance of the LCM scheme is analysed through separate consideration of the coded and uncoded bits in the bit-to-symbol mapping. It is demonstrated that the decoding threshold of the coded bits can be taken to indicate the threshold of the overall system, provided the system parameters are carefully selected to ensure system error rate is not degraded by the performance of the uncoded bits in the high-SNR region. The uncoded BER of the coded bit positions is investigated and an analytical expression is developed. From this, the equivalent bit channel noise variance can be derived, allowing application of EXIT analysis method for a given LDPC code graph, leading to the coded bit (and overall system) decoding threshold.

Separately, the expression for the error rate of the uncoded bit positions in the multi-level modulation scheme is also developed, and is demonstrated to show the system design conditions for capacity-approaching performance, which in combination with the system decoding threshold will inform parameter selection and overall system design for the LCM scheme. Finally, the expressions for bit-channel error rate are used to produce an approximate measure of mutual information for the LCM mapping, allowing both capacity analysis and semi-analytical finite-length error-rate performance predictions for the scheme. A simulation study demonstrates the accuracy of the analysis and performance predictors developed, showing that they characterise the system well. The rest of this article is laid out as follows: In Section II the preliminaries are provided, including a detailed description of the LCM scheme and its demodulation and decoding steps. In Section III the bit channel and uncoded error rate analysis is performed and the analytical expressions characterising uncoded performance are derived, with an illustrated example provided for a particular case. Section IV describes the application of EXIT analysis to the equivalent bit channels and the method to derive the decoding threshold, while Section V covers the mutual information based analysis. Section VI provides a comprehensive simulation study for the scheme and the contributions in this article, and Section VII offers concluding remarks.

II. PRELIMINARIES

In this section, the LCM scheme is described in detail, along with the class of codes used in producing the results throughout this article. The section concludes with some observations on the error rate performance of the LCM system, through a simulation result on the AWGN channel. These observations motivate the following work.

A. LDPC CODED MODULATION

LCM is a coded modulation scheme for M-QAM constellations where B_c bits per symbol are encoded by a rate $R_{LDPC} = \frac{K}{N}$ LDPC code, such that the number of information bits per symbol is given by:

$$B_i = B_u + B_c R_{LDPC}, \tag{1}$$

where $B_u = \log_2(M) - B_c$ is the number of uncoded bits per symbol, and the coded modulation scheme rate is $R_{LCM} = \frac{B_i}{\log_2 M}$. This allows the use of a fixed LDPC code to provide varying levels of protection to constellation symbols, according to constellation order M and number of encoded bits B_c . For example, with $R = \frac{3}{4}$ code and $B_c = 4$, for 64-QAM we have $R_{LCM} = \frac{5}{6}$ and for 2¹⁴-QAM, $R_{LCM} = \frac{13}{14}$. Thus the level of protection afforded by the scheme with fixed code reduces with increasing modulation order, which may beneficial given that higher-order modulation will be selected only in high-SNR scenarios.

The multilevel coded modulation of the LCM scheme is achieved by employing two separate 1-dimensional symbol mappers, as illustrated in Fig. 1. The effect of this mapping is that, in each dimension, the uncoded and coded input bits select subsets of the amplitude constellation and the modulated symbol is comprised of the intersection of these subsets. This concept is more easily described graphically as in Fig. 2a and Fig. 2b, for uncoded and coded bits, respectively. It can be seen in Fig. 2 that the uncoded bits together define a 2-dimensional region of the constellation while the coded bits together describe a disparate subset of the constellation points similar to the cosets found in traditional TCM schemes [13], [14].

Uncoded Symbol	Bits	Coded Symbol
+15 +13 +11 +9	10	+15 +7 -1 -9
+7 +5 +3 +1	11	+13 +5 -3 -11
-1 -3 -5 -7	01	+11 +3 -5 -13
-9 -11 -13 -15	00	+9 +1 -7 -15

TABLE 1. LCM sc	heme one-to-many mapping	g for 256-QAM an	$d B_c = 4,$
in each dimensio	n.		

LCM MAPPING EXAMPLE

The LCM mapping is more clearly described by example. Consider the 256-QAM LCM system with $B_u = 4$ and $B_c = 4$, and a rate $R_{LDPC} = \frac{3}{4}$ LDPC code. Then the number of information bits is $B_i = 7$, with one bit of redundancy per symbol. The bits to be mapped are partitioned into in-phase and quadrature bits denoted by v and w, and further partitioned to uncoded and coded bits denoted by the subscripts as v_u/v_c and w_u/w_c , respectively. These form the inputs to the double-Gray 1-dimensional mapping illustrated in Fig. 1, where in this example each partitioned set contains two bits. A numerical example:

bits =
$$\begin{bmatrix} 0 \\ w_u \end{bmatrix} \begin{bmatrix} 0 \\ w_c \end{bmatrix} \begin{bmatrix} 0 \\ w_c \end{bmatrix} \begin{bmatrix} 1 \\ v_u \end{bmatrix} \begin{bmatrix} 1 \\ v_c \end{bmatrix} \begin{bmatrix} 1 \\ v_c \end{bmatrix}$$
 (2)

Now the mapping of Table 1 applies for each partitioned set of bits, leading from the left-hand mapping to the regional constellation subsets of Fig. 2a, and for the right-hand mapping of Table 1 of the coded bits to the disparate constellation subsets of Fig 2b. In Fig. 2a, the bits $w_u = [0, 1]$ select the amplitudes with marker 'x', and bits $v_u = [1, 0]$ select the amplitudes with marker '*'. The 2-dimensional region associated with the uncoded bits $[w_u, v_u]$ is then formed by the symbols in the intersection marked ' ∇ ', and the resulting increased Euclidean distance is one source of improved error rate performance in the scheme. Likewise, from the right-hand mapping of Table 1 for the coded bits $w_c = [0, 0]$ leads to the disparate subset indicated by '+' markers in Fig. 2b and $v_c = [1, 1]$ produces the subset indicated by '*'. Now, the 2-dimensional coset associated with $[w_c, v_c]$ is produced by the intersection of these sets as shown with '\$'. Finally, the symbol selected transmission for the bits $[w_u, w_v, v_u, v_c]$ is found as the intersection of the uncoded region subset and the coded disparate subset, the symbol (13-7j) indicated in both Figs. 2a and 2b with the red 'o'.



FIGURE 1. Block diagram for the LDPC Coded Modulation scheme.

B. LLR COMPUTATION, DECODING AND DEMODULATION1) RECOVERY OF CODED BITS

Computation of the log-likelihood ratio (LLR) for the LCM scheme is performed with regards to the coded bit subsets only, based on the one-to-many mapping described visually in Fig. 2b for the $[w_c, v_c]$ coset (\diamond). This relationship fully describes the LCM scheme in terms of the LLR computation, and the bit LLRs are computed as standard according to the sets S_0 and S_1 , of symbols associated with bit value 0 or 1 in position *b* of the vector $[w_c, v_c]$, respectively:

$$L(b) = \log \frac{\sum_{s \in S_0} e^{\frac{-1}{\sigma_n^2} |r-s|^2}}{\sum_{s \in S_1} e^{\frac{-1}{\sigma_n^2} |r-s|^2}},$$
(3)

where *r* is the received symbol and σ_n^2 is the noise variance of additive white Gaussian noise (AWGN). Once the bit LLRs have been computed and mapped to original code word bit positions (through de-interleaving, if necessary), decoding also proceeds normally for the selected iterative message-passing algorithm, such as sum-product algorithm (SPA).

2) RECOVERY OF UNCODED BITS

In contrast to the LDPC-coded BICM schemes where all bits are encoded, in LCM a further demodulation step is required for the uncoded bits, in order to take advantage of the improved received symbol estimate provided by the LDPC code and achieve performance improvements beyond those offered simply through the increased average minimum distance afforded by the regional subsets described in Fig. 2a. Following decoding of the coded bits, the scheme reproduces the decoded coset from the decoded bits (which need no further processing). From this decoded coset, the member with minimum Euclidean distance to the original soft received



FIGURE 2. Example of constellation cosets for the LCM scheme.

symbol selects the uncoded regional subset for hard-decision recovery of the uncoded bits.

This leads to a significant performance improvement for the uncoded bits: for standard hard-decision demodulation based only on the regional subsets, uncoded single-bit error events for symbols on the inner boundary of the regions occur for transpositions greater than d (half the inter-symbol spacing) in the in-phase/quadrature directions of the signal space. When the decoded coset is used to recover the uncoded bits, assuming LDPC decoding is successful, in the worst case single-bit error events will only occur for transpositions greater than or equal to half the spacing between the coded coset symbols (which is 4d in the example case of Fig. 2b).

C. LDPC CODE CLASS

The LDPC code class used for all results presented in this work is a generalisation of the code suggested for inclusion in the next generation DSL standard. It is quasi-cyclic, i.e., formed of tiled circulant permutation matrices, and the base matrix has the familiar dual format to allow linearcomplexity encoding. The graphs in this work were generated by progressive edge growth (PEG) algorithm [24] selection of first entry in the circulant.

III. LCM UNCODED ERROR RATE ANALYSIS

In this section, uncoded error rate analysis is provided for the bit-to-symbol mapping of the LCM scheme, for both the coded and uncoded bit positions. Note particularly that while the bit positions of the coded bits are considered, along with their symbol mapping, the analysis here assumes a hard decision at the receiver with no use of any decoder. This approach allows investigation of the equivalent binaryinput/continuous-output bit channels encountered by the LDPC code bits [21]–[23], which will then be exploited



(b) Coded constellation cosets for $w_c = \{00\}, v_c = \{11\}$

for EXIT chart and MI-based analysis of the coded system performance.

Furthermore, for the uncoded bit positions, the error rate analysis is considered for the practical case where the received symbol hard decision is performed after the received symbol estimate is improved by use of the decoded coset, where the assumption is made that the LDPC decoding was successful and so the decoded coset is correct. Simulation results will demonstrate that the analytical expression for the error rate derived under this assumption is valid for the SNR region above the threshold of the LDPC code, as expected.

Since the QAM symbol mapping is fully separable to two identical 1-dimensional mapping operations, the error rate analysis of coded and uncoded bit positions is straightforward, based simply on Euclidean distance of the error regions for each bit position.

A. EXAMPLE FOR 256-QAM WITH 4 CODED BITS1) CODED BIT POSITIONS

The 1-dimensional hard-decision error regions for the coded bits of the 256-QAM LCM scheme with $B_c = 4$ are presented in Fig. 3, corresponding to the right-hand mapping of Table 1 for the example amplitude symbol (-15), and assuming a symbol spacing of 2*d*, the Euclidean distances for the decision regions highlighted are illustrated, for both coded bit positions. For this symbol, in Fig. 3, the regions highlighted in red correspond to a bit error, while the regions in blue correspond to symbol hard decisions which do not lead to an error in the respective bit positions. The illustrated Euclidean distances to the decision boundaries make up the complementary error function coefficients of eqns. (4) and (5). The error rate under additive white Gaussian noise given the transmitted symbol S = -15 is found by application of the complimentary error function, as in [25], [26]:

$$P_{b}^{c,\{S=-15\}}(0) = \frac{1}{2} \left[\operatorname{erfc}\left(\frac{d}{\sqrt{N_{0}}}\right) - \operatorname{erfc}\left(\frac{5d}{\sqrt{N_{0}}}\right) + \operatorname{erfc}\left(\frac{9d}{\sqrt{N_{0}}}\right) - \operatorname{erfc}\left(\frac{13d}{\sqrt{N_{0}}}\right) + \operatorname{erfc}\left(\frac{17d}{\sqrt{N_{0}}}\right) - \operatorname{erfc}\left(\frac{21d}{\sqrt{N_{0}}}\right) + \operatorname{erfc}\left(\frac{25d}{\sqrt{N_{0}}}\right) - \operatorname{erfc}\left(\frac{29d}{\sqrt{N_{0}}}\right) \right], \quad (4)$$

and

$$P_{b}^{c,\{S=-15\}}(1) = \frac{1}{2} \left[\operatorname{erfc}\left(\frac{3d}{\sqrt{N_{0}}}\right) - \operatorname{erfc}\left(\frac{7d}{\sqrt{N_{0}}}\right) + \operatorname{erfc}\left(\frac{11d}{\sqrt{N_{0}}}\right) - \operatorname{erfc}\left(\frac{15d}{\sqrt{N_{0}}}\right) + \operatorname{erfc}\left(\frac{19d}{\sqrt{N_{0}}}\right) - \operatorname{erfc}\left(\frac{23d}{\sqrt{N_{0}}}\right) + \operatorname{erfc}\left(\frac{27d}{\sqrt{N_{0}}}\right) \right],$$
(5)

where S = -15 is the symbol under consideration from the 1-dimensional constellation. For the case of 256-QAM the symbol half-distance term $d = \sqrt{\frac{4 E_b}{85}}$ is found from the general expression for *M*-QAM constellations:

$$d = \sqrt{\frac{3\log_2 M}{2(M-1)}} \frac{E_b}{N_0},$$
 (6)

where E_b is the transmitted energy per bit and N_0 is the noise power. The expressions (4), (5) provide the probability of error for a single symbol in the two coded bit positions. The average bit error rate in each coded position b, $P_b^c(b)$, is simply the average of $P_b^{c,S}(b)$ over all symbols in the 1-dimensional constellation, $S \in C$, presented in brief notation as:

$$P_{b}^{c}(0) = \frac{1}{\sqrt{M}} \sum_{i=0}^{\sqrt{M}-2} \left[(-1)^{\left(\lfloor \frac{i}{2} \rfloor \right)} \left(\frac{\sqrt{M}}{2} - \lfloor \frac{i-1}{2} \rfloor - 1 \right) \operatorname{erfc} \left(\frac{d (2i+1)}{\sqrt{N_{0}}} \right) \right], \quad (7)$$

$$P_{b}^{c}(1) = \frac{1}{\sqrt{M}} \sum_{i=0}^{\sqrt{M}-3} \left[(-1)^{\left(\lfloor \frac{i}{2} \rfloor \right)} \left(\frac{\sqrt{M}}{2} - \lfloor \frac{i}{2} \rfloor - 1 \right) \operatorname{erfc} \left(\frac{d (2i+1)}{\sqrt{N_{0}}} \right) \right], \quad (8)$$

where the expression $\lfloor x \rfloor$ denotes the greatest integer less than *x*.

2) UNCODED BIT POSITIONS

In this section the probability of error of the uncoded bit positions described by the left-hand mapping of Table 1 is considered, under the successful decoded coset constraint. That is, the hard-decision error rate of the uncoded bits given the coded coset. Fig. 4 provides a graphical representation



Coded Bit Position (

of the hard decision regions under the constraint of decoded coset corresponding to $w_c = \{0, 0\}$, the quadrature coset pictured in Fig. 2b. The uncoded amplitude symbol considered here is also S = -15, corresponding to $w_u = \{0, 0\}$. For both bit positions, the upper red and blue regions in Fig. 4 indicate the error regions under standard hard-decision demodulation, while the green and orange regions denote correct and incorrect decision regions under the coded coset constraint. Note that, for the symbol (-15) considered here, the decision boundary distance for both bit positions is worse for the LCM coset constrained decision than for the standard approach shown above, but the worst-case boundary distance (which dominates error rate performance) is much better (4d rather than d for the least-significant bit of the mapping).

Again, the error rate for the bit positions of the example symbol is found by exploiting the complementary error function:

$$P_{b}^{u,\{S=-15\}}(0) = \frac{1}{2} \left[\operatorname{erfc}\left(\frac{4d}{\sqrt{N_{0}}}\right) - \operatorname{erfc}\left(\frac{20d}{\sqrt{N_{0}}}\right) \right], \quad (9)$$
$$P_{b}^{u,\{S=-15\}}(1) = \frac{1}{2} \left[\operatorname{erfc}\left(\frac{12d}{\sqrt{N_{0}}}\right) \right]. \quad (10)$$

And again, the average probability of error P_b^u in each position is the average across all symbols:

$$P_{b}^{u}(0) = \frac{1}{4\sqrt{M}} \begin{bmatrix} 2\operatorname{erfc}\left(\frac{4d}{\sqrt{N_{0}}}\right) + \operatorname{erfc}\left(\frac{12d}{\sqrt{N_{0}}}\right) - \operatorname{erfc}\left(\frac{20d}{\sqrt{N_{0}}}\right) \end{bmatrix}, (11)$$

$$P_{b}^{u}(1) = \frac{1}{4\sqrt{M}} \left[\operatorname{erfc}\left(\frac{4d}{\sqrt{N_{0}}}\right) + \operatorname{erfc}\left(\frac{12d}{\sqrt{N_{0}}}\right) \right]. (12)$$

Thus we have produced closed-form expressions for the hard-decision probability of error for both the coded bits, and the uncoded bits under coset-constrained hard-decision for 256-QAM and $B_c = 4$. The same approach may be applied for any square QAM constellation and any setting of B_c , the number of coded bits used.

B. EXPRESSIONS FOR THE GENERAL CASE

The development presented in the previous section translates in a straightforward fashion to the general case, allowing the



FIGURE 4. Error regions for the uncoded bits given the decoded coset.

general expression for the uncoded and coded error probability under hard decision for bit position b to be presented as

$$P_{b}^{u}(b) = \sum_{x=1}^{2^{b+1}-1} (-1)^{\lfloor \frac{x+1}{2}+1 \rfloor} 2^{\frac{B_{c}}{2}+1} \left(2^{b} - \lfloor \frac{x}{2} \rfloor\right)$$
$$\frac{2^{\frac{B_{u}-B_{u}^{\max}}{2}}}{\sum_{y=1}^{2}} \operatorname{erfc}\left(\frac{d}{N_{0}} 2^{\frac{B_{c}}{2}} \left(2^{\frac{B_{u}-B_{u}^{\max}}{2}+1}x - 2y + 1\right)\right), \quad (13)$$

where $B_{\mu}^{\max} = 2^{b+1}$.

.....

For the coded bits, the variation by bit position requires the less concise expressions:

$$P_{b}^{c \{B,B_{c}\}}(0) = \sum_{x=1}^{2^{\binom{B_{u}}{2}+1}-1} (-1)^{x+1} \left(2^{\binom{B_{u}}{2}+2}-2x\right)$$

$$\frac{2^{\binom{B_{c}}{2}-1}}{\sum_{y=1}^{2^{\binom{B_{c}}{2}}-1}} \operatorname{erfc}\left(\frac{d}{N_{0}}\left(2^{\binom{B_{c}}{2}}x-2y+1\right)\right), (14)$$

$$P_{b}^{c \{B,B_{c}\}}(1) = P_{b}^{c \{B,B_{c}\}}(0) + 2\sum_{x=1}^{2^{\binom{B_{u}}{2}+1}} (-1)^{x+1}$$

$$\frac{2^{\binom{B_{c}}{2}-2}}{\sum_{y=1}^{2^{-2}}} \operatorname{erfc}\left(\frac{d}{N_{0}}\left(2^{\binom{B_{c}}{2}}(x-1)+2y-1\right)\right), (15)$$

and finally:

$$P_b^{c\ \{B,B_c\}}(b) = P_b^{c\ \{B,B_c-2\}}(b-1) \quad \text{for} \quad b > 1, \ B_c > 2.$$
(16)

These expression allow the performance of the LCM system to be analysed for any choice of system parameters, and will be applied in Sections IV and V to develop analysis tools which can characterise the system for any choice of LDPC code, thereby fully characterising the performance of LCM scheme.

IV. EXIT CHARTS FOR LCM SYSTEMS

Prior work in the literature on BICM [21]–[23] has demonstrated that the effect on coded performance of the UEP encountered for bit positions in the constellation symbol mapping can be modelled by use of distinct binaryinput/continuous-output bit channels. In particular, the overall coded system performance can be modelled by producing a composite EXIT chart comprised of the EXIT charts of the LDPC code on each of the bit channels.

Note that the eqns. (4)-(16) are functions of $\frac{E_b}{N_0}$ as well as bit position, and this will be explicitly stated in following. Under the assumption that the channels encountered by individual bit positions may be treated as AWGN channels, the equivalent noise variance for each bit channel is then expressed as:

$$\sigma_{n,\text{equiv}}^2\left(b, \frac{E_b}{N_0}\right) = \frac{1}{2\operatorname{erfc}^{-1}\left(2P_b^c\left(b, \frac{E_b}{N_0}\right)\right)}.$$
 (17)

This equivalent bit channel noise variance is then used in the standard EXIT analysis variable node processor mutual information update given by [10]:

$$I_{E,V}\left(I_{A,V}, d_{v}, \sigma_{n}^{2}\right) = J\left(\sqrt{(d_{v}-1)\left[J^{-1}\left(I_{A,V}\right)\right]^{2} + \frac{4}{\sigma_{n}^{2}}}\right),$$
(18)

for variable node of weight d_v , where $I_{A,V}$ is the a priori mutual information. The term $I_{E,V}$ is the extrinsic mutual information at the output of the node processor, and the function $J(\cdot)$ represents the mutual information, which in this work is approximated analytically. Furthermore, the inverse mutual information update for the check node component decoder is approximated by [10]:

$$I_{A,C}(d_c) \approx 1 - J\left(\frac{J^{-1}(1 - I_{E,C})}{\sqrt{d_c - 1}}\right),$$
 (19)

where $I_{A,C}$ and $I_{E,C}$ are the a priori and extrinsic mutual information of the check node processor for node of weight d_c . Then the composite EXIT chart for the variable node processors across all node weights and bit channels is given by:

$$I_{E,V}\left(I_{A,V}, \frac{E_b}{N_0}\right)$$

= $\sum_{b=0}^{B_c/2} \sum_{i=2}^{d_{max}} \Lambda_i I_{E,V}\left(I_{A,V}, d_{v,i}, \sigma_{n,\text{equiv}}^2\left(b, \frac{E_b}{N_0}\right)\right),$ (20)

where Λ_i is the fraction of edges in the Tanner graph of the LDPC code connected to variable nodes of weight $d_{v,i}$. The proposed composite EXIT chart of (20) offers the advantage of using the standard formulation (18) from the literature with the derived bit channel equivalent noise variances for the LCM scheme, and thus does not require significant effort or increased computational complexity to perform, with respect to the EXIT analysis of a standard irregular LDPC code.

In Fig. 5 an example of the composite EXIT chart and its components is provided, for the threshold SNR point. The



FIGURE 5. EXIT chart for the 256-QAM case with $B_c = 4$.

component EXIT charts for the the variable nodes on both bit channels are provided in dashed and dotted plots, respectively. Note that while the two bit channels for this set of LCM parameters do not differ greatly in uncoded performance, the resulting component EXIT charts differ greatly from those observed for 256-QAM under Gray mapping in a BICM system, and thus correspond to a significant change in observed decoding outcome. Also worth noting in this plot is the effect of the weight-1 variable nodes which do not converge and thus harm overall code performance, which is the cost of their inclusion to allow linear-complexity encoding operation.

V. MUTUAL INFORMATION-BASED ANALYSIS

In this section, we make use of the observation that when the a priori mutual information equals zero, the extrinsic variable node processor information value provides an approximation to the channel mutual information for the considered SNR, i.e.

$$I_{ch}\left(\frac{E_b}{N_0}\right) \approx I_{E,V}\left(0,\frac{E_b}{N_0}\right),\tag{21}$$

at the ordinate axis in Fig. 5. This approximation holds very well for code rates above $\frac{1}{2}$.

A. CAPACITY ANALYSIS

In order to justify this approximation, a comparison will be made between the capacity plots of the standard BICM scheme derived by Monte Carlo simulation with those derived from $I_{E,V}(0, \frac{E_b}{N_0})$ through the same method outlined in Section IV, but for the probabilities of error for the bit channels of the gray mapping used for square QAM BICM, provided analytically in [25]. The plots for Monte Carlo and analytical approximate measures of MI for BICM are included in solid and dashed lines in Fig. 6, respectively. These plots demonstrate that for higher code rates (above 0.5), the approximation holds very well, as expected from the discussion found elsewhere in the literature [27], [28].



FIGURE 6. Capacity plots for the BICM and LCM systems.

Also included in that plot is the analytical approximation of the capacity curve for LCM, based on the development of Sections III and IV, in black dotted line. While this plot differs greatly for the BICM plot for the same QAM modulation order (also shown), it is worth remembering that this curve is for the coded bits of the LCM scheme only, which can be afforded a lower code rate R_{LDPC} to result in a higher overall system information rate R_{LCM} . Examining the curves for the LCM scheme at $R_{LDPC} = 0.75$ and the curve for the BICM scheme at R_{LCM} , we observe that the capacity of the schemes in terms of SNR is the same (indicated in vertical lines on the plot). This is an unsurprising result, the arrangement of bit-tosymbol mapping does not alter the information available from the channel. However, this result in combination with the simulation results presented in Section VI serves to validate the assumptions made in Section IV and here.

B. FINITE-LENGTH PERFORMANCE

The argument applied above to produce the capacity plot of the scheme can also be used to produce a semi-analytical performance predictor for LCM. From (18), when $I_{A,V} = 0$ and considering the BPSK channel we have:

$$\left(\mathrm{MI}_{Approx.}\right)^{\mathrm{BPSK}} = I_{E,V}\left(0, \left(\sigma_n^2\right)^{\mathrm{BPSK}}\right)$$
$$= J\left(\sqrt{\frac{4}{\left(\sigma_n^2\right)^{\mathrm{BPSK}}}}\right). \tag{22}$$

So testing the BPSK performance of a code for a given SNR range equally gives the BER performance in terms of approximate MI. Likewise, under the bit channel assumption of Sections III and IV, we have the noise variance estimates for the coded bit channels of the LCM scheme, as in (17), and the MI approximation follows directly as

$$\left(\mathrm{MI}_{Approx.}\right)^{\mathrm{LCM}} = \frac{1}{B_c} \sum_{b} J\left(\sqrt{\frac{4}{\sigma_{n,\mathrm{equiv}}^2\left(b,\frac{E_b}{N_0}\right)}}\right). \quad (23)$$

Mapping the BER results of the BPSK simulation to these approximate MI values for LCM directly provides the finitelength performance prediction, a semi-analytical method whereby one BPSK simulation is sufficient to characterise the coded bit performance of the LCM scheme across all settings of modulation order and number of coded bits, B_c for a given LDPC code. Of course, the same approach applies to the BICM schemes through the use of the probability of error expressions of [25], [26]. It is worth noting again that, as illustrated in Fig. 6, the approximation in (21) holds very well for code rates above 0.5 and thus, as will be shown in Section VI, the approach outlined here offers a very close approximation to the true performance. At lower code rates the predictor is overly optimistic and should be used only as a lower bound on performance.

It should also be noted that the literature is quite rich in work estimating the finite-length performance of LDPC codes on binary-input channels, for example in works such as [29]–[31], and where these works produce an error rate curve for the binary-input AWGN channel, these analytical results may substitute directly for the simulated results in the discussion above, and thus provide a wholly analytical estimate of finite-length performance for the LCM scheme. In this way, the translation of BPSK to LCM performance proposed in this section may make use not only of newly simulated Monte Carlo results and the archived results for established LDPC codes found in standards, but also provide a fast analytical performance estimation for newly designed codes.

The plots in Fig. 7 illustrate graphically the application of (22) and (23) to a set of simulated results for an LDPC code with block length $N_{\text{LDPC}} = 600$ and rate $R_{\text{LDPC}} = \frac{2}{3}$. In Fig. 7a the BPSK results are given, with MI of (22) on the lower axis and the corresponding $\frac{E_b}{N_0}$ on the upper axis. In Fig. 7b the transform from MI to $\frac{E_b}{N_0}$ is explicitly presented, derived from the capacity relation shown in Fig. 6, with $\frac{E_b}{N_0}$ taken for the $R_{\text{LDPC}} = \frac{2}{3}$ code under consideration rather than $\frac{E_s}{N_0}$ of Fig. 6, to allow visually the BPSK results to be translated to the predicted error rate performance in Fig. 7c. This method neatly allows the characterisation of finitelength performance of the LCM system across modulation order, coded/uncoded bit parameters and LDPC code used. Furthermore, since the translation of performance illustrated in Fig. 7 takes the form of a closed-form expression, the computational cost of producing the finite-length performance prediction for the LCM scheme is simply in the computation of the single set of BPSK performance results, either through Monte Carlo simulation, use of the analytical methods of [29]-[31], or reference to the archived results for known LDPC code graphs. In combination with the work

presented in [27], provided with the analytical expressions of Section III-B and the observations outlined here, the finitelength performance predictor can easily be extended to multicarrier modulation schemes such as OFDM and thus used for bit and power loading under frequency-dependent channels. The derived finite-length error rate performance of Fig. 7c is verified in Fig. 10 of Section VI.

VI. SIMULATION STUDY

In this section, the results from the simulation study of the LCM scheme are presented. The results validate the proposed analyses and offer insight on the relative performance of the LCM scheme, along with emphasising the benefits of the contributions in this work with regards to selection of code and system parameters for in the design of communications systems using the LCM scheme.

Throughout this section, the log-domain sum-product algorithm is used for LDPC decoding, with a maximum of 50 decoder iterations. The additive white Gaussian noise (AWGN) channel is considered in all cases (as presented in [23], the techniques applied in this work extend to other channels easily). The LDPC code rate for Fig. 9 is $R_{LDPC} = \frac{3}{4}$ and for Fig. 10 two code graphs with rate $R_{LDPC} = \frac{2}{3}$ were selected, while in Fig. 11 the LCM scheme uses the same $R_{LDPC} = \frac{3}{4}$ LDPC code of as Fig. 9 and the BICM scheme uses an LDPC code from the same class but with rate $R_{LDPC} = R_{LCM} = \frac{7}{8}$. The LDPC code block length is $N_{LDPC} = 12000$ unless otherwise stated.

In Fig. 8, the analytical expressions for the hard decision (uncoded) probability of error for the code bit mapping positions are verified for a number of LCM parameter settings, both modulation order M and number of coded bits B_c . These results are independent of the LDPC code selected. The solid line plots show the BER simulation results, while the markers represent the analytical probability of error for each case. This figure demonstrates the very close match between simulated and analytical results, as expected, and also demonstrates the UEP experienced by the coded bit positions when $B_c > 2$. This UEP is particularly relevant to the decoding threshold as it has the greatest effect the expected threshold region.

Fig. 9 investigates the LCM scheme performance for fixed modulation order, 256-QAM, and code rate $R_{LDPC} = \frac{3}{4}$ for the allowed settings of number of coded bits, namely $B_c = \{2, 4, 6\}$. Also included, in black vertical lines, are the coded bit decoding thresholds for each case, derived from the EXIT charts as described in Section IV. This serves to demonstrate the effectiveness of the bit channel approximation in characterising the effect of the UEP of the coded bit positions on decoding performance, and also on the assertion that the coded bit position threshold describes the performance threshold of the overall coded system. The EXITbased approach will provide a useful tool in predicting performance of communications systems using the LCM without necessitating long and computationally costly Monte Carlo simulations, allowing system designers to vary parameters as needed. This also presents interesting possibilities in terms



(c) The derived semi-analytical error rate performance of the LCM system for the considered parameters / code.

FIGURE 7. Example of constellation cosets for the LCM scheme.

of adaptive coding and modulation. Finally, for the plot of the $B_c = 2$ case, there is a clearly apparent deterioration in

system performance at high SNR. The uncoded bit position analytical probability of error (under the assumption of correct LDPC decoding), P_b^u , is also included for both the $B_c = 2$ and $B_c = 4$ cases, in red and cyan dashed plots, respectively. The uncoded bit position performance of the $B_c = 4$ system is significantly lower than the overall system performance as the error rate is dominated by the LDPC code, but for the $B_c = 2$ case the simulation results demonstrate that the uncoded bit position errors dominate performance leading to significant deterioration in performance. Thus the derived P_b^u expressions also provide a valuable design tool for the coded system designer.

In Fig. 10, the finite-length performance prediction based on the mutual information mapping, as described in Section V-B, is evaluated. For this plot, two LDPC code graphs with rate $\frac{2}{3}$ and block lengths $N_{LDPC} = 600$ and $N_{LDPC} = 48000$, respectively, were selected from the code class. The LCM parameters were 256-QAM with $B_c = 4$, resulting in $R_{LCM} = 0.83$. The plots labelled Simulation result from directly testing these codes in the LCM scheme through Monte Carlo simulation on the AWGN channel, while for the plots labelled MI-derived, the performance of each code was evaluated on the BPSK channel and then mapped to MI. The MI measure for the LCM scheme under consideration was then prepared and finally the plots produced, as outlined in Fig. 7. These plots demonstrate that the proposed semi-analytical method for performance prediction works remarkably well at short block length, while for the large block length code there is a gap between fully simulated and MI-derived plots of approximately 0.1dB. A possible reason for this gap is the use of the approximate LLR [32] which under the excellent performance offered by the large code may result in the small performance loss observed.

The final plot, Fig. 11, offers a performance comparison between the LCM scheme and other widely-used coded modulation schemes, namely those based on BICM and TCM. The scheme coded rates are indicated in the legend, with both LDPC-based schemes having rates of $\frac{7}{8}$, and the concatenated RS-TCM scheme having a rate close to this, based on the selection of $k_{RS} = 239$ with $n_{RS} = 255$ in the byte-oriented Reed Solomon code. Both TCM schemes employ the wellknown four-dimensional mapping of [14] with the trellis code specified in the G.fast standard [13]. For matched system code rate, as expected from the capacity analysis of Fig. 6 the two LDPC-based schemes should offer effectively the same threshold performance. Also as expected, the LDPCbased schemes offer a strong performance improvement over the classical schemes with a 2dB improvement over the concatenated RS-TCM scheme at an error rate of 10⁻⁷. Note that upon close observation of the LDPC-based scheme performance, while the thresholds are the same and they offer almost identical error rate performance in the BER region of 10^{-6} , below this point the BICM scheme is exhibiting error floor behaviour. This can be explained by the relatively high LDPC code rate required to match the LCM system rate, in conjunction with the constraints imposed by the code class,



FIGURE 8. Simulated and analytical plots for the coded bit positions uncoded probabilities of error.



FIGURE 9. Performance of the LCM system for varying parameter B_c .

particularly the quasi-cyclic circulant constraint. It should be noted that this behaviour is dependent on the particular code graph and class, and was also observed on the BPSK channel for the this code graph, rather than being related to the BICM system implementation. Thus, with careful code design the BICM scheme should perform as well as the LCM scheme. This result simply serves to illustrate another benefit offered through using LCM, namely the ability to achieve a high system rate with a lower rate LDPC code.

VII. CONCLUSION

In this work, the performance of the LCM scheme has been investigated. A number of analytical and semi-analytical expressions and techniques have been developed to describe the LCM performance, and have been demonstrated to very well characterise the scheme, particularly for a number of key properties which will heavily inform parameter selection and



FIGURE 10. Evaluation of the finite length performance prediction based on BPSK results.



FIGURE 11. Comparison of performance of coded modulation systems.

overall system design. Furthermore, the proposed techniques lend themselves to application to different channels and could therefore prove useful in an adaptive coding and modulation based communications system using the LCM scheme. Future contributions may include extension of the proposed methods to a scheme which uses LCM across the subcarriers in a multicarrier type system, which would be of particular interest for the wired community. Another interesting future work would be in the application of the methods developed here to an LCM system which includes trellis shaping for conditioning the transmitted signal, where the EXIT-based analysis may particularly be applicable. The contributions in this work should prove timely as the LCM scheme has been selected for inclusion in the next generation DSL standard.

REFERENCES

 R. G. Gallager, Low-Density Parity-Check Codes. Cambridge, MA, USA: MIT Press, 1963.

- [2] Digital Video Broadcasting (DVB); Second Generation Framing Structure, Channel Coding and Modulation Systems for Broadcasting, Interactive Services, News Gathering and Other Broadband Satellite Applications (DVB-S2), ETSI Standard EN 302 307-1 V1.4.1, 2005.
- [3] G.hn: Unified High-Speed Wire-Line Based Home Networking Transceivers—System Architecture and Physical Layer Specification, document ITU-T Rec. G.9960, 2009.
- [4] WiFi: Local and Metropolitan Area Networks—Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughputs, IEEE Standard 802.11, 2009.
- [5] 10 GBASE-T: Local and Metropolitan Area Networks—Carrier Sense Multiple Access With Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications, IEEE Standard 802.3, 2006.
- [6] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Improved low-density parity-check codes using irregular graphs," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 585–598, Feb. 2001.
- [7] D. Divsalar, H. Jin, and R. McElliece, "Coding theorems for turbo-like codes," in *Proc. 36th Allerton Conf. Commun., Control, Comput.*, 1998, pp. 201–210.
- [8] H. Jin, A. Khandekar, and R. McEliece, "Irregular repeat accumulate codes," in Proc. 2nd Int. Symp. Turbo Codes Rel. Topics, 2000, pp. 1–8.
- [9] M. P. C. Fossorier, "Quasi-cyclic low-density parity-check codes from circulant permutation matrices," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1788–1793, Aug. 2004.
- [10] S. T. Brink, G. Kramer, and A. Ashikhmin, "Design of low-density paritycheck codes for modulation and detection," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670–678, Apr. 2004.
- [11] T. S. M. Richardson and R. Urbanke, "Design of capacityapproaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.
- [12] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
- [13] Fast Access to Subscriber Terminals (G.fast)—Physical Layer Specification, document ITU-T Rec. G.9701, 2014.
- [14] L.-F. Wei, "Trellis-coded modulation with multidimensional constellations," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 4, pp. 483–501, Jul. 1987.
- [15] V. Oksman, R. Strobel, T. Starr, J. Maes, W. Coomans, M. Kuipers, E. B. Tovim, and D. Wei, "MGFAST: A new generation of copper broadband access," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 14–21, Aug. 2019.
- [16] E. Eleftheriou and S. Olcer, "Low-density parity-check codes for digital subscriber lines," in *Proc. IEEE ICC*, New York, NY, USA, vol. 3, Apr. 2002, pp. 1752–1757.
- [17] E. Eleftheriou, S. Olcer, and H. Sadjadpour, "Application of capacity approaching coding techniques to digital subscriber lines," *IEEE Commun. Mag.*, vol. 42, no. 4, pp. 88–94, Apr. 2004.
- [18] K. Narayan and J. Li, "Bandwidth efficient low density parity check coding using multilevel coding and iterative multistage decoding," in *Proc. 2nd Int. Symp. Turbo Codes Rel. Topics*, Brest, France, Sep. 2000, pp. 165–168.
- [19] N. von Deetzen and S. Sandberg, "Design of unequal error protection LDPC codes for higher order constellations," in *Proc. IEEE Int. Conf. Commun.*, Glasgow, Scotland, Jun. 2007, pp. 926–931.
- [20] Y. Li and W. E. Ryan, "Bit-reliability mapping in LDPC-coded modulation systems," *IEEE Commun. Lett.*, vol. 9, no. 1, pp. 1–3, Jan. 2005.
- [21] L. Gong, L. Gui, B. Liu, B. Rong, Y. Xu, Y. Wu, and W. Zhang, "Improve the performance of LDPC coded QAM by selective bit mapping in terrestrial broadcasting system," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 263–269, Jun. 2011.
- [22] K. Yan, T. Cheng, F. Yang, K. Peng, and J. Song, "Improved design of bit mapping based on EXIT-chart analysis for DVB-T2 system," *IEEE Trans. Consum. Electron.*, vol. 57, no. 4, pp. 1579–1585, Nov. 2011.
- [23] C. Healy, A. Al Rawil, and C. Tsimenidis, "Joint interleaver and LDPC code parameter design for beyond G. Fast systems," in *Proc. IEEE GLOBECOM*, Abu Dhabi, UAE, 2018, pp. 1–6.
- [24] X.-Y. Hu, E. Eleftheriou, and D. M. Arnold, "Regular and irregular progressive edge-growth tanner graphs," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 386–398, Jan. 2005.
- [25] K. Cho and D. Yoon, "On the general BER expression of one- and twodimensional amplitude modulations," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1074–1080, Jul. 2002.
- [26] P. K. Vitthaladevuni, M.-S. Alouini, and J. C. Kieffer, "Exact BER computation for cross QAM constellations," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 3039–3050, Nov. 2005.

- [27] Y. Li and W. Ryan, "Mutual-information-based adaptive bit-loading algorithms for LDPC-coded OFDM," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1670–1680, May 2007.
- [28] J. Neckebroek, "Error correction, precoding and bitloading algorithms in high-speed access networks," M.S. thesis, Ghent Univ., Ghent, Belgium, 2016.
- [29] A. Amraoui, A. Montanari, T. Richardson, and R. Urbanke, "Finite-length scaling for iteratively decoded LDPC ensembles," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 473–498, Feb. 2009.
- [30] M. Noor-A-Rahim, K. D. Nguyen, and G. Lechner, "Finite length analysis of LDPC codes," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Istanbul, Turkey, Apr. 2014, pp. 206–211.
- [31] R. Yazdani and M. Ardakani, "Waterfall performance analysis of finitelength LDPC codes on symmetric channels," *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3183–3187, Nov. 2009.
- [32] A. J. Viterbi, "An intuitive justification and a simplified implementation of the MAP decoder for convolutional codes," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 260–264, Feb. 1998.



CORNELIUS TOMAS HEALY received the M.Sc. (by Research) and Ph.D. degrees in communications from the University of York, in 2010 and 2014, respectively. In 2015, he was a Postdoctoral Researcher with the Centre for Telecommunications Research, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil. Since 2016, he has been a Research Associate with the School of Engineering, Newcastle University. His main research interests include error control coding,

coded modulation, and iterative receiver design.



ANAS AL RAWI (Member, IEEE) received the M.Sc. (Hons.) and Ph.D. degrees in communications and signal processing from Newcastle University, U.K., in 2007 and 2011, respectively. From 2010 to 2012, he was a Postdoctoral Researcher with the School of Electrical and Electronic Engineering, Swansea University, Swansea, U.K. In 2012, he was a Research Associate with the Institute of Electronics, Communications and Information Technology, Queens University of

Belfast, Belfast, U.K. He is currently the Research Manager of BT Labs and a Royal Society Industry Fellow with the Cavendish Laboratory, Cambridge University. Since joining BT in 2012, he has been involved in developing novel transmission techniques to push the capacity of wireline and wireless systems beyond current practical limits. His research interests include computational electromagnetics, transmission cross-layer optimization, cooperative networks, and MIMO systems.



CHARALAMPOS C. TSIMENIDIS (Senior Member, IEEE) received the M.Sc. (Hons.) and Ph.D. degrees in communications and signal processing from Newcastle University, in 1999 and 2002, respectively. He is currently a Reader in digital communications with the School of Engineering, Newcastle University. During the last 15 years, he has published more than 180 conference and journal papers, supervised successfully three M.Phil. and 45 Ph.D. students, and made contributions in

the area of receiver design to several European and U.K. funded research projects. His main research interests include adaptive array receivers for wireless communications, including demodulation and error correction algorithms for doubly spread multi-path channels.