Wireless Communications and Mobile Computing

# Error Control Codes for Next-Generation Communication Systems: Opportunities and Challenges

Lead Guest Editor: Zesong Fei
Guest Editors: Jinhong Yuan and Qin Huang

# Error Control Codes for Next-Generation Communication Systems: Opportunities and Challenges

Wireless Communications and Mobile Computing

# Error Control Codes for Next-Generation Communication Systems: Opportunities and Challenges

Lead Guest Editor: Zesong Fei
Guest Editors: Jinhong Yuan and Qin Huang

# Editorial Board

# Contents

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 2643205, 2 pages
https://doi.org/10.1155/2018/2643205

*Editorial*

# Error Control Codes for Next-Generation Communication Systems: Opportunities and Challenges

**Zesong Fei**,[1] **Jinhong Yuan**,[2] **and Qin Huang**[3]

[1]*School of Information and Electronics, Beijing Institute of Technology, Beijing 10081, China*
[2]*School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, NSW 2052, Australia*
[3]*School of Electronic and Information Engineering, Beihang University, Beijing 100191, China*

Correspondence should be addressed to Zesong Fei; feizesong@bit.edu.cn

Received 3 October 2018; Accepted 3 October 2018; Published 2 December 2018



Error control codes are widely applied in modern communication systems to improve the bandwidth-power efficiency and the reliability of data transmissions. Modern error control codes have attracted the interest of scholars and industry partners since Turbo codes were invented. For example, Turbo codes have been used in the 4G cellular mobile systems. Nowadays, LDPC codes and the polar codes are adopted in the 5G standard. The recent development on the theoretic framework of new channel coding theorem for finite code length will provide guidelines for future practical error control codes designs.

In the age of IoT, everything will be connected via communication links. It is expected that the next-generation communication systems need to support many scenarios such as wireless communications, optical communications, distributed storage systems, V2X networks, and sensor networks. These scenarios will impose new requirements to the communication systems ranging from lower complexity encoder/decoder, lower delay or latencies, ultrareliable transmission at rates close to the Shannon capacity, low energy consumptions, etc. In addition to the communication systems, error control codes also find emerging applications in security, flash memories, and deep-space probing.

This special issue is a collection of 11 papers which explore the performance of error control codes for the next generation communication systems and discuss the opportunities and challenges that they will face.

"Superposition Coded Modulation Based Faster-Than-Nyquist Signaling", by S. Li et al., presented a transmission scheme of faster-than-Nyquist signaling combined with superposition coded modulation. The proposed scheme requires a lower SNR than orthogonal signaling with a larger alphabet to achieve a wide range of spectral efficiencies.

"A Novel Design of Downlink Control Information Encoding and Decoding Based on Polar Codes", by C. Sun et al., proposed a novel design on downlink control information encoding and decoding. The proposed scheme could support dynamic configuration of transmission modes with low complexity. It is shown that the proposed scheme can comply with the false alarm rate target of 5G standard.

"Performance Analysis of CRC Codes for Systematic and Nonsystematic Polar Codes with List Decoding", by T. Murata and H. Ochiai, studied the effect of the length and generator polynomials of CRC codes on frame error rate performance of polar codes with list decoding. The authors found that the length of CRC will affect the performance significantly, while the generator polynomials will only affect nonsystematic polar codes.

"Adding a Rate-1 Third Dimension to Parallel Concatenated Systematic Polar Code: 3D Polar Code", by Z. Liu et al., proposed a three-dimensional polar code scheme to improve the error floor performance of parallel concatenated systematic polar codes. The key idea of the proposed scheme is that it makes the best use of the characteristic of parallel concatenation and serial concatenation.

"Construction and Decoding of Rate-Compatible Globally Coupled LDPC Codes", by J. Zhang et al., presented a family of rate-compatible globally coupled LDPC codes which provide more flexibility in code rate and guarantee the structural property of algebraic construction. The authors also proposed a modified low complexity decoding scheme.

"Construction of Quasi-Cyclic LDPC Codes Based on Fundamental Theorem of Arithmetic", by H. Zhu et al., studied the construction of Quasi-Cyclic LDOC codes based on an arbitrary given expansion factor. The constructed codes perform well in AWGN channels when iterative decoding algorithms are used.

"Code-Hopping Based Transmission Scheme for Wireless Physical-Layer Security", by L. Yin and W. Hao, proposed a code-hopping based secrecy transmission scheme that uses dynamic nonsystematic LDPC codes and automatic repeat-request (ARQ) mechanism to jointly encode and encrypt source messages. In this scheme, source message is jointly encoded and encrypted by a parity-check matrix which is dynamically selected from a set of LDPC matrices based on the shared secret key. The authors showed that the key is difficult for the eavesdropper to generate, and the security gap is small.

"Research and Implementation of Rateless Spinal Codes Based Massive MIMO System", by L. Wang et al., proposed a spinal codes-aided massive MIMO scheme and a multilevel puncturing and dynamic block-size allocation scheme. The proposed schemes can approach the maximum achievable rate, and a comparable MIMO testbed is established to demonstrate the effectiveness of the proposed scheme.

"Design and Analysis of Adaptive Message Coding on LDPC Decoder with Faulty Storage", by G. G and L. Yin, discussed the impacts of message errors on LDPC decoders with unreliable memories. An adaptive coding scheme based on the magnitude level of messages was also proposed to improve the robustness.

"A CCM-Based OFDM System with Low PAPR for Sparse Source", by Q. Yang et al., introduced compressive coded modulation scheme to restrain peak-to-average power ratio in OFDM systems, which becomes severe when the source is sparse.

"Efficient Quantization with Linear Index Coding for Deep-Space Images", by R. Mahmood et al., proposed a modified quantization with a linear index coding scheme to improve its spectral efficiency and robustness. The proposed scheme efficiently removes the redundant bit-planes for spectrally efficient linear index coding of images. A multipass decoding scheme is also proposed which provides better gain by using extrinsic information.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

## Acknowledgments

The editors thank all of the authors for their submissions to this special issue. We are also grateful to the anonymous reviewers for their valuable comments to improve the quality of the articles. We hope that this special issue will further encourage research interests and engineering practice in the area of Error Control Codes.

*Zesong Fei*
*Jinhong Yuan*
*Qin Huang*

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 6387214, 13 pages
https://doi.org/10.1155/2018/6387214

*Research Article*

# Efficient Quantization with Linear Index Coding for Deep-Space Images

**Rehan Mahmood**,[1,2] **Zulin Wang**,[1,3] **and Qin Huang**[1]

[1]*School of Electronic and Information Engineering, Beihang University, Beijing 100191, China*
[2]*Institute of Space Technology, Islamabad, Pakistan*
[3]*Collaborative Innovation Center of Geospatial Technology, Wuhan 430079, China*

Correspondence should be addressed to Qin Huang; qinhuang@buaa.edu.cn

Received 24 November 2017; Revised 14 May 2018; Accepted 10 September 2018; Published 11 October 2018

Academic Editor: Gonzalo Vazquez-Vilar



Due to inevitable propagation delay involved in deep-space communication systems, very high cost is associated with the retransmission of erroneous segments. Quantization with linear index coding (QLIC) scheme is known to provide compression along with robust transmission of deep-space images, and thus the likelihood of retransmissions is significantly reduced. This paper aims to improve its spectral efficiency as well as robustness. First, multiple quantization refinement levels per transmitted source block of QLIC are proposed to increase spectral efficiency. Then, iterative multipass decoding is introduced to jointly decode the subsource symbol-planes. It achieves better PSNR of the reconstructed image as compared to the baseline one-pass decoding approach of QLIC.

## 1. Introduction

The deep-space communication is more challenging than its near-Earth counterpart due to the associated huge inter-transceiver propagation delay [1, 2]. The inevitable propagation delay increases the likelihood of larger fluctuations in channel SNR during transmission, e.g., due to antenna pointing errors, atmospheric conditions, etc. [3, 4]. The error correcting codes defined in [5, 6] provide excellent performance for both near-Earth and deep-space communication systems. However, the postdecoding bit-error rate (BER) increases dramatically when the channel SNR degrades slightly below the decoding threshold of the code selected for transmission. The retransmission of erroneous frames is possibly required with a low rate code, if the postdecoding BER is higher than the value acceptable by the target application. Therefore, visually acceptable reconstruction quality of the images transmitted from deep-space despite moderate-to-high BER is considered as a desirable feature.

It is well-known that the state-of-the-art image compression standards (JPEG2000 and ICER) are sensitive to postdecoding residual errors [7–9] due to their embedded fixed-to-variable length entropy coders. Even a single error after the channel decoder makes the rest of the bitstream unsuitable for decoding. Although this catastrophic propagation of errors is spatially confined by partitioning the image into segments, the complete loss of certain segments may reduce the visual quality of reconstructed image. Considering the strict integrity requirement of deep-space scientific images, retransmission of the lost segments is necessary.

In order to provide error resilient source transmission, a vast number of schemes are proposed in literature under the category of joint source-channel coding (JSCC) [10–18]. A comprehensive survey of such schemes aiming towards robust transmission of images can be found in [19–24] and references therein. Particularly, this paper is interested in the JSCC based linear index coding scheme, which is known to reconstruct better visual quality images on mismatched deep-space channels. The scheme, referred to as quantization with linear index coding (QLIC), is proposed in [24] for the transmission of deep-space images using Raptor codes [25]. It replaces the concatenation of entropy coding and channel

coding with a single linear encoding map, and channel codes are used to provide both source compression and error protection.

Indeed, QLIC is able to withstand significant SNR fluctuations while still preserving the visual quality of the image, but its spectral efficiency is inferior to the currently used deep-space image transmission system [26, 27]. Therefore, this paper is aimed towards enhancing the spectral efficiency as well as the robustness of QLIC. Firstly, the proposed encoder utilizes multiple refinement levels to efficiently reduce the redundant symbol-planes thus increasing the overall spectral efficiency. Then, we show that the multipass decoding of QLIC provides significant gain, if the information which maximizes the virtual channel capacity is utilized in the subsequent decoding passes. Our iterative decoding provides better PSNR of the reconstructed image as compared to the baseline one-pass decoding.

On the encoder side, QLIC [24] determines the minimum number of symbol-planes from partitioned subbands in order to achieve a target distortion. The baseline QLIC uses the same block length $K$ for partitioned subbands as well as for the channel codeword. On one hand, large $K$ is required for capacity achieving performance of channel codes. On the other hand, small $K$ is necessary to efficiently identify the redundant transform coefficients. As a result, the spectral efficiency is compromised in both the cases. This paper proposes to partition the subbands in small block lengths while still using large block length for channel coding. Consequently, the proposed arrangement decreases the entropy rate of symbol-planes and thus less budget is required for a given channel capacity. Simulation results show that the proposed approach reduces the transmission overhead by up to 40% while still preserving the robustness feature of QLIC. Besides, it is shown that the degree distribution of the Raptor codes can be optimized by only considering the virtual correlation channel between the symbol-planes.

On the decoder side, QLIC [24] utilizes one-way virtual correlation channel among the symbol-planes. Consequently, higher significant symbol-planes provide their decoding decisions to the subsequent lower significant symbol-planes in a multistage manner. In this paper, we show that lower significant symbol-planes can also provide new extrinsic information to the higher significant symbol-planes by executing multiple decoding passes. In fact, different from [28, 29], only the reliably recovered symbol of a lower level is used to provide effective extrinsic information, if its corresponding symbol of higher level was recovered with low reliability in the previous decoding pass. Therefore, we propose to only utilize the information from those combinations of the symbol-planes which results in the maximum capacity of its observed virtual correlation channel. As a result, the decoding performance as well as the quality of the reconstructed image will improve with multipass decoder. It is shown in Section 4.2 that multipass decoding provides a gain of up to 1.5 dB in terms of PSNR of the reconstructed image as compared to the baseline one-pass decoding approach.

The main contributions of this paper are summarized below:

(i) An efficient quantization scheme is proposed for QLIC to assign quantization precision to the transformed image considering relatively small blocks. The proposed scheme achieves better overall transmission bandwidth efficiency.

(ii) A multipass decoding scheme is proposed to utilize the redundancy left in the lower level symbol-planes in order to improve the BER of the higher level symbol-planes. The improved BER of the higher levels eventually results in improved reconstruction quality.

The rest of this paper is organized as follows: In Section 2, we present the background of QLIC and notations used in the paper. The proposed idea of multiple refinement levels per subblock is discussed in Section 3 whereas the details of multipass decoding approach are described in Section 4. Section 5 concludes the paper.

## 2. Quantization with Linear Index Coding

In this section, we briefly outline QLIC scheme, the optimization problem, the solution of which is used in Section 3 for the assignment of multiple refinement levels to each source block. We also introduce the baseline encoding-decoding approach and the notations used throughout the paper which are consistent with the notations in [24, 31].

In classical separated source-channel coding systems the aim of the source coding is to reduce the redundancy from the source as much as possible and then the error protection is provided by the channel codes. In this way, controlled amount of redundancy is added to the transmitted data stream in order to protect it against channel induced errors. However, in QLIC, channel codes are used to provide compression as well as error protection. The source data after quantization is arranged into bit-planes and the bit-planes are directly mapped to the channel codewords. Consequently, the rate-budget assigned to each encoded bit-plane is directly proportional to the conditional entropy rate of the bit-plane given the higher level bit-planes. The direct mapping of the bit-planes has been shown to be more robust against channel induced errors [24].

Let us consider that an image after $W$ level biorthogonal discrete wavelet transform (DWT) is divided into $s = 2^{2W}$ parallel source components with block length $K$ equal to the length of LL0 subband, where LL0 represents the low pass subband after 2D wavelet transform. In fact, LL0 is the subsampled version of the original image after DWT. Let $\mathbf{z}^{(i)} = (z_1^{(i)}, \ldots, z_K^{(i)})$ represent the sequence of the $i$th source transform coefficients, where $i = 1, \ldots, s$. All the operations in the context of the QLIC are essentially the same for every source component except for the LL0 subband. Therefore, similar to [24], we do not consider the encoding and decoding of the LL0 subband and assume that it is available error-free at the receiver. However, in Sections 3.3 and 4.2, we consider the transmission overhead of LL0 subband in comparing the

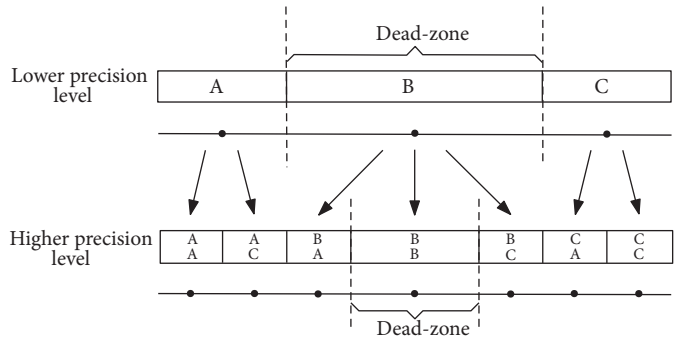



Figure 1: Dead-zone uniform scaler quantizer (DZUSQ) $Q_k$.

simulation results of the proposed enhancements with the baseline QLIC scheme.

Let $Q_k : \mathbb{R} \longrightarrow \{A, B, C\}^p$, for $k = 1, \ldots, K$, be an embedded dead-zone uniform scaler quantizer (DZUSQ) applied to the transform coefficients $\mathbf{z}$, where $p = P$ is the highest level of refinement and $p = 1$ is the lowest one. The alphabets and decision regions of DZUSQ are shown in Figure 1. We denote $D^{(i)}_{\mathcal{Q},p}$ as the quantization distortion of the $i$th source component at the $p$th quantization level and $\mathbf{u} = \mathcal{Q}(\mathbf{z})$ as the block of ternary quantization indices arranged in a two-dimensional $P \times K$ array. The $p$th row of $\mathbf{u}$, denoted by $\mathbf{u}^{(p)} = (u_{p,1}, \ldots, u_{p,K})$, is referred to as the $p$th "symbol-plane." All the symbols-planes from 1 to $p$ are included for a refinement level of $p$. According to the chain rule of mutual information, the entropy rate of every subsource $H^{(i)} = \sum_{p=1}^{P} H^{(i)}_p$, where

$$H_p = \frac{1}{K} H\left(\mathbf{u}^{(p)} \mid \mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(p-1)}\right), \tag{1}$$

for $p = 1, \ldots, P$, is the entropy rate of each symbol-plane in bits/source symbol.

Let $r_i(d)$ denote the operational rate-distortion (R-D) function of the $i$th source component with respect to the mean-square error (MSE), where $d = (1/K)\mathbb{E}[||\mathbf{z} - \widehat{\mathbf{z}}||^2]$ and $\widehat{\mathbf{z}}$ is the estimate of $\mathbf{z}$ after transmission using a suitable source-channel code. $r_i(d)$ is then given by the lower convex envelope of the following set of points considering the concatenation of various refinement levels of quantizer and ideal entropy coding:

$$\left(\sum_{j=1}^{p} H^{(i)}_j, D^{(i)}_{\mathcal{Q},p}\right), \quad p = 0, \ldots, P, \tag{2}$$

where $D^{(i)}_{\mathcal{Q},0} = \sigma_i^2$ by definition. Since $r_i(d)$ is a piecewise linear and convex function, it is possible to represent it with a family of straight lines $a_{i,p}d + b_{i,p}$ obtained by joining the consecutive R-D points in (2).

Let us consider that a successive refinement source-channel code exists which encodes all the subsources up to a particular refinement level such that the overall distortion is given by

$$D = \frac{1}{s}\sum_{i=1}^{s} \nu_i d_i, \tag{3}$$

where $\{\nu_i\}$ is a set of nonnegative weights. The nonnegative weights are due to the fact that MSE distortion in the pixel domain is not equal to the MSE distortion in the transform domain for a biorhtogonal transform. Further, if the source-channel code transmits the data in $N$ channel uses, the resulting transmission bandwidth efficiency for the corresponding scheme is given by $b = N/sK$. This notation is analogous to the "bit per pixel" (bpp) concept used in image coding and we will use $b$ to compare the spectral efficiency of our proposed enhancement in Section 3.3 with the baseline scheme.

Consequently, the refinement level $P$ which is necessary for every subsource to achieve an overall distortion $D$ is then the result of the following weighted MSE distortion linear program

$$\text{minimize} \quad \frac{1}{s}\sum_{i=1}^{s} \nu_i d_i \tag{4}$$

$$\begin{aligned} \text{subject to:} \quad & \frac{1}{s}\sum_{i=1}^{s} \gamma_i \le bC \\ & D^{(i)}_{\mathcal{Q},P} \le d_i \le \sigma_i^2, \quad \forall i \\ & \gamma_i \ge a_{i,p} d + b_{i,p}, \quad \forall i, p. \end{aligned} \tag{5}$$

The above program considers idealized channel codes; however, practical channel codes require an overhead $\theta$ for successful decoding. The modified set of R-D points for the practical channel codes is given by

$$\left(\sum_{j=1}^{p} H^{(i)}_j \left(1 + \theta^{(i)}_j\right), D^{(i)}_{\mathcal{Q},p}\right), \quad p = 0, \ldots, P, \tag{6}$$

where the overhead $\theta$ can be determined experimentally for a particular family of channel codes. This modified set of R-D points is then used to determine $a_{i,p}$ and $b_{i,p}$ in (4) which are in fact the coefficients of the straight line.

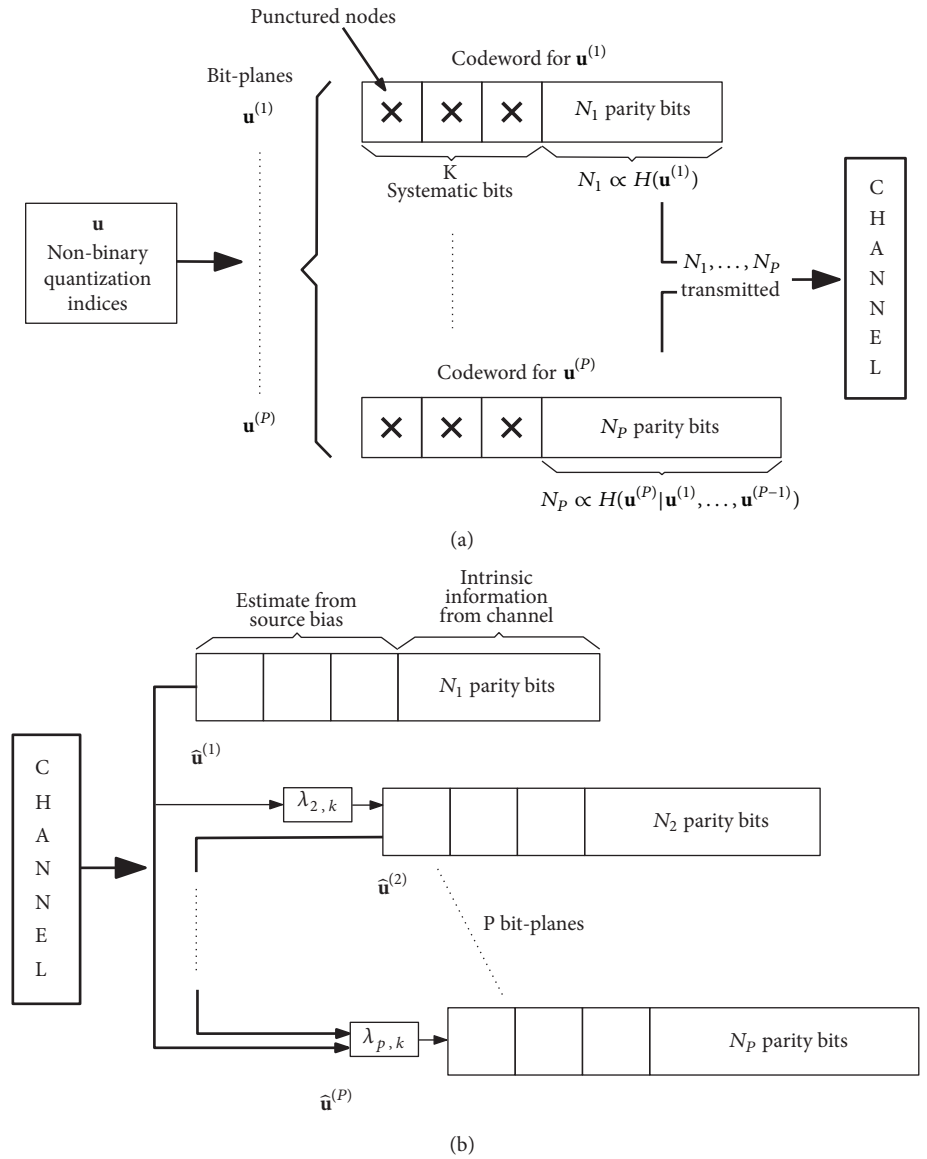



Figure 2: (a) Multilayer encoding approach of QLIC. (b) Baseline one-pass multistage decoder. The higher significant bit-planes provide decoding decisions to the lower significant bit-planes.

The solution of the above program is then used to quantize each subsource up to a refinement level of $P$. In order to use binary Raptor codes, the symbol-planes are decomposed into bit-planes by considering ternary to binary alphabets mapping, the details of which are given in [31]. Therefore, in the rest of the paper we use the term "bit-planes" instead of "symbol-planes." The bit-planes of each block $\mathbf{u}$ are directly encoded to the channel codeword with multilayer encoding as shown in Figure 2(a). The systematic bits are punctured and only coded bits are transmitted. On the decoder side, the decoding is performed in a multistage manner starting from the highest significant bit-plane as shown in Figure 2(b). Each $p$th stage decoder makes use of the source probability model $\Pr(\mathbf{u}^{(p)} \mid \mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(p-1)})$ which is already conveyed to the decoder as a part of the header. Consequently, the coded nodes receive intrinsic information from the channel. The systematic nodes, however, receive the message

$$\lambda_{p,k} = \log \frac{P\left(u_{p,k} = 0 \mid \widehat{u}_{1,k}, \ldots, \widehat{u}_{p-1,k}\right)}{P\left(u_{p,k} = 1 \mid \widehat{u}_{1,k}, \ldots, \widehat{u}_{p-1,k}\right)}, \tag{7}$$

where $\widehat{u}_{1,k}, \ldots, \widehat{u}_{p-1,k}$ denote the hard decisions obtained from the already decoded higher significant bit-planes.

The soft information can also be used instead of hard decisions; however, no significant improvement in performance was observed as mentioned in [11]. The errors, if any, in the higher significant bit-planes will eventually feed the lower levels with false log-likelihood ratios (LLRs), though

the error propagation is not as catastrophic as for the variable-to-fixed length decoding. Both hard and soft reconstruction are possible considering the posterior probability $\Pr(u_{p,k} \mid \mathbf{y})$, where $\mathbf{y}$ is the channel output. In the rest of the paper, we refer to this multistage decoder as the one-pass baseline *multistage decoder.*

## 3. Multiple Refinement Levels

Different from JPEG2000 and ICER which use sophisticated context based probability models to derive the entropy coder, the pure compression performance of QLIC is derived by the solution of optimization problem in (4). The solution identifies the refinement levels $p$ for every subsource block to achieve a target overall distortion. The $p$ bit-planes are then directly encoded to the channel codewords for transmission. The pure compression performance of QLIC, i.e., using ideal channel codes, is inferior to the compression performance provided by the state-of-the-art image compression systems. This difference, however, may become significant in practical transmission scenario due to the overhead associated with the nonideal performance of finite-length channel codes.

The high level and low level transform coefficients are usually not uniformly distributed within the subband due to various image features. The choice of $K$, corresponding to $I/2^W \times I/2^W$ portion of the subband, is thus critical for better solution of the optimization problem, where $I \times I$ are the dimensions of the image. Since modern block codes tend to be capacity achieving asymptotically, a large $K$ favors increased spectral efficiency due to relatively low $\theta$. Therefore, a relatively larger block length of $K = 16384$ is used in the numerical results of [24, 31].

However, large block length is associated with a serious drawback. The optimization problem to find out the refinement levels considers each subsource as a single unit. Consequently, the solution assigns the same refinement level to all the transform coefficients within a subsource. However, due to different spatial location of high and low level coefficients within a subsource, this assignment of refinement levels is usually not optimal. For example, let us consider a $128 \times 128$ length quantized subsource of an image taken from Mars exploration rover as shown in Figure 3. A refinement level of $p$ is used for quantization. The zero values are represented with black pixels whereas the rest of the values are represented with white pixels. Let us focus on quadrant II of the image. Very few pixels in quadrant II have nonzero values and are quantized with a refinement level of $p$. However, it is likely that if these values are quantized with a refinement level lower than $p$, it will not significantly affect the reconstruction quality of the image. In that case, the conditional entropy rate of the bit-planes also decreases which increases the overall spectral efficiency.

*3.1. Proposed Encoding Approach.* A possible solution is to use small block lengths to find the quantization refinement levels. Figure 4 compares the pure compression performance of QLIC for block lengths of $K = 16384$ and $K = 4096$. It is clear from Figure 4 that the pure compression performance for $K = 4096$ is superior; e.g., PSNR of 48 dB is achieved

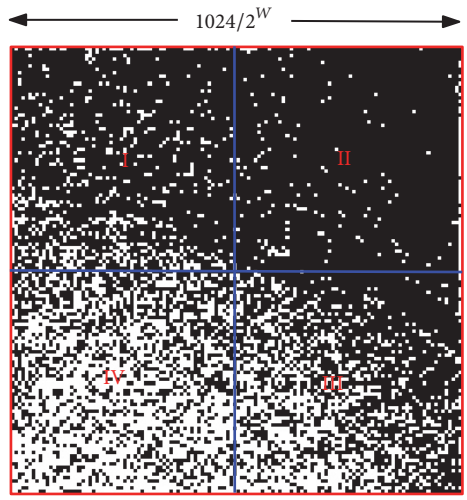



Figure 3: Binary map of a single subsource of an image from Mars exploration rover. The nonzero quantization indices are represented with white pixels.

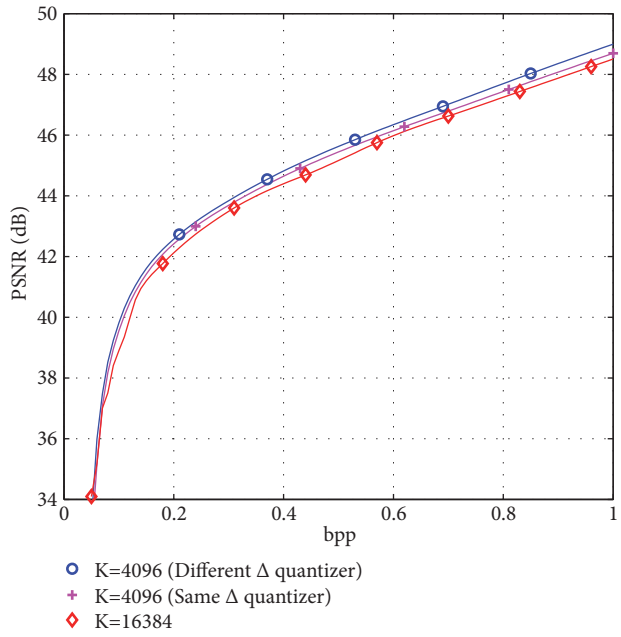



Figure 4: The pure compression performance of QLIC using ideal channel codes for an image taken from Mars exploration rover. The compression performance for short block lengths is superior to that of the large block lengths.

at 0.86 bpp for $K = 4096$, whereas 0.92 bpp is required in case of $K = 16384$ to achieve the same PSNR. Although the pure compression performance for $K = 4096$ is better, the encoding of bit-planes with short block length channel codes requires significant overhead for convergence. The improved compression performance is thus easily overwhelmed by the accumulated overhead of the channel codes.

Another possible solution is to combine the promising features of both the block lengths; i.e., use small $K$ to find the solution of the optimization problem whereas use large block

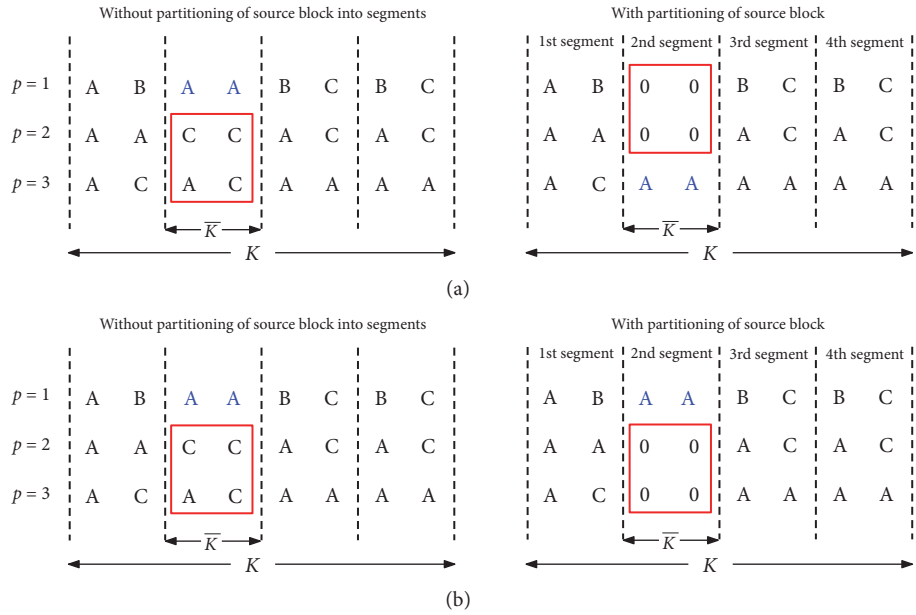



FIGURE 5: (a) Typical index assignment after reduced refinement level. (b) Proposed approach of index assignment.

length channel codes for transmission. A trivial approach is to further partition each subsource into $j = 2^L$ segments, where $L$ is a power of 2 and the length of each segment is $\overline{K} = K/2^L$. The optimization problem (4) is then solved by feeding $s \times j$ subsource segments as input. The obtained refinement levels are eventually used to quantize each segment of length $\overline{K}$ thus generating $p$ bit-planes per segment. The $j$ segments of the same subsource can then be concatenated together to make the overall block length $K$. For example, if $L = 2$, there will be $j = 4$ segments per subsource which corresponds to $K = 16384$ and $\overline{K} = 4096$. Consequently, each bit-plane can be encoded and transmitted using a channel code of block length $K$.

The above scheme is straightforward but unable to improve the spectral efficiency in practice. Every segment associated with a subsource may have different range of transform coefficients. Therefore, the quantization step size $\Delta$ which depends on the dynamic range of transform coefficients in every segment is different. In other words, the quantization of every $j$th segment may result in a different codebook. Thus, the quantization index assignment is also different even for the same value coefficients but associated with different segments of the same subsource. Consequently, the overall entropy rate increases after concatenation of all the segments of the subsource as compared to the case when the same refinement level is used by considering the subsource as a whole, i.e., composed of all the $j$ segments. Since the rate budget $N$ assigned to each bit-plane is dependent on its entropy rate, the overall spectral efficiency decreases. In fact, the resulting spectral efficiency is even worse as compared to the large block length case.

Therefore, we propose to use the overall dynamic range of the transform coefficients of a subsource to determine the quantization step size $\Delta$ for every $j$th segment. In other words, the codebook for all the $j$ segments remains the same. Although this approach is suboptimal in the sense that a wider $\Delta$ is used, we observed that it works quite well as input to the optimization problem (4). The solution of the optimization problem may identify segments to be quantized with lower $p$ as compared to the case when optimization is performed by considering the whole subsource as a single segment. As shown in Figure 4, although the pure compression in this case is reduced relative to the previous arrangement, the conditional entropy rate of the bit-planes is also decreased. Apparently, less budget is required for encoding which ultimately increases the overall spectral efficiency. The conditional entropy rate of the bit-planes is decreased due to the fact that certain quantization symbols are replaced with "0" ("B" in our assignment of quantization alphabets). This process is equivalent to the one as shown in Figure 5(a) which depicts that the 2nd segment is originally to be quantized with $p = 3$ but later reduced to $p = 1$.

The above solution enables getting rid of redundant symbols-planes quite effectively and thus increases the overall spectral efficiency. However, there are also some associated drawbacks. Let us assume that $p = 3$ is the subsource refinement level whereas $p_{seg} = 1$ is the refinement level of its $j$th segment obtained with help of the procedure explained in the earlier paragraphs. Figure 5(a) shows the $p \times K$ arrangement of the quantization alphabets without partitioning the source block into segments as well as with partitioning. The red square in the figure shows the quantization alphabets which become insignificant due to the partitioning of the source block; i.e., they are represented with "zeros." In other words, it is necessary to quantize the 2nd segment with $p = 3$ without partitioning of the source block but after partitioning its quantization precision requirement is reduced to $p = 1$. Consequently, in the overall arrangement of the quantization

alphabets of the $K$-length source block, e.g., "A A" appears as if it is reduced in precision with reference to the other segments. Therefore, it is necessary to convey information to the receiver about its updated refinement level. This overhead becomes significant particularly in case of small $\overline{K}$. Since lower bit-planes are more prone to errors in case of mismatched channel SNR, the quantization alphabets of the 2nd segment are more susceptible to channel noise for the partitioning case although they originally correspond to higher significant level. The quantization symbols with decreased significance are thus more vulnerable to errors and robustness of QLIC is compromised.

In order to overcome this issue, we propose to update the $p \times K$ arrangement of the quantization alphabets as shown in Figure 5(b). This arrangement is robust in the sense that it preserves the significance of higher significant bit-planes. Consequently, it alleviates the problem associated with the multistage decoding in case of mismatched channel capacity. Added to that, it is no longer required to convey the information about the updated refinement levels to the receiver. The reason can be quickly explained by considering the ternary quantization alphabets and the DZUSQ such that $B$ is the quantization alphabet which corresponds to 0. By design [31], the quantizer cells corresponding to alphabets "A" and "C" can only be subdivided into "A" and "C" for further refinement. Therefore, the assignment of "B" (0) to the higher refinement level where the lower refinement level is alphabet "A" or "C" is illegal and can easily be spotted by the dequantizer.

*3.2. Robustness of Raptor Codes.* The multilayer encoding approach of QLIC directly maps the symbols-planes to the channel codewords as shown in Figure 2. The rate budget $N_p$ assigned to each bit-plane is in proportion to the conditional entropy rate of the bit-plane given the higher significant bit-planes and is given by

$$N_p = K\left(\frac{H_p}{C} + \theta_p\right), \tag{8}$$

where $\theta_p > 0$ is a small rate margin associated with the particular family of channel codes used for encoding. The layered encoding uses systematic channel codes such that the systematic symbols are punctured before transmission and only $N_p$ parity bits are transmitted. Since $N_p$ is dependent on $H_p$, discrete set of coding rates may result in reduced bandwidth efficiency as it is difficult to generate matching code rates for every possible $H_p$. However, this issue can be alleviated with the use of rateless codes which have an added advantage of virtually generating infinite number of coding rates.

The nonuniversality of Raptor codes is well-known for general noisy channels [32]. Due to multilayer encoding, every bit-plane observes a composite channel at the decoder. The parity symbols observe a physical channel with capacity $C$ whereas the systematic symbols observe a virtual correlation channel with capacity $(1 - H_p)$. For capacity achieving performance, the output degree distribution of Raptor codes must be optimized for the observed composite channel.

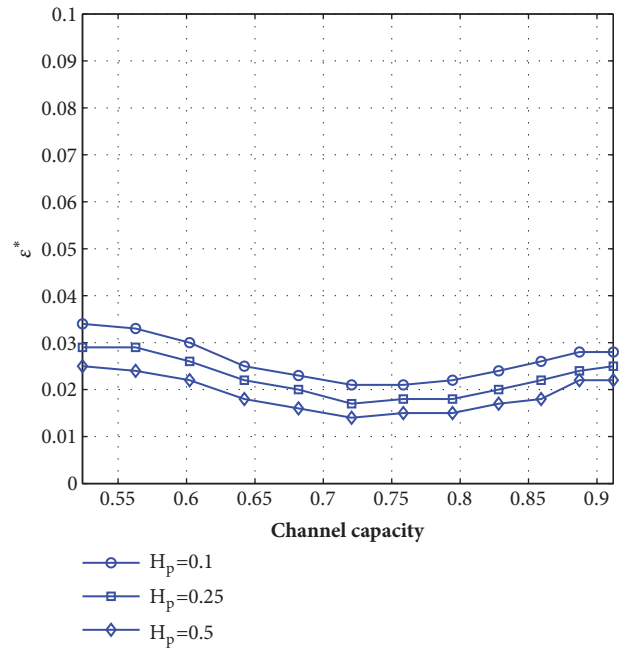



Figure 6: The decoding threshold of Raptor codes for different channel capacities. The output degree distribution for all plots is optimized for SNR of 3dB. It appears that there is no big difference in the decoding thresholds of Raptor codes at various channel SNRs.

This is achieved by finding the matching distribution for various $(H_p, C)$ pairs on a sufficiently fine grid as proposed in [24]. However, it is quite impractical to optimize the output degree distribution for every $(H_p, C)$ pair. Even on a grid with uniform spacing of 0.01 bits/sec for both $H_p$ and $C$, the number of output degree distribution will be in terms of thousands. Therefore, in the following we show that the output degree distribution of the Raptor codes optimized for a particular $(H_p, C)$ pair is robust to the variations in $C$. According to [33], this is particularly true when the joint decoding of the constituent codes (LT and LDPC) is performed.

Inspired by the work of [33], we numerically compute the threshold $\epsilon^*$ of systematic binary Raptor codes for the case of composite channel, i.e., by considering both the virtual correlation channel and the physical channel. In order to observe the effect of channel capacity $C$ on a degree distribution optimized for a particular $(H_p, C)$ pair, we fix $H_p$ and numerically find the threshold for different values of $C$. If the decoding is successful, we increase the code rate and repeat the process with a new ensemble of Raptor codes and different noise realization. Similarly, if the decoding fails, we decrease the code rate and repeat the same process. The consistency of $\epsilon^*$ is then established by repeating the same process a large number of times and averaging the results. The decoding threshold of the Raptor codes degree distribution optimized for the $(H_p, C)$ pairs: $(0.1, 0.72)$, $(0.25, 0.72)$, and $(0.5, 0.72)$ are shown in Figure 6. The numerical results show that there is no big difference in the decoding threshold of Raptor codes, optimized for a particular physical channel capacity, over the range of channel capacities relevant in deep-space

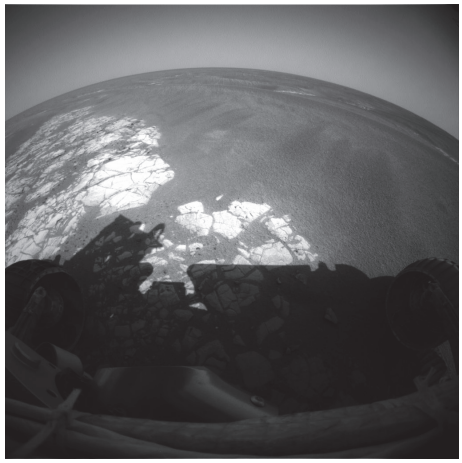


Figure 7: Image taken from Mars exploration rover: MER1 [30].

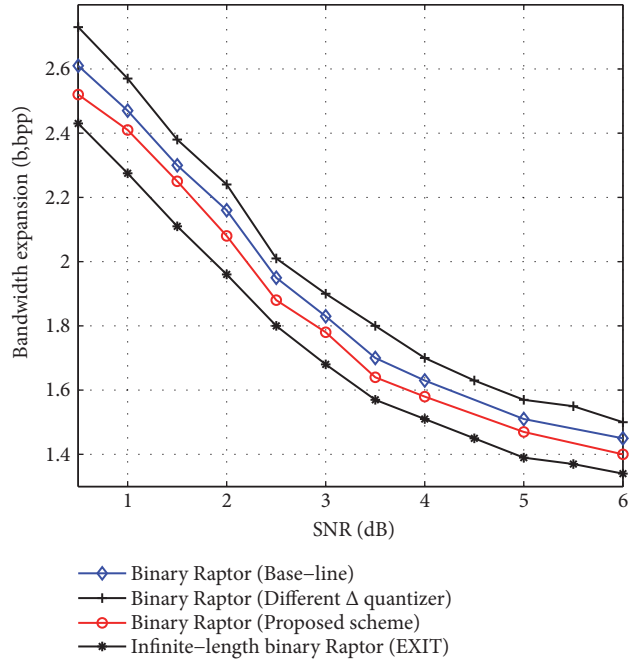



Figure 8: The spectral efficiency of QLIC at various SNR levels for MER-B image. The plots depict that the proposed QLIC scheme with multiple quantization precision per source block possesses superior spectral efficiency.

communication. Therefore, it is not necessary to optimize the degree distribution of Raptor codes for every value of $C$. It is reasonable to perform the optimization for every $(H_p, C_{constant})$ pair, which leads to system simplification. The channel capacity $C_{constant}$ may correspond to the middle of the range of relevant channel SNRs. In fact, in all the numerical results, we used the degree distribution of Raptor codes optimized for SNR of 3 dB. As an example, the degree distribution of the LT part of the Raptor codes optimized for the $(H_p, C)$ pair $(0.1, 0.72)$ is given by

$$\begin{aligned}\Omega(x) = {} & 0.0043x^1 + 0.4856x^2 + 0.1341x^3 + 0.1607x^4 \\ & + 0.10976x^5 + 0.0140x^8 + 0.0547x^{16} \\ & + 0.0015x^{41} + 0.0355x^{42}.\end{aligned} \tag{9}$$

A high rate regular LDPC code with degree (2,100) and rate 0.98 is used as a precode in all the Raptor codes.

*3.3. Spectral Efficiency Comparison.* In this subsection, we present the simulation results depicting the spectral efficiency of QLIC using the multiple refinement levels approach. We also compare these results with the numerical simulations of [24] in order to highlight the improved spectral efficiency of our proposed approach. The test image used in all the simulations results is from Mars exploration rover and referred to as MER1. The image is shown in Figure 7. It is the same image used in [24]. It is a BW uncoded image and bears a resolution of 1024x1024 pixels with 12-bit pixel value. The image is compressed using a 3-level Cohen–Daubechies–Feauvea (CDF) 9/7 wavelet transform as defined in JPEG2000 standard for lossy compression [8]. The transform coefficients $\mathbf{z}$ are then quantized with a UDZSQ, producing ternary quantization indices. The ternary quantization indices are further decomposed into binary indices to facilitate the use of binary Raptor codes [31]. Considering the resolution of the image, it is divided into $s = 64$ subsources of 128x128 pixels each, to make it compatible with the packet length defined in CCSDS recommendations [26], i.e., $K = 16384$. Every subsource is further subdivided into $j = 4$ segments to apply the multiple refinement levels concept.

A target PSNR of 49 dB is defined for the reconstructed image. The LL0 subband [8] is not transmitted using the QLIC approach. Instead, it is transmitted along with the header using a rate 1/3 channel code which results in negligible overhead as explained in [24]. We included this overhead in the simulation results although it accounts for only 0.004 in $b$ (bit-plane). We reproduce the results of [24] using binary Raptor codes for comparison purpose. Figure 8 compares the spectral efficiency of the baseline as well as the proposed QLIC approach achieved at various values of channel SNR. The following comments are in order:

(i) The ($\diamond$)-curve corresponds to spectral efficiency achieved by the baseline QLIC scheme using binary Raptor codes. Similar to the numerical results of [24], the degree distribution of the Raptor codes is optimized by considering both the virtual correlation channel and the physical channel on a sufficiently fine scale. Therefore, this curve serves as a benchmark to compare the spectral efficiency of the proposed scheme.

(ii) The (+)-curve corresponds to the spectral efficiency for the case when small block length is used in the encoding process. The length of each segment $\overline{K} = 4096$; however, the quantization step size $\Delta$ is assigned according to the dynamic range of the transform coefficient within the segment. As explained earlier, due

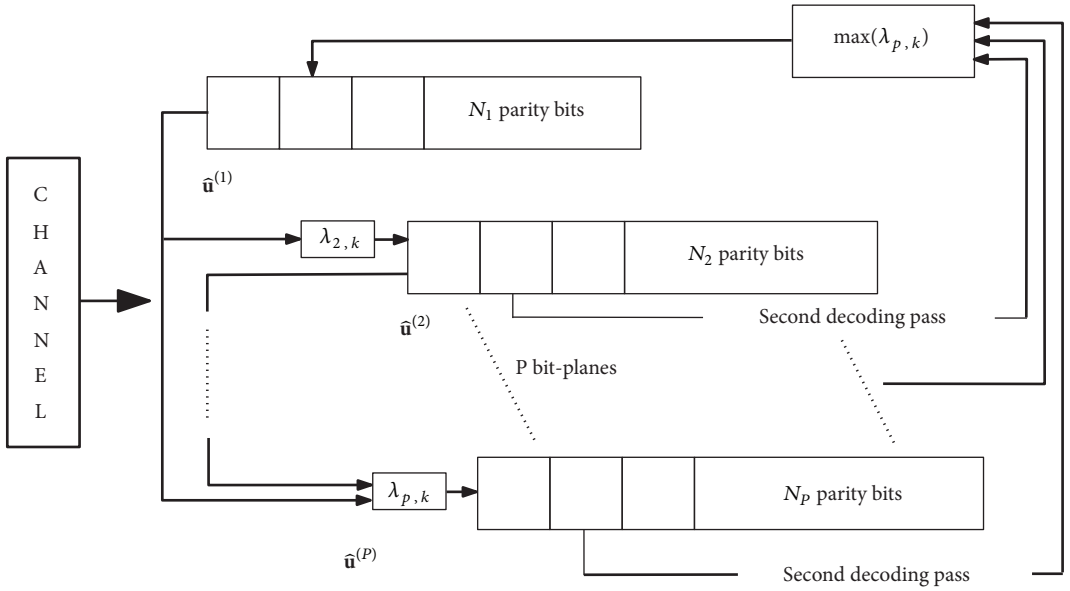



Figure 9: Proposed multipass decoding approach.

to the increased randomness of the data the resulting spectral efficiency is even worse as compared to the baseline scheme.

(iii) The (○)-curve corresponds to the spectral efficiency achieved by the proposed approach. We would like to highlight that, in order to obtain the spectral efficiency at various values of channel SNR, only Raptor codes optimized for 3 dB SNR are used. The gap between the proposed scheme (○)-curve and the baseline scheme (◇)-curve is significant in terms of transmission efficiency. For example, 2.16 bpp is required by the baseline scheme at SNR of 2 dB. However, only 2.08 bpp is required by the proposed scheme in order to achieve the same performance. This corresponds to 40% reduced transmission overhead with reference to the performance of infinite length channel codes at 2 dB SNR denoted by (∗) in Figure 8.

(iv) The (∗)-curve shows the bandwidth expansion for the infinite length channel codes using EXIT charts.

## 4. Multipass Decoding

In this section, we first establish the importance of observed virtual correlation channel in multipass decoding. Then, we present the analysis of multipass decoding. It reveals that certain symbols recovered in the first decoding pass can only provide effective extrinsic information in the subsequent decoding passes. In particular, a reliably recovered symbol of a lower level can only provide effective extrinsic information, if its corresponding symbol of higher level was recovered with low reliability in the previous decoding pass. Therefore, we propose to utilize information from only those combinations of the bit-planes which results in the higher extrinsic information. The block diagram of the proposed multipass decoder is shown in Figure 9.

*4.1. Proposed Decoding Approach.* During decoding, every level of the multistage decoder observes a composite channel. A fraction $\gamma^{(p)}$ of the systematic output nodes (output nodes in accordance with the Raptor codes) in the decoding bipartite graph are connected to the virtual correlation channel with capacity $C_v$. The remaining fraction of output nodes $(1-\gamma^{(p)})$ are connected to the physical channel with capacity $C_{phy}$, where $\gamma^{(p)} = K/(K+N)$. A higher value of the ratio $r_c^{(p)} = \gamma^{(p)}/(1-\gamma^{(p)})$ indicates more dependence on the decoding decisions of higher significant bit-planes than the physical transmission channel output $\mathbf{y}^{(p)}$.

The overall observed channel capacity at the $p$th decoding stage can be given by $C_{total}^{(p)} = C_{phy}^{(p)} + C_v^{(p)}$, where $C_v^{(p)} = (1 - H_p)$. Since the assignment of parity budget $N_p$ is probably different for every bit-plane, the channel capacity $C_{total}^{(p)}$ observed by each bit-plane for convergence is also different. Therefore, in order to compare the overall observed channel capacity at various decoding levels, we introduce the normalized average channel capacity $\mathbf{C}$ in the following, considering the BER of the higher levels.

Let the multistage decoder at every decoding level provide the estimate $\hat{\mathbf{u}}^{(p)}$ of $\mathbf{u}^{(p)}$ after running sufficient iterations of belief-propagation algorithm. Further, $P_c^{(p)} = \Pr(\hat{\mathbf{u}}^{(p)} = \mathbf{u}^{(p)})$ is the probability that the decoding decisions of the higher $(p-1)$ decoding levels are correct. The normalized bit-level average channel capacity $\mathbf{C}^{(p)}$ observed at every $p$th level can then be defined as

$$\mathbf{C}^{(p)} = \left(\frac{\left(C_{phy} \times N_p\right)}{K} - \boldsymbol{\delta}_{phy}\right) + \left(1 - H_p\right) \times \prod_{i=1}^{p-1} P_c^{(i)}, \quad (10)$$

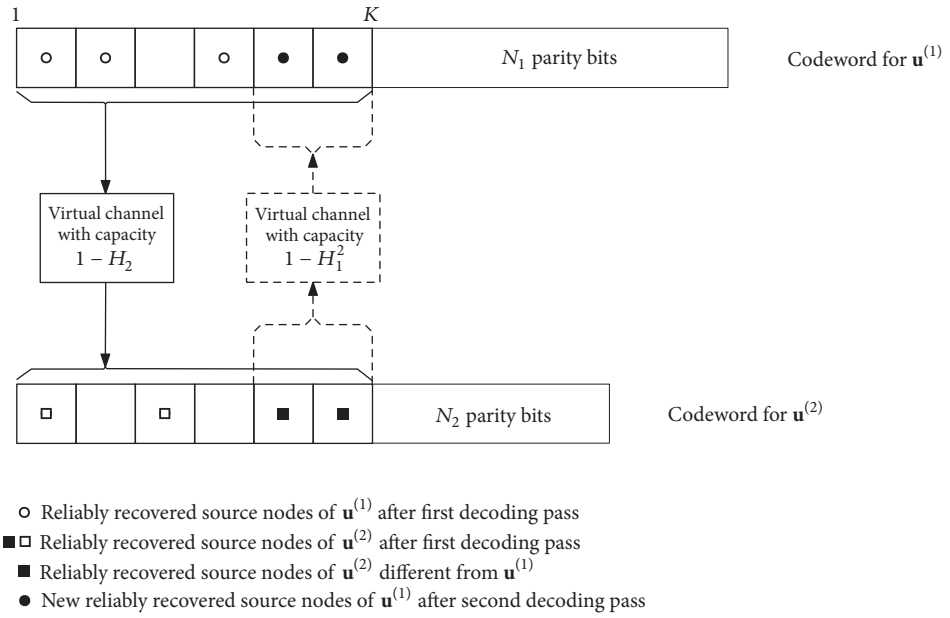



Figure 10: Multipass decoding approach depicting the decoding of two correlated sources $\mathbf{u}^{(1)}$ and $\mathbf{u}^{(2)}$.

where $\boldsymbol{\delta}_{phy}$ denotes the degradation in channel capacity from its nominal value. $\mathbf{C}^{(p)}$ becomes 1 provided that $\boldsymbol{\delta}_{phy} = 0$ and $\prod_{i=1}^{p-1} P_c^{(i)} = 1$.

Now let us consider the transmission using QLIC for $P = 2$ such that $\mathbf{u}^{(1)}$ and $\mathbf{u}^{(2)}$ are the bit-planes to be transmitted over a binary input-output symmetric channel with nominal capacity $C_{phy}$. The rate budget assigned to $\mathbf{u}^{(1)}$ is proportional to the marginal entropy rate of $\mathbf{u}^{(1)}$, whereas $\mathbf{u}^{(2)}$ is assigned a rate budget in proportion to its conditional entropy rate given $\mathbf{u}^{(1)}$. $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$ are received corresponding to the transmission of $\mathbf{u}^{(1)}$ and $\mathbf{u}^{(2)}$, respectively, over a mismatched transmission channel with capacity $(C_{phy} - \boldsymbol{\delta}_{phy})$. The baseline multistage decoder first decodes $\mathbf{y}^{(1)}$ and outputs the estimate $\hat{\mathbf{u}}^{(1)}$. The decoding of $\mathbf{y}^{(2)}$ is then followed such that $\hat{\mathbf{u}}^{(1)}$ is available and LLRs according to (7) are provided to the systematic nodes of $\mathbf{u}^{(2)}$.

The normalized average channel capacities $\mathbf{C}_{\longrightarrow 1}^{(1)}$ and $\mathbf{C}_{\longrightarrow 1}^{(2)}$ observed by $\mathbf{u}^{(1)}$ and $\mathbf{u}^{(2)}$, respectively, during the first decoding pass (baseline multistage decoder), are given by

$$\begin{aligned} \mathbf{C}_{\longrightarrow 1}^{(1)} &= \left( \frac{\left(C_{phy} \times N_1\right)}{K} - \boldsymbol{\delta}_{phy} \right) + \left(1 - H_1\right) \times P_c^{(0)}, \\ \mathbf{C}_{\longrightarrow 1}^{(2)} &= \left( \frac{\left(C_{phy} \times N_2\right)}{K} - \boldsymbol{\delta}_{phy} \right) + \left(1 - H_2\right) \times P_c^{(1)}, \end{aligned} \tag{11}$$

where $P_c^{(0)} = 1$ due to the independent decoding of $\mathbf{u}^{(1)}$.

Let us define $\mathscr{A}^{(p)} = \{x : x \in \mathscr{K}\}$, where $\mathscr{K} = \{x : x = 1, \ldots, K\}$ is an index set which labels the $K$ systematic nodes of the decoding bipartite graph. Now consider that every decoding stage of the multistage decoder is equipped with a genie aided decoder which is capable of correctly identifying the reliably recovered source nodes after sufficient iterations of BP algorithm. Consequently, it is possible to only provide the true LLRs $\{\lambda_{x,k} : x \in \mathscr{A}^{(1)}\}$ to the corresponding systematic nodes of $\mathbf{u}^{(2)}$. The LLRs $\{\lambda_{x,k} : x \in \mathscr{K} : x \notin \mathscr{A}^{(1)}\}$ are assumed as erased and thus set to zero. Obviously, such a decoder will restrict the propagation of errors to the subsequent decoding stages. Further, this genie aided decoder is somehow similar to the case when the soft information recovered from the $(1 - p)$ decoding levels is used to scale $\lambda_{p,k}$. Indeed for practical multipass decoding, we used this later approach.

The decoded output $\hat{\mathbf{u}}^{(1)}$ for any arbitrary value $\boldsymbol{\delta}_{phy} > 0$ can only be improved by integrating new extrinsic information at the systematic nodes of the decoding bipartite graph. In other words, $P_c^{(1)}$ can only be increased by enhancing the capacity of the virtual channel $C_v^{(1)}$ by executing a second decoding-pass and utilizing the extrinsic information available from the decoded output $\hat{\mathbf{u}}^{(2)}$. Now let us consider the following two hypothetical scenarios for the second decoding pass of $\mathbf{u}^{(1)}$ with the assumption that $|\mathscr{A}^{(1)}| > |\mathscr{A}^{(2)}|$, where $|.|$ is the cardinality of $\mathscr{A}$. This assumption is true in case of layered QLIC encoding approach.

(i) Scenario 1: $\mathscr{A}^{(2)} \cap \mathscr{A}^{(1)} = \mathscr{A}^{(2)}$; i.e., the reliably recovered nodes indices of $\hat{\mathbf{u}}^{(2)}$ are exactly the same as of $\hat{\mathbf{u}}^{(1)}$ in the first decoding pass. Now if a second decoding pass is executed for $\mathbf{u}^{(1)}$, new extrinsic information from $\hat{\mathbf{u}}^{(2)}$ is available only for those systematic nodes of $\mathbf{u}^{(1)}$ which were already recovered reliably in the first decoding pass as shown in Figure 10. Effectively, negligible improvement in $C_v^{(1)}$ is observed and hence no further improvement in the decoding

performance of $\mathbf{u}^{(1)}$ is possible in the subsequent decoding passes.

(ii) Scenario 2: now consider that all the nodes recovered in the first decoding pass of $\mathbf{u}^{(2)}$ are different from the recovered nodes of $\mathbf{u}^{(1)}$; i.e., $\mathscr{A}^{(2)} \cap \mathscr{A}^{(1)} = \emptyset$. In this case, new extrinsic information from $\widehat{\mathbf{u}}^{(2)}$ is available corresponding to the nodes of $\mathbf{u}^{(1)}$ which were not recovered in the first decoding pass. Consequently, a second decoding pass of $\mathbf{u}^{(1)}$ is likely to increase $P_c^{(1)}$ as compared to the first decoding pass.

Consequently, let us define $P_d^{(2)} = \Pr(\mathscr{A}^{(2)} \cap \mathscr{A}^{(1)})$ as the probability with which the reliably recovered nodes indices of $\widehat{\mathbf{u}}^{(2)}$ are different from the reliably recovered nodes indices of $\widehat{\mathbf{u}}^{(1)}$ after the first decoding pass. We can estimate $P_d^{(p)}$ for the case of two correlated sources as

$$P_d^{(p)} = \left(P_c^{(p)}\left(1 - P_c^{(p-1)}\right)\left(1 - \gamma_p\right)\right). \tag{12}$$

The updated normalized average capacity observed by $\mathbf{u}^{(1)}$ in the second decoding pass is thus given by

$$\mathbf{C}_{\longrightarrow 2}^{(1)} = \left(\frac{\left(C_{phy} \times N_1\right)}{K} - \boldsymbol{\delta}_{phy}\right) + \left(\left(1 - H_1\right) P_c^{(0)}\right) + \left(\left(1 - H_1^2\right) P_d^{(2)}\right), \tag{13}$$

where $H_1^2$ is the conditional entropy rate of $\mathbf{u}^{(1)}$ w.r.t. $\mathbf{u}^{(2)}$. Consequently, if $((1 - H_1^2)P_d^{(2)}) > 0$, $\mathbf{C}_{\longrightarrow 2}^{(1)} > \mathbf{C}_{\longrightarrow 1}^{(1)}$ and the decoding performance of $\mathbf{u}^{(1)}$ will be improved. Further, $\mathbf{C}_{\longrightarrow 2}^{(2)}$ is given by

$$\mathbf{C}_{\longrightarrow 2}^{(2)} = \left(\frac{\left(C_{phy} \times N_2\right)}{K} - \boldsymbol{\delta}_{phy}\right) + \left(\left(1 - H_2\right) P_c^{(1)}\right). \tag{14}$$

According to (13), multipass decoding gain is dependent not only on the correlation among the bit-planes but also on the probability $P_d$. However, contrary to the genie aided decoding case, it is not possible to identify the reliably recovered nodes for practical decoders. Therefore, the practical solution is to use the reliability information which is implict in the soft decisions. The reliability information can then be used to scale the LLRs of the systematic nodes in the subsequent decoding passes. Therefore, different from [24], the decoding with soft decisions is necessary for the multipass approach. Therefore, in the following we use the soft decisions estimates for practical case.

The extrinsic LLR $L_{ext}^{(p)}$ for the $p$th bit-plane in each decoding pass can be represented as

$$L_{ext}^{(p)} = \frac{P\left(\widehat{u}_{p,k} = 0\right)}{P\left(\widehat{u}_{p,k} = 1\right)}, \tag{15}$$

where $P(\widehat{u}_{p,k} = 0)$ and $P(\widehat{u}_{p,k} = 1)$ can be expressed with reference to all the bit-planes except $p$ as follows.

$$P\left(\widehat{u}_{p,k} = 0\right) = P\left(u_{p,k} = 0 \mid u_{(\mathscr{X}_1,k)}, \ldots, u_{(\mathscr{X}_z,k)}\right) \times P\left(\widehat{u}_{(\mathscr{X}_1,k)}, \ldots, \widehat{u}_{(\mathscr{X}_z,k)}\right) \tag{16}$$

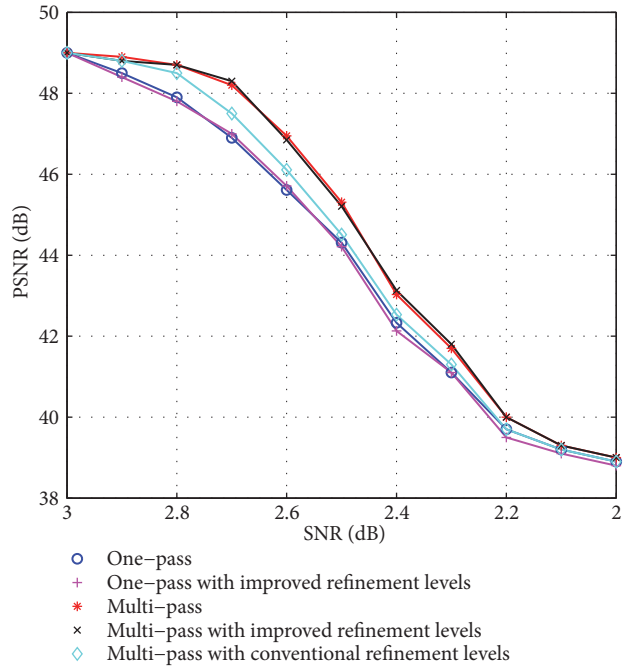



Figure 11: Reconstructed PSNR of MER1 when channel SNR varies form its nominal value of 3 dB.

and

$$P\left(\widehat{u}_{p,k} = 1\right) = P\left(u_{p,k} = 1 \mid u_{(\mathscr{X}_1,k)}, \ldots, u_{(\mathscr{X}_z,k)}\right) \times P\left(\widehat{u}_{(\mathscr{X}_1,k)}, \ldots, \widehat{u}_{(\mathscr{X}_z,k)}\right) \tag{17}$$

where $\mathscr{X} = \{x : x \in \mathscr{P} : x \neq p\}$, $z = |\mathscr{X}|$ and $\mathscr{P} = \{1, \ldots, P\}$ is an index set which labels all the bit-planes. The multipass decoding thus improves the decoding performance by providing $L_{ext}^{(p)}$ to each bit-plane according to (15). However, the performance improvement even by using the soft decisions is not very significant as shown in Figure 11 ((∗)-curve). We explain the reason in the following and then improve the multipass decoding.

In case of more than two bit-planes, different amount of correlation exists between the combinations of the bit-planes. For example, in case of 3 bit-planes, $P(\mathbf{u}^{(1)} \mid \mathbf{u}^{(2)})$, $P(\mathbf{u}^{(1)} \mid \mathbf{u}^{(3)})$, and $P(\mathbf{u}^{(1)} \mid \mathbf{u}^{(2)}, \mathbf{u}^{(3)})$ are known at the decoder for the most significant bit-plane. Due to the multistage decoding, certain symbols are recovered with very low reliability in case of mismatched channel SNR. Particularly, this is true for bit-planes with high $r_c^{(p)}$. Added to that, the layered encoding approach makes the lower significant bit-planes more susceptible to channel noise. It is mainly because the bit-planes are transmitted over a combination of virtual and physical channel and thus affected by the noise on both the channels.

Therefore, it is likely that including symbols from every bit-plane without any criterion in calculating the extrinsic information may decrease $L_{ext}^{(p)}$. This issue becomes more significant as the number of bit-planes increases. Therefore, excluding such low reliability symbols for the calculation of

extrinsic information may increase the multipass decoding performance. Indeed, it results in higher capacity of the virtual correlation channel observed by the $p$th bit-plane as compared to the former case. However, it is difficult to define a threshold to identify the unreliable symbols. Therefore, in order to get maximum benefit of multipass decoding for every $p$th bit-plane, we propose to calculate (15) for all the $2^{(P-1)} - 1$ combinations of the bit-planes belonging to the set $\mathcal{X}$. Then, the maximum value of $|L_{ext}^{(p)}|$ is used as the extrinsic information. In the next subsection, we show that this approach outperforms the general approach.

*4.2. Robustness Comparison.* In this subsection, we compare the simulation results for the baseline one-pass and the multipass decoding approach. The simulation setup similar to the one explained in Section 3.2 is used. Figure 11 shows the reconstructed PSNR of the MER1 image. Originally, the rate budget assigned to the image corresponds to the nominal channel SNR of 3 dB. However, as the SNR decreases from its nominal value, the reconstruction quality of the image also decreases in terms of PSNR. The following comments are in order with reference to Figure 11:

(i) (○)-curve shows the performance of one-pass decoding scheme as the channel degrades from its nominal value. The PSNR degrades gradually as the channel SNR decreases similar to the results of [24]. The baseline QLIC encoder is used to generate these results.

(ii) (∗)-curve corresponds to reconstructed PSNR of the image at various values of channel SNR by using the multipass decoding approach. Three iterations of multipass decoder are executed to generate these results. It appears that multipass decoding provides gain over the one-pass decoding approach. Particularly, the gain is significant for low values of mismatched channel SNR. For example, the decoding gain is 1.5 dB when channel degrades by 0.3 dB from its nominal value. Similar to the (○)-curve, the baseline encoder is used to generate these simulations results.

(iii) (◇)-curve corresponds to the typical multipass decoding, i.e., without using the approach proposed in the previous subsection. The multipass decoding provides a maximum gain of 0.6 dB at SNR of 2.8 dB. The gain decreases sharply in case of further channel degradation. It is due to the reason explained earlier that the symbols recovered with low reliability are unable to provide significant extrinsic information in subsequent decoding passes.

(iv) (+)-curve is similar to the (○)-curve. However, the proposed multiple refinement level encoder is used. The simulation results confirm that the robustness of QLIC is retained by using the multiple refinement levels per subsource.

(v) (×)-curve corresponds to the multipass decoding and the proposed multiple refinement encoder. The multipass decoding performance is similar to the (∗)-curve. It is shown that the multirefinement level approach outperforms the multipass decoding approach.

## 5. Conclusion

In this paper, we propose to efficiently remove the redundant bit-planes for spectrally efficient linear index coding of images. Further, the bit-planes are arranged to preserve their significance, and hence the similar robustness performance is achieved even with higher spectral efficiency. Then, multipass decoding is used to iteratively decode the bit-planes. We show that the multipass decoding provides better gain by using extrinsic information from selected bit-planes.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by National Natural Science Foundation of China under Grant 61471022, and NSAF under Grant U1530117.

## References

[1] G. Maral and M. Bousquet, *Satellite Communications Systems: Systems, Techniques and Technology*, John Wiley & Sons, 2011.

[2] J. Taylor, *Deep Space Communications*, Jet Propulsion Laboratory, California Institute of Technology, 2014.

[3] J. L. Massey, "Deep-space communications and coding: A marriage made in heaven," in *Advanced Methods for Satellite and Deep Space Communications*, pp. 1–17, Springer, 1992.

[4] T. De Cola, E. Paolini, G. Liva, and G. P. Calzolari, "Reliability options for data communications in the future deep-space missions," *Proceedings of the IEEE*, vol. 99, no. 11, pp. 2056–2074, 2011.

[5] CCSDS, "TM synchronization and channel coding," Tech. Rep. 131.0-B-3 Blue Book, Consultative Committee for Space Data Systems (CCSDS), 2017.

[6] K. S. Andrews, D. Divsalar, S. Dolinar, J. Hamkins, C. R. Jones, and F. Pollara, "The development of turbo and LDPC codes for deep-space applications," *Proceedings of the IEEE*, vol. 95, no. 11, pp. 2142–2156, 2007.

[7] O. Y. Bursalioglu, M. Fresia, G. Caire, and H. V. Poor, "Lossy joint source-channel coding using raptor codes," *International Journal of Digital Multimedia Broadcasting*, vol. 2008, Article ID 124685, 18 pages, 2008.

[8] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: an overview," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 4, pp. 1103–1127, 2002.

[9] A. Kiely and M. Klimesh, "The ICER progressive wavelet image compressor," *Interplanetary Network Progress Report*, vol. 42, no. 155, pp. 1–46, 2003.

[10] O. Y. Bursalioglu, M. Fresia, G. Caire, and H. V. Poor, "Lossy multicasting over binary symmetric broadcast channels," *IEEE*

*Transactions on Signal Processing*, vol. 59, no. 8, pp. 3915–3929, 2011.

[11] M. Fresia and G. Caire, "A linear encoding approach to index assignment in lossy source-channel coding," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 56, no. 3, pp. 1322–1344, 2010.

[12] Z. Yang, S. Zhao, X. Ma, and B. Bai, "A new joint source-channel coding scheme based on nested lattice codes," *IEEE Communications Letters*, vol. 16, no. 5, pp. 730–733, 2012.

[13] M. Fresia, F. Perez-Cruz, H. Poor, and S. Verdu, "Joint source and channel coding," *IEEE Signal Processing Magazine*, vol. 27, no. 6, pp. 104–113, 2010.

[14] J. Hagenauer, "Source-Controlled Channel Decoding," *IEEE Transactions on Communications*, vol. 43, no. 9, pp. 2449–2457, 1995.

[15] I. Shahid and P. Yahampath, "Distributed joint source-channel coding using unequal error protection LDPC codes," *IEEE Transactions on Communications*, vol. 61, no. 8, pp. 3472–3482, 2013.

[16] J. Li, S. Lin, K. Abdel-Ghaffar, W. E. Ryan, and J. Costello, "Integrated code design for a joint source and channel LDPC coding scheme," in *Proceedings of the IEEE Information Theory and Applications Workshop*, pp. 1–9, San Diego, Calif, USA, 2017.

[17] P. Wang, L. Yin, and J. Lu, "An efficient helicopter-satellite communication scheme based on check-hybrid LDPC coding," *Tsinghua Science and Technology*, vol. 599, pp. 10–26, 2018.

[18] L. Yin and W. Hao, "Code-hopping based transmission scheme for wireless physical-layer security," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 7063758, 12 pages, 2018.

[19] J. Kliewer and N. Gortz, "Iterative source-channel decoding for robust image transmission," in *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing ICASSP-02*, Minneapolis, Minn, USA, April 1993.

[20] L. Xu, L. Wang, S. Hong, and H. Wu, "New results on radiography image transmission with unequal error protection using protograph double LDPC codes," in *Proceedings of the 2014 8th International Symposium on Medical Information and Communication Technology (ISMICT)*, pp. 1–4, Firenze, Italy, April 2014.

[21] J. Kliewer, N. Goertz, and A. Mertins, "Iterative source-channel decoding with Markov random field source models," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3688–3701, 2006.

[22] R. Hamzaoui, V. Stanković, and Z. Xiong, "Optimized error protection of scalable image bit streams," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 91–107, 2005.

[23] A. Gabay, M. Kieffer, and P. Duhamel, "Joint source-channel coding using real BCH codes for robust image transmission," *IEEE Transactions on Image Processing*, vol. 16, no. 6, pp. 1568–1583, 2007.

[24] O. Y. Bursalioglu, G. Caire, and D. Divsalar, "Joint source-channel coding for deep-space image transmission using rateless codes," *IEEE Transactions on Communications*, vol. 61, no. 8, pp. 3448–3461, 2013.

[25] A. Shokrollahi, "Raptor codes," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2551–2567, 2006.

[26] CCSDS, "Low density parity check codes for use in near-earth and deep space applications," Tech. Rep. 131.1-O-2 Orange Book, Consultative Committee for Space Data Systems (CCSDS).

[27] S. P. Protocol, "Recommendation for space data system standards," Tech. Rep. 133.0-B-1. Blue Book, CCSDS, 2003.

[28] R. Mahmood, Q. Huang, and W. Zulin, "A novel decoding method for linear index joint source-channel coding schemes," in *Proceedings of the 13th International Bhurban Conference on Applied Sciences and Technology, IBCAST 2016*, pp. 595–600, Islamabad, Pakista, January 2016.

[29] R. Mahmood, Z. Wang, and Q. Huang, "Multi-pass decoding for the robust transmission of deep-space images," in *Proceedings of the 2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, Sydney, NSW, Australia, June 2017.

[30] "Images, mars exploration program," https://mars.nasa.gov/.

[31] O. Y. Bursalioglu, G. Caire, and D. Divsalar, "Joint source-channel coding for deep space image transmission using rateless codes," in *Proceedings of the 2011 Information Theory and Applications Workshop (ITA)*, pp. 1–10, La Jolla, Calif, USA, Feburary 2011.

[32] O. Etesami and A. Shokrollahi, "Raptor codes on binary memoryless symmetric channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2033–2051, 2006.

[33] A. Venkiah, C. Poulliat, and D. Declercq, "Jointly decoded raptor codes: analysis and design for the biawgn channel," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Article ID 657970, 2009.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 4181626, 10 pages
https://doi.org/10.1155/2018/4181626

*Research Article*

# Superposition Coded Modulation Based Faster-Than-Nyquist Signaling

**Shuangyang Li**,[1,2] **Baoming Bai**,[1] **Jing Zhou**,[3] **Qingli He**,[1] **and Qian Li**[1]

[1]*State Key Laboratory of ISN, Xidian University, Xi'an, China*
[2]*Science and Technology on Communication Networks Laboratory, Shijiazhuang, China*
[3]*Department of EEIS, University of Science and Technology of China, Hefei, China*

Correspondence should be addressed to Baoming Bai; bmbai@mail.xidian.edu.cn

Received 24 November 2017; Accepted 31 March 2018; Published 16 May 2018

Academic Editor: Luca Reggiani



A structure of faster-than-Nyquist (FTN) signaling combined with superposition coded modulation (SCM) is considered. The so-called FTN-SCM structure is able to achieve the constrained capacity of FTN signaling and only requires a low detection complexity. By deriving a new observation model suitable for FTN-SCM, we offer the power allocation based on a proper detection method. Simulation results show that, at any given spectral efficiency, the bit error rate (BER) curve of FTN-SCM lies clearly outside the minimum signal-to-noise ratio (SNR) boundary of orthogonal signaling with a larger alphabet. The achieved data rates are also close to the maximum data rates of the certain shaping pulse.

## 1. Introduction

With the demand and the growth of advanced signal processing capabilities at base stations, the need of efficient backhauling solutions to transmit a large amount of data increases significantly. Thus, as one of the most important parts of deploying the fifth-generation (5G) cellular network, more efficient backhauling techniques need to be applied [2]. Conventionally, the capacity of networks is enlarged by consuming more time/bandwidth/spatial resources. However, this solution may not always be possible, due to the practical reasons. Hence, as an alternative method to gain more capacity, FTN signaling has recently received a lot of attention. An overview of FTN signaling for 5G communication systems was provided in [3].

FTN signaling is an extension of traditional linear modulation and a classical way of nonorthogonal signal transmission, which was first proposed by Mazo in 1975 [4]. Mazo discovered that, with *sinc* pulse as the shaping pulse, the minimum squared Euclidean distance of binary phase shift keying (BPSK) modulated pulses remains the same even when the symbol rate is, to some extent, higher than the Nyquist criterion. His work indicates that there are roughly 25% more bits that could be transmitted in the same bandwidth compared to that of Nyquist signaling, with almost the same error performance over additive white Gaussian noise (AWGN) channels. Recently, Rusek et al. proved that FTN signaling is able to bring more degrees of freedom (DoF) over the AWGN channel [5, 6] compared to orthogonal signaling. As a result, a higher spectral efficiency is expected for FTN signaling and, indeed, fascinating simulation results have already been reported. In [7], a precoded FTN system with quadrature phase shift keying (QPSK) modulation was presented, which, as simulation results imply, requires lower SNR to reach the BER $< 10^{-5}$ compared to that of the constrained capacity of 8-PSK for orthogonal signaling with the same spectral efficiency. However, there is still no such signaling method existing in the literature that is able to outperform orthogonal signaling constrained by a larger alphabet at any preferred spectral efficiency. The reason for this problem lies in the complexity of maximum-likelihood (ML) detection for FTN signals growing exponentially with the size of the alphabet and with the number of taps of intersymbol interference (ISI), respectively. When the system requires high spectral efficiency, conventional FTN signaling systems need either an alphabet with a larger size or a compression factor of a smaller

value to meet the requirements. Consequently, the required ML detection complexity becomes prohibitively high and a suboptimal detection method has to be utilized, which in return somehow damages the performance. Hence, in this paper, we attempt to solve such an issue by considering SCM [8–13].

SCM is a special case of multilevel coding (MLC) [8], which offers an excellent solution to transmissions with severe interference. With the use of the fast Fourier transform- (FFT-) based technique proposed in [9], the detection complexity of SCM system is $O(\log N_{\text{frame}})$, where $N_{\text{frame}}$ is denoted as the frame length [10]. Moreover, with proper Gaussian assumption, the optimization for SCM systems is easier than that of conventional bit interleaved coded modulation (BICM) systems [10]. SCM has also been proven to have promising performance over a variety of channels [11, 12]. More advantages of SCM can be found in [10] and the references therein.

The idea of combining SCM with FTN signaling first appeared in [14], where FTN signals are treated as the sum of several orthogonal signals with different time delays; thus it allows the successive interference cancellation (SIC) detection at the receiver. However, in [1], it has been proven that the aforementioned structure cannot bring any gain in terms of DoF. Hence, a so-called "full-FTN" structure has been proposed in [1] along with its proof of achieving the constrained capacity of FTN signaling. In this structure, the signals are viewed as the sum of several FTN signals of the same compression factor and the SIC is also utilized to reduce the detection complexity. Different from the traditional FTN signaling, to gain a higher spectral efficiency, this structure utilizes more layers rather than a small compression factor. Since, with SIC detection, the detection complexity grows linearly with the number of layers and exponentially with the number of ISI taps, the overall detection complexity of this structure is normally very low. On the other hand, since at each layer, the symbol rate still exceeds the signal bandwidth, the DoF gain of FTN signaling is maintained. However, this structure still lacks a well-designed equalizer to perform SIC, because the common equalizers for FTN signaling, such as the one in [15], require the utilization of the whitening filter in the receiver. This is rather difficult and even impossible when the FTN signal, at each layer, is corrupted by both the colored noise and the signals from other layers. Hence, it is needed to derive a new observation model, which allows the SIC and the detection for each individual layer at the same time. It should also be noted that the combination of FTN signaling and SCM bypasses the obstacle of designing the channel code in terms of different compression factors. Similar to the traditional SCM, an identical code can be utilized for all layers of FTN-SCM, which makes the design and implementation of FTN-SCM system very easy. By simply adjusting the number of superposition layers and the power allocation, FTN-SCM is able to provide a wide range of spectral efficiencies with excellent performance.

The main contributions of this paper are summarized in the following:

(1) We adapt the idea from [1] and provide a more generalized FTN-SCM scheme.

(2) A new channel observation model suitable for FTN-SCM is introduced.

(3) The detection method and the corresponding power allocation for FTN-SCM are also discussed.

(4) Simulation results show that, for BER $< 10^{-5}$, FTN-SCM requires lower SNR than that of the orthogonal signaling with a larger alphabet at any given spectral efficiency.

The rest of this paper is organized as follows. The diagram of FTN-SCM is provided in Section 2. Then, the new channel observation is derived in Section 3. In Section 4, the detection method is described, along with the power allocation derivation. Our numerical results are presented in Section 5, and finally a brief conclusion is provided in Section 6.

## 2. System Model

The transmitter structure of FTN-SCM is illustrated in Figure 1. Assume that the sequence $\mathbf{u}$ carrying information bits is separated into $K$ substreams, namely, $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{K-1}$. Each subsequence of $\mathbf{u}$, say $\mathbf{u}_k$, is then encoded by its corresponding encoder generating the codeword $\mathbf{c}_k$ of length $N$. $\mathbf{c}'_k$ is the permuted version of $\mathbf{c}_k$, which is afterward modulated in the form of BPSK with an average symbol energy $E_k = P_k \tau T$, where $P_k$ is a pregiven power, $\tau$ is the compression factor, and $\tau T$ is the symbol time. $\mathbf{x}_k = \{x_k[1], x_k[2], \ldots, x_k[n], \ldots, x_k[N]\}^{\mathrm{T}}$ represents the modulated symbols at the $k$th layer, which are then superposed directly with the modulated symbols from other layers. The transmitted symbol sequence $\mathbf{x}$ is obtained as the superposition is finished, where the $n$th symbol of $\mathbf{x}$ is given as $x[n] = \sum_{k=0}^{K-1} x_k[n]$. The FTN modulator is able to shape the transmitted signal $s(t)$ for the given input $\mathbf{x}$ based on a certain $T$-orthogonal pulse $h(t)$. Without loss of generality, an FTN-SCM signal can be expressed as

$$s(t) = \sum_{n=1}^{N} x[n]\, h(t - n\tau T) = \sum_{k=0}^{K-1} \sum_{n=1}^{N} x_k[n]\, h(t - n\tau T). \quad (1)$$

A brief diagram of FTN-SCM signal is given in Figure 2, where $K = 2$ and $\tau = 0.5$. As shown in the figure, the pulse of each individual symbol is interfered by the pulses from both the current layer and the other layers. It should be mentioned that, in this case, a symbol rate that is higher than the Nyquist criterion is maintained at each layer. Thus, the capacity gain of FTN signaling is preserved. Note that the different power assignment for each layer is not the only way of performing SIC in the receiver, similar to that of the orthogonal signaling; choosing codes of different rate for each layer may also do the work.

Figure 3 illustrates the receiver structure of FTN-SCM. In this paper, as we only focus on AWGN channels, the received signal $r(t)$ is presented as $r(t) = s(t) + w(t)$, where $w(t)$ has one side power spectral density (PSD) $N_0$. Let $g_n$

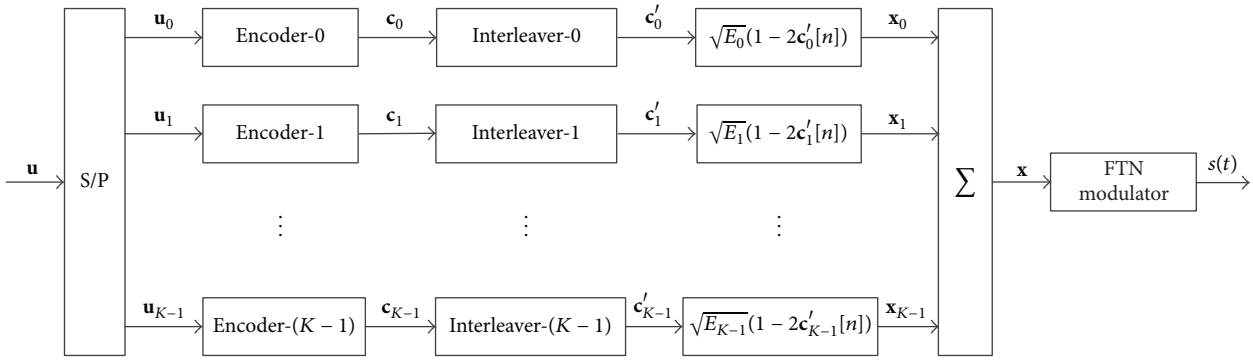



Figure 1: The transmitter structure of FTN-SCM.

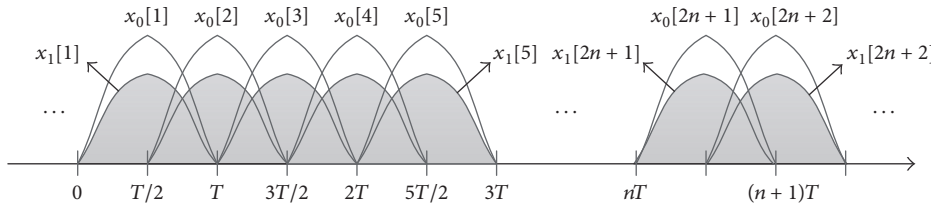



Figure 2: A brief diagram of an FTN-SCM signal with $K = 2$ and $\tau = 0.5$.

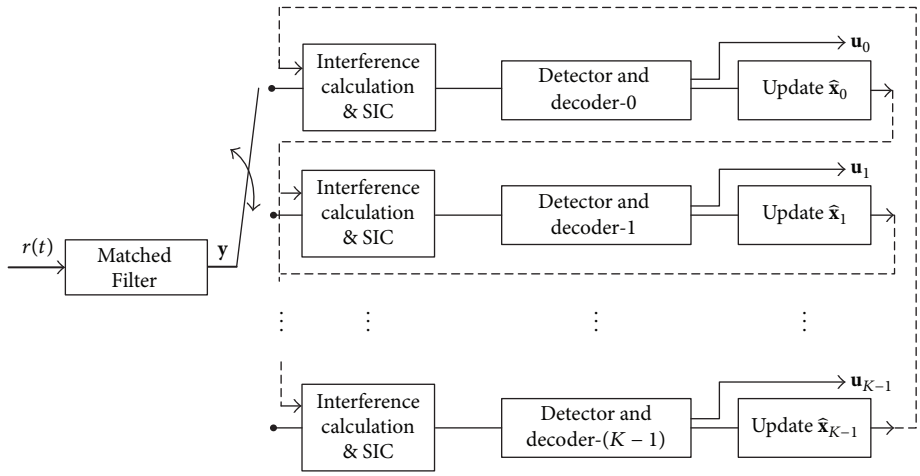



Figure 3: The receiver structure of FTN-SCM.

represent the autocorrelation function samples of $h(t)$. We have

$$g_n = \int_{-\infty}^{\infty} h(t)\, h^*(t - n\tau T)\, \mathrm{d}t, \quad -L_\mathrm{I} \le n \le L_\mathrm{I}, \tag{2}$$

where $(\cdot)^*$ denotes the complex conjugation and $L_\mathrm{I}$ is the length of finite ISI tap. The output of the matched filter is denoted as $\mathbf{y}$. We thus have

$$\mathbf{y} = \mathbf{G}\mathbf{x} + \boldsymbol{\eta}, \tag{3}$$

where $\mathbf{G}$ is a Toeplitz matrix given as

$$\mathbf{G} = \begin{pmatrix} 1 & g_{-1} & \cdots & g_{-L_\mathrm{I}} & 0 & 0 & 0 & 0 & \cdots \\ g_1 & 1 & g_{-1} & \cdots & g_{-L_\mathrm{I}} & 0 & 0 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & & \ddots & & \ddots & \\ g_{L_\mathrm{I}} & \cdots & g_1 & 1 & g_{-1} & \cdots & g_{-L_\mathrm{I}} & 0 & \cdots \\ 0 & g_{L_\mathrm{I}} & \cdots & g_1 & 1 & g_{-1} & \cdots & g_{-L_\mathrm{I}} & \cdots \\ \vdots & & \ddots & & \ddots & \ddots & \ddots & & \ddots \end{pmatrix} \tag{4}$$

and $\boldsymbol{\eta}$ is the colored-noise vector with zero mean and the covariance matrix $\mathrm{E}[\boldsymbol{\eta}\boldsymbol{\eta}^\mathrm{H}] = (N_0/2)\mathbf{G}$. Here, $\mathrm{E}[\cdot]$ is the expectation operator and $(\cdot)^\mathrm{H}$ is the Hermitian (conjugate) transpose.

Without loss of generality, the detection can start at the first layer, and after the detection of each layer, the estimation of current layer inputs, say $\widehat{\mathbf{x}}_k$, is stored for the interference calculation of the following layers. Note that, after one complete sweep over all layers, the updated estimation can be reused to perform other sweeps, which is able to further improve the interference calculation. Normally, three complete sweeps would be enough for FTN-SCM systems.

## 3. Channel Observation Model

We consider the minimum distance detector in this paper. Based on the receiver structure, for detecting the $k$th layer, we have

$$\widehat{\mathbf{x}}_k = \arg\min_{\mathbf{x}_k} \int_{-\infty}^{\infty} \left| r(t) - \sum_{\substack{i=0 \\ i\neq k}}^{K-1} \widehat{s}_i(t) - s_k(t) \right|^2 \mathrm{d}t, \tag{5}$$

where

$$\widehat{s}_i(t) = \sum_{n=1}^{N} \widehat{x}_i[n]\, h(t - n\tau T) \tag{6}$$

represents the estimation of the signal of the $i$th layer. By expanding the equation and dropping the irrelevance, (5) yields

$$\widehat{\mathbf{x}}_k = \arg\max_{\mathbf{x}_k} \int_{-\infty}^{\infty} \left\{ \mathrm{Re}\left\{ \left( r(t) - \sum_{\substack{i=0 \\ i\neq k}}^{K-1} \widehat{s}_i(t) \right) s_k^*(t) \right\} - \frac{1}{2} s_k(t)\, s_k^*(t) \right\} \mathrm{d}t, \tag{7}$$

where $\mathrm{Re}(\cdot)$ represents the real part of a complex number. By switching the integral sequence, (7) can be further derived as

$$\widehat{\mathbf{x}}_k = \arg\max_{\mathbf{x}_k} \mathrm{Re}\left\{ \sum_{n=1}^{N} x_k^*[n] \cdot \int_{-\infty}^{\infty} \left( r(t) - \sum_{\substack{i=0 \\ i\neq k}}^{K-1} \widehat{s}_i(t) \right) h^*(t - n\tau T)\, \mathrm{d}t \right\} - \frac{1}{2} \int_{-\infty}^{\infty} s_k(t)\, s_k^*(t)\, \mathrm{d}t. \tag{8}$$

Thus, with respect to the matched filter outputs, we get

$$\begin{aligned} \widehat{\mathbf{x}}_k = \arg\max_{\mathbf{x}_k} & \sum_{n=1}^{N} \mathrm{Re}\left\{ \left( y[n] - a_k[n] \right) x_k^*[n] \right\} \\ & - \frac{1}{2} \sum_{m=1}^{N} \sum_{n=1}^{N} x_k^*[m]\, x_k[n]\, g_{m-n}, \end{aligned} \tag{9}$$

where

$$a_k[n] = \sum_{\substack{i=0 \\ i\neq k}}^{K-1} \sum_{j=-L_1}^{L_1} \widehat{x}_i[n+j]\, g_j. \tag{10}$$

It is obvious that (9) enables the implementation of the Viterbi algorithm [16]. Similarly, as for soft-in soft-out (SISO) algorithms, for example, the BCJR algorithm [17], (9) implies the recursive probabilistic factorization of the form

$$\begin{aligned} P(\mathbf{y} \mid \mathbf{x}_k, \mathbf{a}_k) \propto \prod_{n=1}^{N} \exp & \left\{ \frac{1}{N_0/2 + \sigma_a^2[n]} \right. \\ & \left. \cdot \mathrm{Re}\left\{ x_k^*[n] \cdot \left( y_n - a_k[n] - \frac{1}{2} g_0 x_k[n] - \sum_{l=1}^{L_1} g_l x_k[n-l] \right) \right\} \right\}, \end{aligned} \tag{11}$$

in which $a_k[n]$ is assumed to be Gaussian, and its variance is denoted as $\sigma_a^2[n]$. Hence, the derivation of the channel observation model is complete.

## 4. Detection Method and Power Allocation

To detect FTN-SCM signal, a well-designed equalizer that accepts nonwhite noise is necessary. Thus, we choose the original method from [18] as the method detecting each layer. The method has been proven to offer promising performance based on the Ungerboeck observation model [19]. As an extension of the traditional $M$-algorithm BCJR ($M$-BCJR) algorithm, the detection method selects the best $M$ states not only based on the current symbol but also considering the influence of some "future" symbols. At each trellis section, for each possible state $S_n$, the method calculates the metrics of the path $\mathbf{v} = x_1^n$ that induces $S_n$ and all possible paths from the section $n+1$ till $n+L$ that are extended from $S_n$. However, new concerns arise due to the presence of other layers. In the following, we aim to offer a performance analysis for the detection of each layer and we further give the power allocation of each layer.

We believe slightly abusing the notation is acceptable. We henceforth use $x_n$ and $a_n$ in place of $x_k[n]$ and $a_k[n]$, respectively, and then the sequence of $\{x_1, x_2, \ldots, x_N\}$ can be represented as $x_1^N$. Without loss of generality, we assume

$x_n$ equiprobably taking values in the alphabet. Hence, in our case, for detecting the $k$th layer, based on the description in [18] and the aforementioned observation model, the metric of path $v_1^{n+L} = x_1^{n+L} + e_1^{n+L}$ with a random error pattern $e_1^{n+L}$ is represented as

$$J\left(v_1^{n+L}\right) = \mathrm{Re}\left\{\left(v_1^{n+L}\right)^{\mathrm{H}}\left(y_1^{n+L} - a_1^{n+L}\right) - \frac{1}{2}\left\|v_1^{n+L}\right\|^2 - \left(v_1^{n+L}\right)^{\mathrm{H}}\mathbf{G}_{\mathrm{L}_{(n+L)\times(n+L)}}\left(v_1^{n+L}\right) + \left(v_1^{n+L}\right)^{\mathrm{H}}\eta_1^{n+L}\right\}, \tag{12}$$

where $\|\cdot\|^2$ is the norm operator and $\mathbf{G}_{\mathrm{L}_{(n+L)\times(n+L)}}$ is the lower triangular matrix with zero main diagonal of the size $(n+L)\times(n+L)$ that is denoted as

$$\mathbf{G}_{\mathrm{L}_{(n+L)\times(n+L)}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ g_1 & 1 & 0 & 0 & 0 & \cdots \\ \vdots & \ddots & 1 & 0 & 0 & \cdots \\ g_{L_{\mathrm{I}}} & \cdots & g_1 & 1 & 0 & \cdots \\ 0 & g_{L_{\mathrm{I}}} & \cdots & g_1 & 1 & \ddots \\ \vdots & & \ddots & & \ddots & \ddots \end{pmatrix}. \tag{13}$$

As we define size A $\triangleq (n+L)\times(n+L+L_{\mathrm{I}})$ and size B $\triangleq (n+L)\times(n+L)$, by substituting (3) into (12), we have

$$J\left(v_1^{n+L}\right) = \mathrm{Re}\left\{\left(x_1^{n+L} + e_1^{n+L}\right)^{\mathrm{H}}\mathbf{G}_{\mathrm{A}}x_1^{n+L+L_{\mathrm{I}}} - \frac{1}{2}\left\|x_1^{n+L} + e_1^{n+L}\right\|^2 - \left(x_1^{n+L} + e_1^{n+L}\right)^{\mathrm{H}}\mathbf{G}_{\mathrm{L}_{\mathrm{B}}}\left(x_1^{n+L} + e_1^{n+L}\right) + \left(x_1^{n+L} + e_1^{n+L}\right)^{\mathrm{H}}\mathbf{G}_{\mathrm{B}}b_1^{n+L} + \left(x_1^{n+L} + e_1^{n+L}\right)^{\mathrm{H}}\eta_1^{n+L}\right\}, \tag{14}$$

in which $b_n$ represents the accuracy of the estimation and is given as

$$b_n = \sum_{\substack{i=0 \\ i\neq k}}^{K-1} x_i[n] - \widehat{x}_i[n], \tag{15}$$

with the variance $\sigma_b^2[n]$.

Furthermore, we consider the difference of the metrics of two individual erroneous paths. Define the two paths as $v_1^{n+L} = x_1^{n+L} + e_1^{n+L}$ and ${v'}_1^{n+L} = x_1^{n+L} + {e'}_1^{n+L}$ and further define $m_1^{n+L} = {e'}_1^{n+L} - e_1^{n+L}$. After several manipulations [18], we obtain

$$J\left({v'}_1^{n+L}\right) - J\left(v_1^{n+L}\right) = \mathrm{Re}\left\{\left(m_1^{n+L}\right)^{\mathrm{H}}\mathbf{T}x_{n+L+1}^{n+L+L_{\mathrm{I}}} - \left(m_1^{n+L}\right)^{\mathrm{H}}\mathbf{G}_{\mathrm{B}}\left(e_1^{n+L} + \frac{1}{2}m_1^{n+L}\right)\right\} + \mathrm{Re}\left\{\left(m_1^{n+L}\right)^{\mathrm{H}}\mathbf{G}_{\mathrm{B}}b_1^{n+L} + \left(m_1^{n+L}\right)^{\mathrm{H}}\eta_1^{n+L}\right\}, \tag{16}$$

where

$$\mathbf{T} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \\ g_{-L_{\mathrm{I}}} & 0 & 0 & \cdots & 0 \\ g_{-(L_{\mathrm{I}}-1)} & g_{-L_{\mathrm{I}}} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ g_{-2} & g_{-3} & \cdots & g_{-L_{\mathrm{I}}} & 0 \\ g_{-1} & g_{-2} & \cdots & g_{-(L_{\mathrm{I}}-1)} & g_{-L_{\mathrm{I}}} \end{pmatrix}. \tag{17}$$

In the following, we focus on the detection performance of each stage. Since the detection at each stage is interfered by the signals from the other stages, it is necessary to make sure that the algorithm is still able to offer a correct detection, for which we offer the following theorem.

**Theorem 1** (correct detection criterion). *In FTN-SCM systems, the kth stage can be successfully detected without the presence of noise, if the maximum variance of $b[n]$, say $\sigma_{max}^2 = \max\{\sigma_b^2[n],\ 1 \le n \le N\}$, satisfies*

$$\sqrt{2\sigma_{max}^2}\sum_{l=-L_{\mathrm{I}}}^{L_{\mathrm{I}}}|g_l| < \frac{1}{2}\sqrt{P_k\tau T}d_{min}^2 - 2\sqrt{P_k\tau T}\sum_{l=1}^{L_{\mathrm{I}}-L} l\left|g_{-(l+L)}\right|, \tag{18}$$

*where $(d_{min}^2)$ represents the minimum squared Euclidean distance of BPSK modulated FTN signals with normalized signal energy, i.e. BPSK modulated FTN signal $s'(t)$ satisfies $\int_{-\infty}^{\infty}|s'(t)|^2 = 1$.*

*Proof.* The proof is given in Appendix A. □

Clearly, Theorem 1 is a sufficient condition for the $k$th stage being successfully detected. Thus, we have proved that the algorithm is able to provide a correct detection, as long as (18) is satisfied. It is straightforward to offer the power allocation based on (18). However, this may not be a good choice for three reasons. Firstly, the derivation for (18) involves the scaling of inequalities; thus the power allocation based on (18) is not the best. Secondly, the algorithm operates on a reduced ISI trellis, where the certain error patterns that have a larger metric may not be included during the detection. Thirdly, practically speaking, Theorem 1 is not the necessary condition for the successful detection. Since error-correcting codes are normally implemented in FTN-SCM systems which helps in the detection in a certain level, thus the power allocation should also take the influence of the corresponding codes into account.

All these three reasons suggest that the power allocation may not necessarily be derived in such a strict way. Hence, we slightly adjust the detection criterion by calculating

Table 1: Power allocation for the simulations in Figure 4.

| $K$ | $P_1/P$ | $P_2/P$ | $P_3/P$ | $P_4/P$ | $P_5/P$ | $P_6/P$ | $P_7/P$ |
|---|---|---|---|---|---|---|---|
| 2 | 0.6714 | 0.3286 | - | - | - | - | - |
| 3 | 0.5774 | 0.2837 | 0.1389 | - | - | - | - |
| 4 | 0.5403 | 0.2655 | 0.1304 | 0.0638 | - | - | - |
| 5 | 0.5237 | 0.2573 | 0.1264 | 0.0621 | 0.0304 | - | - |
| 6 | 0.5159 | 0.2535 | 0.1246 | 0.0612 | 0.0301 | 0.0147 | - |
| 7 | 0.5122 | 0.2517 | 0.1237 | 0.0608 | 0.0299 | 0.0147 | 0.0072 |

the probability of $P(x_n \mid y_1^N, b_1^N, S_{n-1})$ instead of $P(x_n \mid y_1^N, b_1^N)$, where the detection algorithm is assumed with $M = 1$ and $S_{n-1}$ is the correct state that is preserved at the $(n-1)$th section. Thus, at $n$th section, the log likelihood ratio (LLR) of the input $L(x_n)$ can be obtained by the following theorem.

**Theorem 2** (the correct tail path). *We claim that* $\mathbf{v}$ *is the correct tail path (CTP) if and only if the last L elements are correct, which is* $e_{n+1}^{n+L} = \mathbf{0}^{\mathrm{T}}$. *Then at nth section, for the CTPs* $\mathbf{v}$ *and* $\mathbf{v}'$ *of the states* $s_+$ *and* $s_-$ *that are induced by the correct state* $S_{n-1}$, *we have*

$$\begin{aligned} L(x_n) &\triangleq \ln \frac{P\left(x_n = +\sqrt{P_k \tau T} \mid y_1^N, b_1^N, S_{n-1}\right)}{P\left(x_n = -\sqrt{P_k \tau T} \mid y_1^N, b_1^N, S_{n-1}\right)} \\ &= \ln \frac{P\left(s_+ \mid y_1^N, b_1^N, S_{n-1}\right)}{P\left(s_- \mid y_1^N, b_1^N, S_{n-1}\right)} \propto J(\mathbf{v}) - J\left(\mathbf{v}'\right). \end{aligned} \tag{19}$$

*Proof.* The proof is given in Appendix B. □

With Theorem 2, it is possible to evaluate the error event rate (EER) of each layer. We assume the two CTPs are $v_1^{n+L} = x_1^{n+L}$ and $v'^{n+L}_1 = x_1^{n+L} + e_1^{n+L}$ with $e_1^{n+L} = [0, \ldots, 0, e_n, 0, \ldots, 0]^{\mathrm{T}}$, respectively. Thus, for the error event $\varepsilon \triangleq e_1^{n+L}$, we have

$$\begin{aligned} P(\varepsilon) &= P\left(J\left(v'^{n+L}_1\right) - J\left(v_1^{n+L}\right) > 0\right) \\ &= P\left(e_n \left(\sum_{l=L+1}^{L_1} g_{-l} x_{n+l}\right) - |e_n|^2 + e_n \eta_n \right. \\ &\quad \left. + e_n \left(\sum_{l=-L_1}^{L} g_{-l} b_{n+l}\right) > 0\right). \end{aligned} \tag{20}$$

By considering the Gaussian assumption, and the steep decrease of $Q$ function, (20) can be simplified as

$$P(\varepsilon) \simeq Q\left(\sqrt{\frac{P_k \tau T}{\sigma^2}}\right), \tag{21}$$

where

$$\sigma^2 = \frac{N_0}{2} + \sum_{l=L+1}^{L_1} |g_{-l}|^2 P_k \tau T + \sum_{l=-L_1}^{L} |g_{-l}|^2 \sigma_b^2[n]. \tag{22}$$

In the following, we offer a power allocation with respect to individual error-correcting codes. Without loss of generality, we assume that the code at the $k$th layer successfully recovers the information sequence at SNR $= \rho_k$ over the FTN channel. Meanwhile, according to (21), the signal-to-interference plus noise ratio (SINR) for the $k$th layer is defined as

$$\begin{aligned} \mathrm{SINR}_k &\triangleq \frac{E\left(x_k^2[n]\right)}{E\left(\sigma_b^2[n]\right) + N_0/2} \\ &= \frac{P_k \tau T}{N_0/2 + \sum_{l=L+1}^{L_1} |g_{-l}|^2 P_k \tau T + \sum_{l=-L_1}^{L} |g_{-l}|^2 \left(\sum_{i=k+1}^{K-1} P_i \tau T\right)}. \end{aligned} \tag{23}$$

Therefore, to successfully decode $u_k$, $\mathrm{SINR}_k \geq \rho_k$ must hold. Thus, we have

$$P_k \geq \frac{\rho_k \left(N_0/2 + \sum_{l=-L_1}^{L} |g_{-l}|^2 \left(\sum_{i=k+1}^{K-1} P_i \tau T\right)\right)}{\tau T \left(1 - \rho_k \sum_{l=L+1}^{L_1} |g_{-l}|^2\right)}. \tag{24}$$

Hence, by noticing the natural power assignment constraint that $P_0 + P_1 + \cdots + P_{K-1} = 1$, the required $P_k$ for all $k$ can be obtained recursively starting from $k = K - 1$. The numerical results based on the above power allocation are demonstrated in the next section.

## 5. Numerical Results

We choose the root raised cosine (RRC) (with roll-off factor $\beta = 0.3$ and a time-truncation to $\pm 15T$ around $t = 0$) (without loss of generality, we assume $T = 1$) as the shaping pulse $h(t)$. Meanwhile, the outer code is chosen as the code rate $R = 1/3$ asymmetric Turbo code in [18], with the generator polynomial $g_1(D) \triangleq [1 \quad (1 + D + D^2)/(1 + D^2)]$ and $g_2(D) \triangleq [1 \quad (1 + D + D^3)/(1 + D^2 + D^3)]$. As we transmit 20000 information bits per layer, we have $N = 60010$ as the codeword length, wherein 10 redundant bits are included to terminate the trellis. The two-dimensional normalized spectral efficiency is defined as $\eta = 2RK/\tau(1 + \beta)$.

The simulation results of FTN-SCM with $K = 2$ to 7 and $\tau = 2/3$ are plotted in Figure 4, wherein the power allocation is shown in Table 1. The parameters for the detection method are chosen as $M = 4$ and $L = 2$. There are 50 Turbo iterations between the FTN and Turbo code at each layer and 3 complete sweeps in total. Figure 5 shows the corresponding achieved data rate (for details on the data rate, please refer to [5]) at BER $< 10^{-5}$. As figures imply, the BER results of FTN-SCM

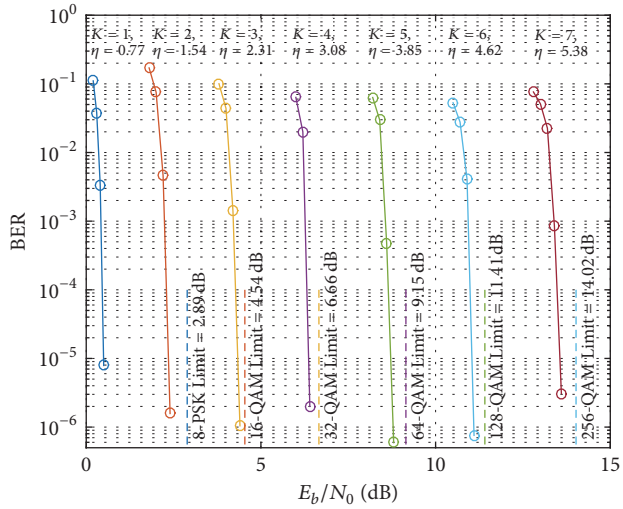



FIGURE 4: BER results of FTN-SCM with $K$ layers and $\tau = 2/3$ compared to constrained capacities of orthogonal signaling, where $\eta$ (bits/s/Hz) represents the two-dimensional normalized spectral efficiency.

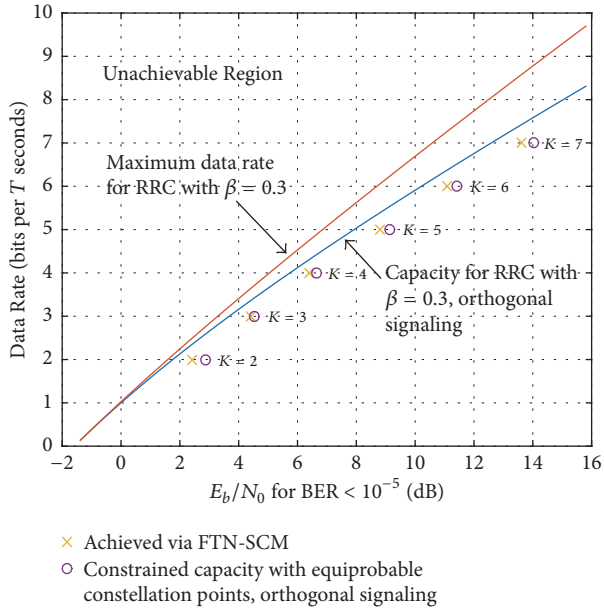



FIGURE 5: Achieved data rate (BER < $10^{-5}$) compared to both the constrained and unconstrained capacity of orthogonal signaling as well as the ultimate capacity of RRC with $\beta = 0.3$.

lie clearly outside the constrained capacity boundary with a larger alphabet of orthogonal signaling. We also observe that FTN-SCM is able to achieve a wide range of spectral efficiencies with a simple binary modulation format at each layer. In order to better illustrate the performance of FTN-SCM, we make a comparison between our method and the method in [1]. The BER results of two methods are given in Figure 6, where the simulation parameters are the same as the case of $K = 2$ in Figure 4, except the power allocation for method in [1] is given as $P_1/P = 0.6705$ and $P_2/P = 0.3295$. Note that, in [1], the authors utilize an *optimal* FTN equalizer in order to gain a very good performance. However, such equalizers are normally impractical, especially when there are a lot of layers; for example, $K = 7$. On the other hand, for BER < $10^{-5}$, our method only needs no more than 0.1 dB to achieve the same performance as the *optimal* result, but with much less complexity, which proves that our method exhibits a better trade-off between performance and complexity than the method in [1]. For a detailed complexity comparison, please refer to [18]. It should be mentioned that the BER results can be improved by choosing a better outer code or a better detection method, which is a future topic for us.

## 6. Conclusion

In this paper, we considered the FTN-SCM structure. Based on the transceiver structure, we derived a new observation model and further offered the power allocation with respect

to the detection method of each layer. Simulation results show that, in a wide range of spectral efficiencies, FTN-SCM requires lower SNR than that for orthogonal signaling with a larger alphabet. It should be noted that the proposed scheme is easy to be extended to nonbinary modulation cases and other types of channels.

## Appendix

## A. Proof of Theorem 1

According to [18], the probabilities of the correct state $S_n = s$ and the wrong state $S_n = s'$ with the error sequence $e_1^n$ satisfy the following equation:

$$\ln \frac{P\left(S_n = s', y_1^N, b_1^N\right)}{P\left(S_n = s, y_1^N, b_1^N\right)} \propto \sum_{i=1}^{2^L} J\left(\mathbf{v}'_i\right) - J\left(\mathbf{v}_i\right), \tag{A.1}$$

where $\mathbf{v}'_i$ and $\mathbf{v}_i$ represent the $i$th path of the probability calculation for $S_n = s'$ and $S_n = s$, respectively. Without loss of generality, we assume that $\mathbf{v}'_i$ and $\mathbf{v}_i$ have the same error pattern $e_{n+1}^{n+L}$. Thus, by considering (16), (A.1) can be further simplified as

$$\begin{aligned} \ln \frac{P\left(S_n = s', y_1^N, b_1^N\right)}{P\left(S_n = s, y_1^N, b_1^N\right)} &\propto 2^L \\ &\times \left\{ \mathrm{Re}\left\{ \left(e_1^n\right)^{\mathrm{H}} \mathbf{T}' x_{n+L+1}^{n+L_{\mathrm{I}}} - \frac{1}{2} d^2\left(e_1^n\right) \right\} \right. \\ &\left. + \mathrm{Re}\left\{ \left(e_1^n\right)^{\mathrm{H}} \mathbf{G}' b_1^{n+L} + \left(e_1^n\right)^{\mathrm{H}} \eta_1^n \right\} \right\}, \end{aligned} \tag{A.2}$$

where

$$\mathbf{T}' = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & & & \\ 0 & & & \vdots \\ g_{-L_{\mathrm{I}}} & 0 & & \\ g_{-(L_{\mathrm{I}}-1)} & g_{-L_{\mathrm{I}}} & \ddots & \\ \vdots & & \ddots & 0 \\ g_{-(L+1)} & g_{-(L+2)} & \cdots & g_{-L_{\mathrm{I}}} \end{pmatrix},$$

$$\mathbf{G}' = \begin{pmatrix} 1 & g_{-1} & \cdots & g_{-L} & g_{-L-1} & \cdots & g_{-L_{\mathrm{I}}} & 0 & \cdots & 0 \\ g_1 & 1 & g_{-1} & \cdots & g_{-L} & g_{-L-1} & \cdots & g_{-L_{\mathrm{I}}} & 0 & \cdots \\ g_2 & g_1 & 1 & g_{-1} & \cdots & g_{-L} & g_{-L-1} & \cdots & g_{-L_{\mathrm{I}}} & 0 \\ & & \ddots & \ddots & \ddots & & \ddots & \ddots & & \\ \ddots & & & \ddots & \ddots & \ddots & & \ddots & \ddots & \\ \cdots & 0 & g_{L_{\mathrm{I}}} & \cdots & g_1 & 1 & g_{-1} & \cdots & g_{-L} & g_{-L-1} \\ 0 & \cdots & 0 & g_{L_{\mathrm{I}}} & \cdots & g_1 & 1 & g_{-1} & \cdots & g_{-L} \end{pmatrix}, \tag{A.3}$$

and $d^2(e_1^n) = (e_1^n)^{\mathrm{H}} \mathbf{G}_{n\times n} e_1^n$, representing the squared Euclidean distance between the erroneous path and the correct path at the current stage. As $b[n]$ is assumed to be Gaussian with variance $\sigma_b^2[n]$, the right-hand side of (A.2) can be upper-bounded in the noiseless regime by

$$\begin{aligned} 2^L &\times \left\{ \mathrm{Re}\left\{ \left(e_1^n\right)^{\mathrm{H}} \mathbf{T}' x_{n+L+1}^{n+L_{\mathrm{I}}} - \frac{1}{2} d^2\left(e_1^n\right) \right\} \right. \\ &\left. + \mathrm{Re}\left\{ \left(e_1^n\right)^{\mathrm{H}} \mathbf{G}' b_1^{n+L} \right\} \right\} < 2^L \\ &\times \left\{ \mathrm{Re}\left\{ \max_{e_1^n, x_{n+L+1}^{n+L_{\mathrm{I}}}} \left\{ \left(e_1^n\right)^{\mathrm{H}} \mathbf{T}' x_{n+L+1}^{n+L_{\mathrm{I}}} \right\} \right\} \right. \\ &\quad - \frac{1}{2} \mathrm{Re}\left\{ \min_{e_1^n} \left\{ d^2\left(e_1^n\right) \right\} \right\} \\ &\left. \quad + \mathrm{Re}\left\{ \max_{e_1^n} \left\{ \left(e_1^n\right)^{\mathrm{H}} \mathbf{G}' b_1^{n+L} \right\} \right\} \right\}. \end{aligned} \tag{A.4}$$

By noticing the fact that $e_n \in \{-\sqrt{2P_k\tau T}, 0, 2\sqrt{P_k\tau T}\}$, the right-hand side of (A.4) can further be extended as

$$\begin{aligned} 2^L &\times \left\{ \mathrm{Re}\left\{ \max_{e_1^n, x_{n+L+1}^{n+L_{\mathrm{I}}}} \left\{ \left(e_1^n\right)^{\mathrm{H}} \mathbf{T}' x_{n+L+1}^{n+L_{\mathrm{I}}} \right\} \right\} \right. \\ &\quad \left. - \frac{1}{2} \mathrm{Re}\left\{ \min_{e_1^n} \left\{ d^2\left(e_1^n\right) \right\} \right\} \right. \end{aligned}$$

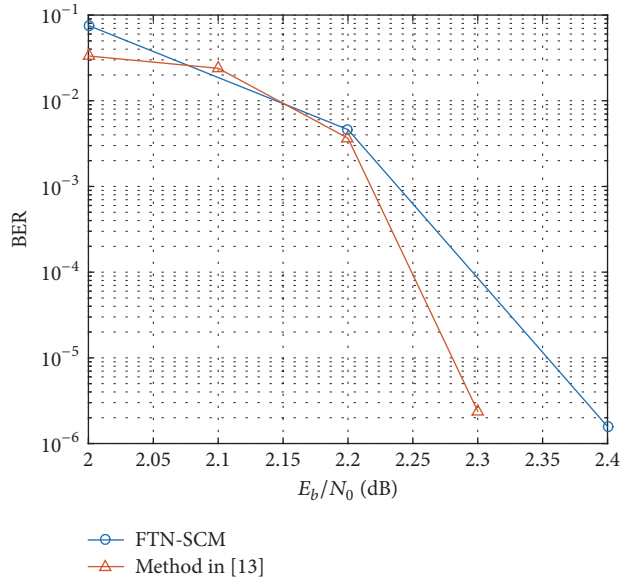



FIGURE 6: BER result of FTN-SCM with $K = 2$ compared to the result of the method in [1].

$$+ \operatorname{Re}\left\{\max_{e_1^n}\left\{\left(e_1^n\right)^{\mathrm{H}}\mathbf{G}'b_1^{n+L}\right\}\right\} < 2^L \times \left\{P_k\tau T\left(2\sum_{l=1}^{L_1-L} l\left|g_{-(L+l)}\right| - \frac{1}{2}d_{\min}^2\right) + \sqrt{2\sigma_{\max}^2 P_k\tau T}\sum_{l=-L_1}^{L_1}\left|g_l\right|\right\}. \tag{A.5}$$

This completes the proof of Theorem 1.

## B. Proof of Theorem 2

Clearly, since $S_{n-1}$ is the correct state at section $(n-1)$ and the states in the trellis are Markovian, the LLR of the input $x_n$ is determined by the probabilities of the states $s_+$ and $s_-$. According to the description in [18], in our case, the probability of $S_n = s$ follows

$$P\left(S_n = s \mid y_1^N, b_1^N, S_{n-1}\right) \propto \sum_{i=1}^{2^L}\exp\left[J\left(\mathbf{v}_i\right)\right], \tag{B.1}$$

where $\mathbf{v}_i = x_1^{n+L} + e_1^{n+L}$, with $e_1^{n+L} = [0,\ldots,0,e_n,e_{n+1},\ldots,e_{n+L}]^{\mathrm{T}}$, representing the $i$th possible path that is extended from $S_n = s$. The calculation implies a process of generating the marginal probability from all joint probabilities, as $2^L$ combinations are all taken into account.

For derivation brevity, we use $\mathbf{v}$ and $\mathbf{v}'$ representing the paths extended from $s_+$ and $s_-$, respectively. We further require that the same subscript $i$ represents the same error pattern. Thus, it is fair to assume that $\mathbf{v}_k$ and $\mathbf{v}'_k$ are the CTPs from states $s_+$ and $s_-$, respectively. Hence, we obtain

$$\frac{P\left(s_+ \mid y_1^N, b_1^N, S_{n-1}\right)}{P\left(s_- \mid y_1^N, b_1^N, S_{n-1}\right)} \propto \exp\left[J\left(\mathbf{v}_k\right) - J\left(\mathbf{v}'_k\right)\right] \times \frac{\exp\left[J\left(\mathbf{v}_1\right) - J\left(\mathbf{v}_k\right)\right] + \exp\left[J\left(\mathbf{v}_2\right) - J\left(\mathbf{v}_k\right)\right]\cdots}{\exp\left[J\left(\mathbf{v}'_1\right) - J\left(\mathbf{v}'_k\right)\right] + \exp\left[J\left(\mathbf{v}'_2\right) - J\left(\mathbf{v}'_k\right)\right]\cdots}. \tag{B.2}$$

Without loss of generality, we consider $J(\mathbf{v}_i) - J(\mathbf{v}_k)$, where $\mathbf{v}_i$ is not the CTP. Recall (16); we have

$$J\left(\mathbf{v}_i\right) - J\left(\mathbf{v}_k\right) = \operatorname{Re}\left\{\left(m_{n+1}^{n+L}\right)^{\mathrm{H}}\mathbf{T}''x_{n+L+1}^{n+L+L_1} - \left(m_{n+1}^{n+L}\right)^{\mathrm{H}}\mathbf{G}_{\mathrm{B}}\left(\frac{1}{2}m_{n+1}^{n+L}\right)\right\} + \operatorname{Re}\left\{\left(m_{n+1}^{n+L}\right)^{\mathrm{H}}\mathbf{G}_{L\times L}b_{n+1}^{n+L} + \left(m_{n+1}^{n+L}\right)^{\mathrm{H}}\eta_{n+1}^{n+L}\right\}, \tag{B.3}$$

where

$$\mathbf{T}'' = \begin{pmatrix} g_{-L} & g_{-(L+1)} & \cdots & g_{-L_1} & 0 & \cdots & 0 \\ \vdots & \ddots & & & \ddots & & \vdots \\ g_{-2} & \cdots & g_{-L} & g_{-(L+1)} & \cdots & g_{-L_1} & 0 \\ g_{-1} & g_{-2} & \cdots & g_{-L} & g_{-(L+1)} & \cdots & g_{-L_1} \end{pmatrix}. \tag{B.4}$$

Since $e_1^{n+L}$ is no longer part of the equation, the only variable in the equation is $m_{n+1}^{n+L}$. Thus, as all the combinations of $m_{n+1}^{n+L}$ are included in (B.2), we can safely draw the conclusion that

$$\frac{\exp\left[J\left(\mathbf{v}_1\right) - J\left(\mathbf{v}_k\right)\right] + \exp\left[J\left(\mathbf{v}_2\right) - J\left(\mathbf{v}_k\right)\right]\cdots}{\exp\left[J\left(\mathbf{v}'_1\right) - J\left(\mathbf{v}'_k\right)\right] + \exp\left[J\left(\mathbf{v}'_2\right) - J\left(\mathbf{v}'_k\right)\right]\cdots} = 1. \tag{B.5}$$

This completes the proof of Theorem 2.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 61771364, 91438101, and 61601346, by the Science and Technology on Communication Networks Laboratory under Grant 614210403070717, by the Fundamental Research Funds for the Central Universities under Grant WK2100060020, by the Key Research Program of Frontier Sciences of CAS under Grant QYZDY-SSW-JSC003, by the China Postdoctoral Science Foundation under Grant 2015M580819, and by the Shaanxi Province Postdoctoral Science Foundation.

## References

[1] Y. J. D. Kim, J. Bajcsy, and D. Vargas, "Faster-Than-Nyquist Broadcasting in Gaussian Channels: Achievable Rate Regions and Coding," *IEEE Transactions on Communications*, vol. 64, no. 3, pp. 1016–1030, 2016.

[2] CISCO, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015-2020, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.pdf, 2016.

[3] J. B. Anderson, "Faster-than-Nyquist signaling for 5G communication," in *Signal Processing for 5G: Algorithms and Implementations*, pp. 24–46, Wiley-IEEE Press, 2016.

[4] J. E. Mazo, "Faster-than-Nyquist signaling," *Bell Labs Technical Journal*, vol. 54, no. 8, pp. 1451–1462, 1975.

[5] F. Rusek and J. B. Anderson, "Constrained capacities for faster-than-Nyquist signaling," *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 764–775, 2009.

[6] J. B. Anderson, F. Rusek, and V. Öwall, "Faster-than-Nyquist signaling," *Proceedings of the IEEE*, vol. 101, no. 8, pp. 1817–1830, 2013.

[7] F. Rusek and J. B. Anderson, "Serial and parallel concatenations based on faster than nyquist signaling," in *Proceedings of the IEEE International Symposium on Information Theory*, pp. 1993–1997, Seattle, WA, USA, 2006.

[8] T. M. Cover, "Broadcast Channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 2–14, 1972.

[9] X. Yuan, Q. Guo, and L. Ping, "Low-complexity iterative detection in multi-user MIMO ISI channels," *IEEE Signal Processing Letters*, vol. 15, pp. 25–28, 2008.

[10] L. Ping, J. Tong, X. Yuan, and Q. Guo, "Superposition coded modulation and iterative linear MMSE detection," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 6, pp. 995–1004, 2009.

[11] K. Wu and L. Ping, "A quasi-random approach to space-time codes," *IEEE Transactions on Information Theory*, vol. 54, no. 3, pp. 1073–1085, 2008.

[12] X. Chen, X. Wang, and X. Ma, "Superposition convolutional coded modulation for the Rayleigh fading channel," in *Proceedings of the International Conference on Communications, Circuits and Systems, ICCCAS 2007*, pp. 37–41, IEEE, Kokura, Japan, 2007.

[13] X. Ma and L. Ping, "Coded modulation using superimposed binary codes," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3331–3343, 2004.

[14] Y. J. Kim and J. Bajcsy, "Faster than Nyquist broadcast signaling," in *Proceedings of the 2012 26th Biennial Symposium on Communications (QBSC)*, pp. 186–189, Kingston, Canada, May 2012.

[15] G. Colavolpe, G. Ferrari, and R. Raheli, "Reduced-state BCJR-type algorithms," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 5, pp. 848–859, 2001.

[16] G. D. Forney, "The Viterbi Algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.

[17] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 284–287, 1974.

[18] S. Li, B. Bai, J. Zhou, P. Chen, and Z. Yu, "Reduced-Complexity Equalization for Faster-than-Nyquist Signaling: New Methods Based on Ungerboeck Observation Model," *IEEE Transactions on Communications*, vol. 66, no. 3, pp. 1190–1204, 2018.

[19] G. Ungerboeck, "Adaptive Maximum-Likelihood Receiver for Carrier-Modulated Data-Transmission Systems," *IEEE Transactions on Communications*, vol. 22, no. 5, pp. 624–636, 1974.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 5957320, 7 pages
https://doi.org/10.1155/2018/5957320

*Research Article*

# A Novel Design of Downlink Control Information Encoding and Decoding Based on Polar Codes

**Ce Sun**,[1] **Zesong Fei**,[1] **Jiqing Ni**,[2] **Wei Zhou**,[2] **and Dai Jia**[1]

[1]*Beijing Institute of Technology, Beijing 100081, China*
[2]*China Mobile Research Institute, Beijing 100081, China*

Correspondence should be addressed to Zesong Fei; feizesong@bit.edu.cn

Received 25 November 2017; Revised 1 March 2018; Accepted 1 April 2018; Published 13 May 2018

Academic Editor: Luca Reggiani



In legacy long term evolution (LTE) networks, multiple transmission modes are defined to cater to diverse wireless environment and improve the spectrum utilization. However, constrained by user equipment (UE) processing capability on blind detection of downlink control information (DCI), two transmission modes are allowed to be configured to UE simultaneously. In recent 5G standardization, the polar codes have supplanted the tail biting convolution codes (TBCC), becoming the channel coding scheme for downlink control information (DCI). Motivated by its successive decoding property, a novel design of DCI encoding and decoding is proposed in this paper. The proposed scheme could support dynamic configuration of transmission modes with decreasing the complexity of blind detection. Evaluation results from link level simulations show that the performance loss compared to conventional encoding/decoding scheme is generally negligible and the proposed scheme can comply with the false alarm rate (FAR) target of 5G standardization.

## 1. Introduction

In the long term evolution (LTE) system, transmitter, and receiver communicate with each other by different transmission scheme. Each transmission scheme is corresponding to a transmission mode (TM) [1]. In the LTE, the multiple TMs are defined to cater to diverse wireless environment and improve the spectrum utilization. The LTE system supports nine TMs; the difference among those is the special structure of the antenna mapping, the reference signal of demodulation, and the feedback type [2].

The downlink control information (DCI) transited by base station to users can be used to schedule the downlink/uplink data transmission and convey essential configurations [3]. Specifically, the different DCI formats correspond to the different transmission modes. The length of DCI format will be adjusted with the different system configuration. Constrained by UE processing capability of blind detection of DCI [4], two transmission modes are allowed to be configured to a UE simultaneously.

The polar codes have been adopted as the channel coding scheme of the control channel in the next-generation communication networks [5]. It is based on the polarization theory and can achieve the capacity of arbitrary binary-input discrete memoryless channel (B-DMC) [6]. Furthermore, the complexity of polar codes is low with some optimized decoding algorithms, such as low-complexity list successive cancellation (LCLSC) decoding algorithm [7]. Blind detection of polar codes has been researched in [8]; that work focuses on fitting within the 5G parameters. A low-complexity blind-detection algorithm for polar-encoded frames is proposed in [9]. That scheme decreased the complexity of polar decoding in blind detection. Our scheme decreased the complexity of the process of blind detection.

Being different from the tail biting convolution code (TBCC) which is the coding scheme in LTE, all the existing decoders of polar codes are based on successive cancellation (SC) decoder [10], which allows the encoded bits to be decoded in given order. Taking advantage of successive property of polar decoders, decoding process can be paused after the first several bits being decoded and continued accordingly based on the value of first several bits. Based on the successive property of polar decoders, a novel design on DCI encoding and decoding is proposed in this paper. The proposed scheme

could support dynamic configuration of transmission modes with decreasing the complexity of blind detection.

The rest of this paper is organized as follows. In Section 2 we introduce the foundation of proposed scheme. The scheme of DCI design is proposed in Section 3. In Section 4, we analyze the complexity, and simulation results are given. Finally, Section 5 concludes the paper.

## 2. Preliminary

In this section, we introduce polar codes and DCI design of the LTE system briefly; these are the foundation of the proposed scheme.

*2.1. Polar Codes.* Polar codes are based on channel polarization theory which is described as follows.

**Theorem 1.** *For any B-DMC $W$, the channels $W_N^{(i)}$ polarize in the sense that, for any fixed $\delta \in (0,1)$, as $N$ goes to infinity through powers of two, the fraction of indices $i \in \{1,\ldots,N\}$ for which $I(W_N^{(i)}) \in (1-\delta,1]$ goes to $I(W)$ and the fraction for which $I(W_N^{(i)}) \in [0,\delta)$ goes to $1-I(W)$, where $N$ is the length of code word which is equal to the length of polarized subchannels, $W_N^{(i)}$ denotes the $i$th subchannel of $N$ subchannels, and $I$ denotes the channel capacity.*

According to Theorem 1, we set the information bits in the subchannel set in which $I(W_N^{(i)}) \in (1-\delta,1]$ and set the frozen bits in the other subchannels to construct the information block $u$. Before setting the information bits and frozen bits, we should calculate the reliability of $N$ subchannels and decide which subchannels are good to be set as information bits. The common algorithms to calculate the reliability include algorithm based on Bhattacharyya parameters [11], density evolution (DE) [12], and Gaussian approximation (GA) [13]. And then send $u$ into polar encoder to be encoded. The polar encoding is denoted as $x_1^N = u_1^N \mathbf{G}_N$, where $x_1^N = x_1, x_2, \ldots, x_N$ is the code word, $u_1^N = u_1, u_2, \ldots, u_N$ is the information block, and $\mathbf{G}_N$ is the generator matrix of order $N$. The recursive definition of $G_N$ is given by

$$G_N = B_N F_2^{\otimes n}, \quad F_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \tag{1}$$

where $B_N$ is a permutation matrix.

*2.2. Successive Cancellation Decoder.* All the existing decoders of polar codes are based on successive cancellation (SC) decoder. After receiving $y_1^N$, the SC decoder generates its decision $u_1^N$ by computing

$$\widehat{u}_i \triangleq \begin{cases} 0 & \text{if } i \in \mathscr{A}^c \\ h_i\left(y_1^N, \widehat{u}_1^{i-1}\right) & \text{if } i \in \mathscr{A}, \end{cases} \tag{2}$$

where

$$h_i\left(y_1^N, \widehat{u}_1^{i-1}\right) \triangleq \begin{cases} 0, & \text{if } \dfrac{W\left(0 \mid \widehat{u}_1^{i-1}, y_1^N\right)}{W\left(1 \mid \widehat{u}_1^{i-1}, y_1^N\right)} \geq 1 \\ 1, & \text{otherwise} \end{cases} \tag{3}$$

and $y_1^N = \mathbf{h}x_1^N + \mathbf{n}$ denotes the received message, $\mathbf{h}$ is the channel matrix, and $\mathbf{n}$ denotes the noise of channel.

Through the above formula, we can see that the polar decoder decodes the information bit by bit from $u_1$ to $u_N$. When we need to decode the $i$th bit $u_i$, it is decided as zero if $u_i$ is frozen bit; otherwise $u_i$ is decided by (2) with the prior information of $u_1, u_2, \ldots, u_{i-1}$ and $y_1, y_2, \ldots, y_N$. This property of polar decoder is defined as successive property which makes it is possible to suspend the process of decoding when $u_1, u_2, \ldots, u_{i-1}$ have been decoded.

*2.3. DCI Design.* According to the latest MIMO-related progress in the 3rd-generation partnership project (3GPP), only one code word (CW) is transmitted for 1 to 4 layers and two CWs are transmitted for 5 to 8 layers. Thus, the actual number of transmission layers could implicitly indicate the number of CWs. Moreover, compared to 1-CW case, 2 CWs would add an additional block of bit fields to DCI, possibly containing MCS/RV/NDI and CBGTI/CBGFI if CBG-based transmission is configured, as shown in Figure 1. This is where the difference between DCI payload sizes mainly rises. Consequently, DCI formats with 1- to 4-layer transmission could strive to have the same DCI payload size and so are the DCI formats with 5- to 8-layer transmission. Different transport layers corresponded to different DCI formats. UE does not know which DCI format of information is selected by base station; therefore blind detection is needed. Constrained by UE processing capability on blind detection of DCI, two DCI formats are allowed to be configured to UE simultaneously. UE attempted to decode the information with one DCI format, if it can not perform decoding correctly, UE will attempt to decode the information with the other DCI format.

## 3. Proposed Scheme

Based on the above discussion, a potential design is to add an explicit rank indicator (RI) field in DCI and utilize this field to implicitly indicate DCI payload size during decoding process. Specifically, the RI with fixed length could be decoded firstly, and then the number of CWs and possibly DCI payload size (this depends on the detailed DCI content) could be implicitly identified.

This design is feasible from technical perspective as polar codes have been adopted for PDCCH [1]. Based on the successive property of polar decoder, the number of RI can be decoded first before decoding of the CWs. Motivated by this property, the number of CWs could be informed dynamically by RI. If the number of transmission layer exceeds a threshold (e.g., 4-layer transmission based on current agreements), the decoder would continue the decoding process based on a long bit length, otherwise based on a short bit length. It is evident that such DCI format and decoding design are feasible with decreasing the blind decoding complexity at UE side.

In this section, the proposed scheme is described in detail. The encoding and decoding of polar codes are adjusted as the proposed scheme.

*3.1. Encoding.* Since each scheduled transmission may contain 1 or 2 CWs, we define two DCI formats with different

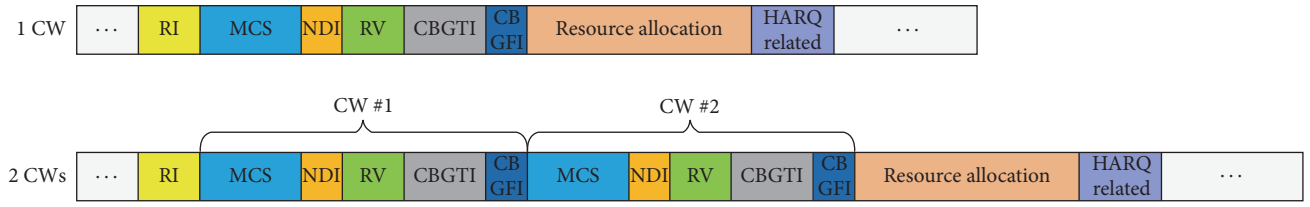
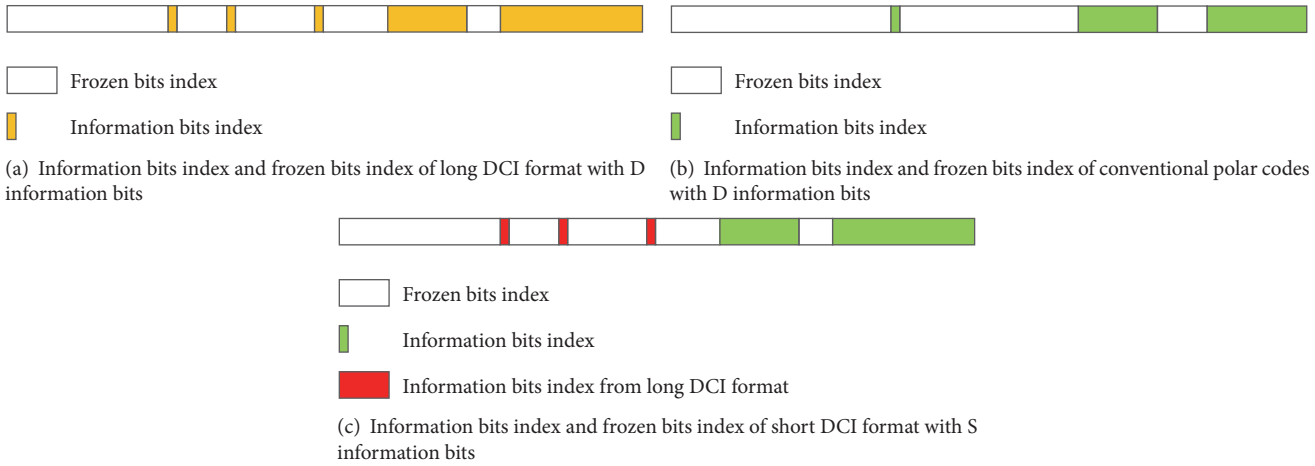



Figure 1: An illustrative DCI format for 1 CW and 2 CWs.




(a) Information bits index and frozen bits index of long DCI format with D information bits

(b) Information bits index and frozen bits index of conventional polar codes with D information bits

(c) Information bits index and frozen bits index of short DCI format with S information bits

Figure 2: The information bits index and the frozen bites index.

lengths. Assume that the long DCI format is $D$ bits and the short DCI format is $S$ bits. Both the DCI formats are encoded to an $N$-bit code words for PDCCH. The RI field in DCI is utilized to implicitly indicate the DCI payload size. The base station sets a threshold $T$ for RI. If RI exceeds the threshold, the base station shall transmit the short DCI format or, otherwise, transmit the long DCI format. The encoding procedure for two DCI formats is illustrated as follows, respectively.

*Long DCI Format.* The base station knows which DCI format should be transmitted and encode the DCI format for different length. Step 1: the reliability of $N$-bit polarized subchannels is calculated. Transmitter selects the $(D + I)$ most reliable subchannels which are noted as $A_{(D+I)} = [A_1, A_2, \ldots, A_{(D+I)}]$. Step 2: the transmitter maps the $(D+I)$-bit information on the selected $A_{(D+I)}$; the first $I$ bits of $(D + I)$-bit information are the RI field and the remaining $D$ bits are the control information. Step 3: the transmitter maps the frozen bits (usually all being zero sequence) on the other $N - (D + I)$ subchannels to construct an $N$-bit sequence $u$. Last, input the sequence $u$ into the polar encoder.

*Shot DCI Format.* Step 1 for short DCI format is the same as that for long DCI format. In step 2, instead of selecting the most reliable $(S+I)$ subchannels from the set, the transmitter selects the first $I$ elements from the set $A_{(D+I)}$ as RI field which is the same as the long DCI format. Then select the most reliable $S$ subchannels from the remaining sets as the information set $A_S$. Then the information bits of short DCI format are mapped on selected $S$ subchannels and the frozen bits are mapped on the other $N - (S + I)$ subchannels.

The construction of sequence $u$ is shown as in Figure 2. Figure 2(a) denotes the sequence $u$ of long DCI format. Figure 2(b) denotes the sequence $u$ of conventional polar codes with $S$ information bits. And Figure 2(c) denotes the sequence $u$ of short DCI formats. The black blocks are information bits which are selected by the reliability of subchannels, the white blocks are the frozen bits, and the red blocks are the information bits selected by the long DCI format.

The selection of thresholds $T$ and $I$ is not fixed, that is, depended on the practical situation. When the channel condition is good, two CWs can be transmitted in one block, and the small value of $T$ and $I$ can be set; otherwise set a large value of $T$ and $I$.

*3.2. Decoding.* The proposed decoding process is generally based on the SCL decoder with adding a step called pause-and-judge (PJ). The detail of the PJ step is described as follows.

Based on successive decoding property, the polar decoder can decode the information from $u_1$ to $u_N$ bit by bit. That is, the polar decoder can pause when $u_1, u_2, \ldots, u_I$ have been decoded and proceed with other operations.

In the proposed scheme, user receives message from the base station without the knowledge of DCI format used. The difference between coded information of long DCI format and that of the short DCI format is selection of information bits set, but the first $I$ elements of RI field are the same. The

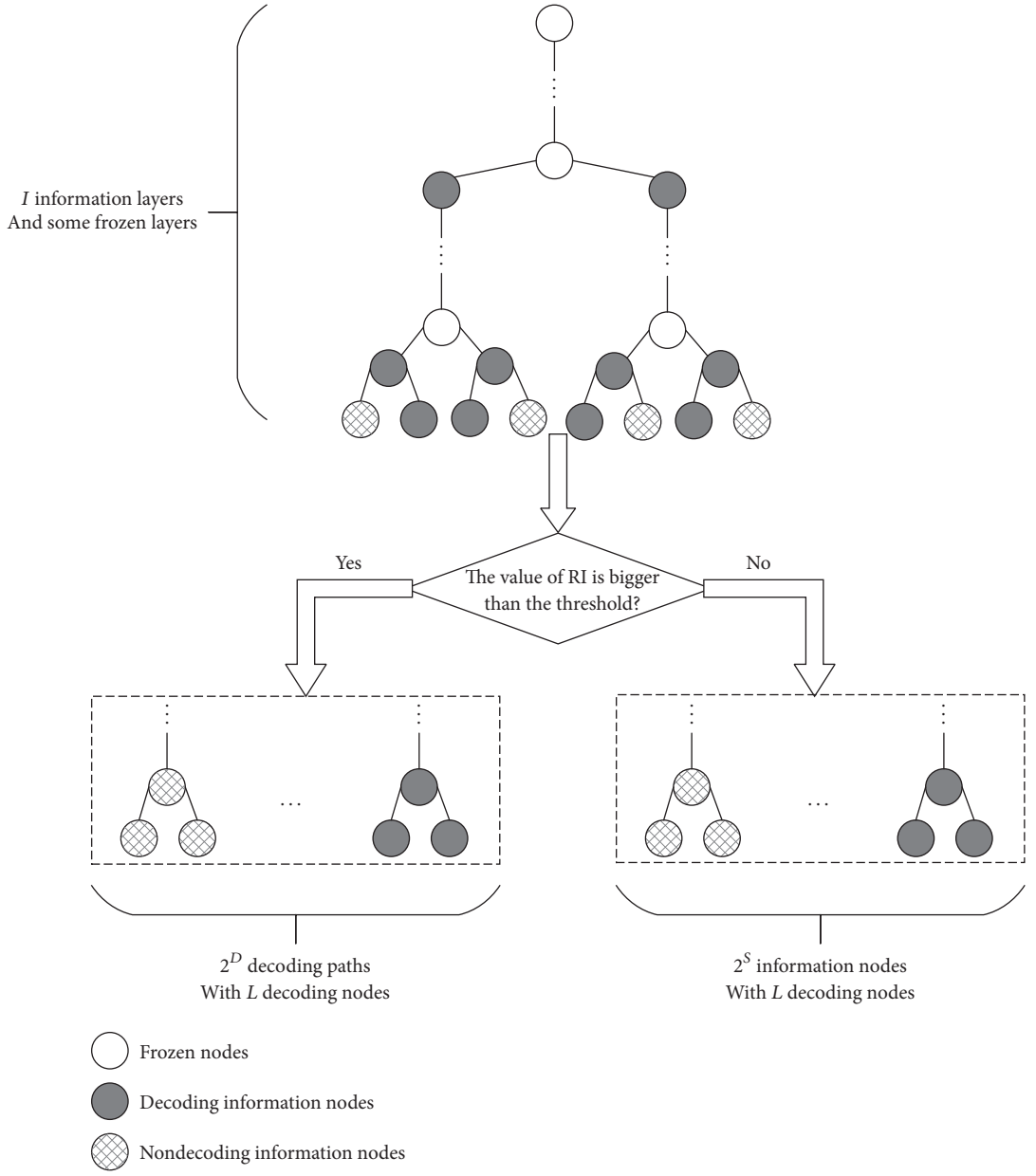



Figure 3: The proposed SCL decoding with $L$ decoding paths.

user decodes information from $u_1$ with the SCL algorithm, when $u_1$ to $u_I$ have been decoded, the decoding proceeding is paused and user selects the most reliability path to decode $u_1, u_2, \ldots, u_I$ and judges the value of $u_1, u_2, \ldots, u_I$. If it is bigger than threshold $T$, user continues decoding according to the long information set $A_{(D+I)}$. Otherwise, user continues decoding according to the short information set $A_S$. The decoding process is shown as in Figure 3.

## 4. Analysis and Simulation Results

*4.1. Complexity of Blind Detection.* In the LTE system, the blind detection of PDCCH can be divided into two parts, detection of search space and detection of DCI format. The DCI can be placed in various valid locations which form the so-called search space. There are 22 candidate search spaces. And two DCI formats are allowed to be configured to UE simultaneously. UE does not know which search space and DCI format are transmitted. Therefore, the most amount of blind detection of PDCCH is $22 * 2 = 44$ times. Our work saves the complexity of blind detection which is used to determine DCI format, the most amount of blind detection of PDCCH is decreased to 22 times.

With proposed scheme, before the blind detection, we can decide which DCI format base station uses by the dynamic configuration of DCI format. That is, we can save up to 50% for complexity of blind detection.

*4.2. Block Error Rate (BLER).* In this section we provide numerical examples to illustrate that the proposed scheme

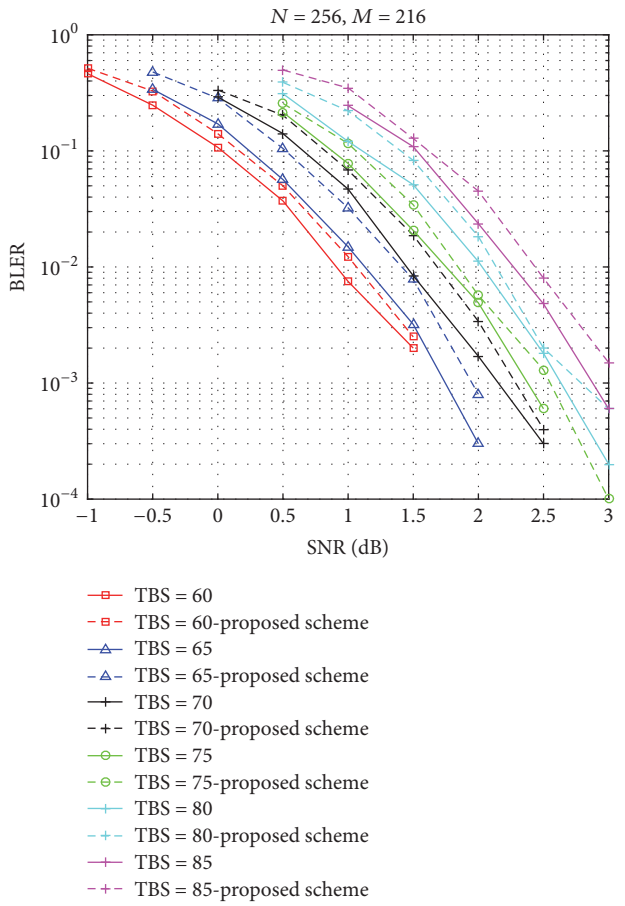



Figure 4: The BLER performance for proposed scheme with $N = 256$, $M = 216$, and various TBS.

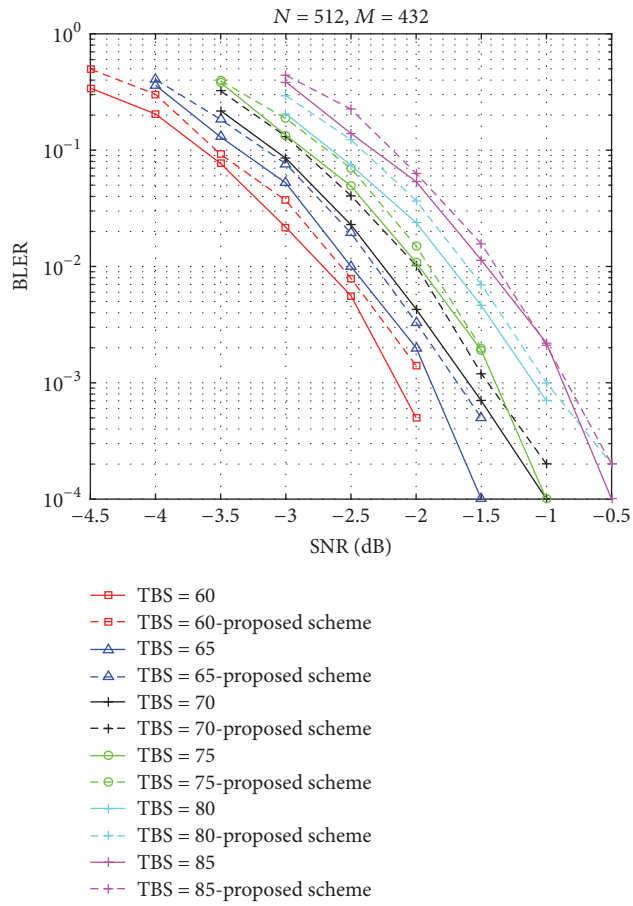



Figure 5: The BLER performance for proposed scheme with $N = 512$, $M = 432$, and various TBS.

has lower complexity and the degradation caused by dynamic indication of code word number is generally negligible. When DCI format uses long information set, the selection of information set of proposed scheme is the same as that of common polar codes; there is hardly any BLER performance loss. Therefore the BLER of short information set with $I$ RI field is needed to be simulated.

We consider the 216-, 432-, and 864-bit DCI payload with the various length of the short information from 60 bits to 85 bits. Note that $M = 216/432/864$ corresponds to aggregation level 2/4/8 with 1/4 RS density per resource element group (REG), respectively. And the mother code lengths are $N = 256/512/1024$, respectively. The proposed scheme is compared with common polar codes. The simulation results are shown as in Figures 4–6. Figure 4 shows that, with a shorter DCI payload length ($M = 216$), a considerable performance degradation on BLER is caused by dynamic indication, less than 0.1 dB, when BLER is $10^{-2}$. Figures 5 and 6 show that, with a longer encoded bit length ($N = 432/864$), the performance loss is generally negligible. It is obvious that a larger aggregation level would be more possibly used for a larger DCI payload size (e.g., DCI format with 2 CWs) to ensure the reliable reception of NR-PDCCH on UE side, under which the BLER performance is hardly impacted by dynamic indication.

*4.3. False Alarm Rate (FAR).* The false alarm ratio has been one of the most important measures of channel coding in 3GPP. The FAR is defined as FAR $= N_u/N_t$, where $N_u$ denotes the number of undetected erroneous packets and $N_t$ denotes the number of total packets. Figures 7 and 8 show that the FAR for the proposed scheme over various $(K, M)$ pairs can comply with the FAR target of $1.5 * 2^{-21}$.

## 5. Conclusion

In this paper, we proposed the dynamic configuration of DCI format based on polar codes. In the encoder of proposed scheme, the RI field is used to indicate the number of MCS fields and other related fields. Then the pause-and-judge step is added to the SCL scheme. Analysis and the simulation results illustrate that the proposed scheme can reduce the complexity of the blind detection and the degradation caused by dynamic indication of code word number is generally negligible.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

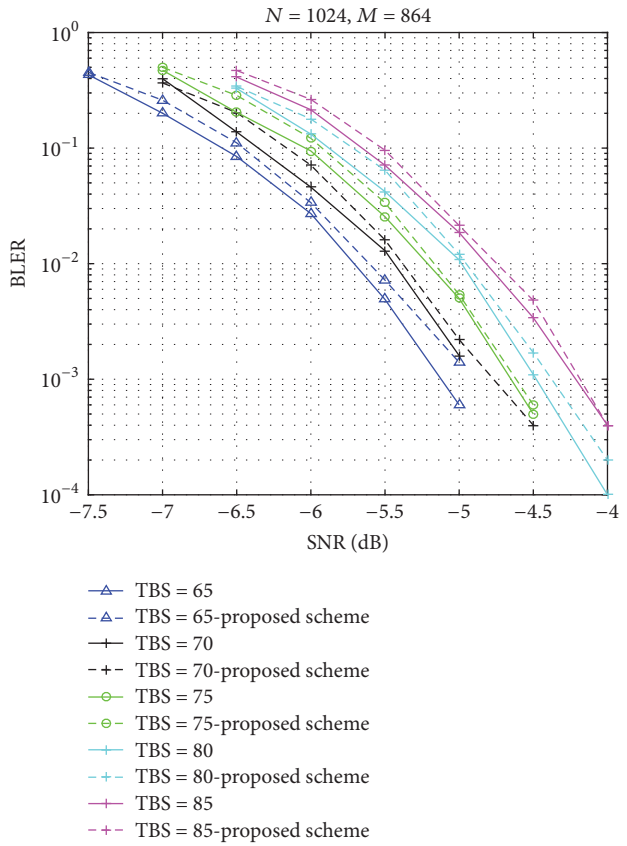



FIGURE 6: The BLER performance for proposed scheme with $N = 1024$, $M = 864$, and various TBS.

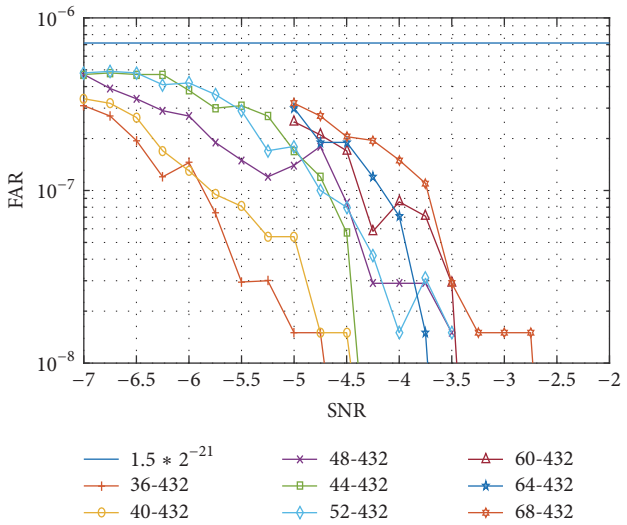



FIGURE 7: FAR evaluation results for $M = 432$.

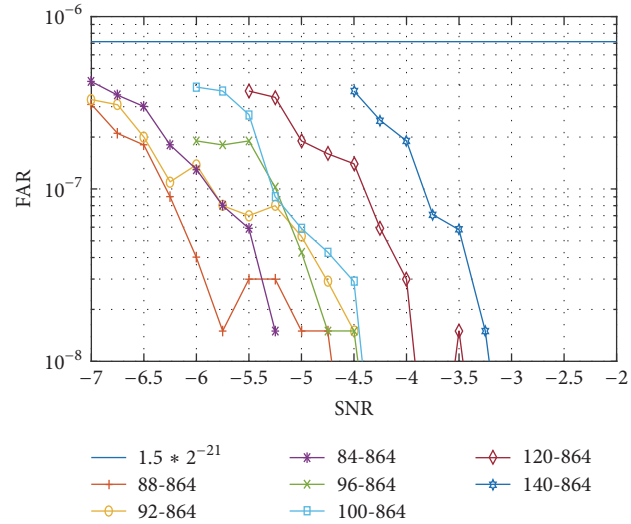



FIGURE 8: FAR evaluation results for $M = 864$.

## Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grant no. 61371075 and Beijing Municipal Science and Technology Project under Grant D171100006317001.

## References

[1] R. Bajracharya, R. Shrestha, Y. B. Zikria, and S. W. Kim, "LTE or LAA: Choosing Network Mode for My Mobile Phone in 5G Network," in *Proceedings of the 2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, pp. 1–4, Sydney, NSW, June 2017.

[2] G. Pecoraro, S. Di Domenico, E. Cianca, and M. De Sanctis, "LTE signal fingerprinting localization based on CSI," in *Proceedings of the 2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 1–8, Rome, October 2017.

[3] B. Tahir and M. Rupp, "New construction and performance analysis of Polar codes over AWGN channels," in *Proceedings of the 24th International Conference on Telecommunications, ICT 2017*, Cyprus, May 2017.

[4] H. Mzoughi, F. Zarai, M. S. Obaidat, and L. Kamoun, "3GPP LTE-Advanced Congestion Control Based on MIH Protocol," *IEEE Systems Journal*, vol. 11, no. 4, pp. 2345–2355, 2017.

[5] E. Arikan, "Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[6] C. Cao, Z. Fei, J. Yuan, and J. Kuang, "Low complexity list successive cancellation decoding of polar codes," *IET Communications*, vol. 8, no. 17, pp. 3145–3149, 2014.

[7] K. Niu, K. Chen, J. Lin, and Q. T. Zhang, "Polar codes: Primary concepts and practical decoding algorithms," *IEEE Communications Magazine*, vol. 52, no. 7, pp. 192–203, 2014.

[8] C. Condo, S. A. Hashemi, and W. J. Gross, "Blind Detection with Polar Codes," *IEEE Communications Letters*, vol. 21, no. 12, pp. 2550–2553, 2017.

[9] P. Giard, A. Balatsoukas-Stimming, and A. Burg, "Blind detection of polar codes," in *Proceedings of the 2017 IEEE International Workshop on Signal Processing Systems (SiPS)*, pp. 1–6, Lorient, October 2017.

[10] S. A. Hashemi, C. Condo, and W. J. Gross, "Fast Simplified Successive-Cancellation List Decoding of Polar Codes," in *Proceedings of the 2017 IEEE Wireless Communications and Networking Conference Workshops, WCNCW 2017*, USA, March 2017.

[11] Q. Zhang, A. Liu, X. Pan, and K. Pan, "CRC Code Design for List Decoding of Polar Codes," *IEEE Communications Letters*, vol. 21, no. 6, pp. 1229–1232, 2017.

[12] R. Mori and T. Tanaka, "Performance of polar codes with the construction using density evolution," *IEEE Communications Letters*, vol. 13, no. 7, pp. 519–521, 2009.

[13] P. Trifonov, "Efficient design and decoding of polar codes," *IEEE Transactions on Communications*, vol. 60, no. 11, pp. 3221–3227, 2012.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 7286909, 8 pages
https://doi.org/10.1155/2018/7286909

*Research Article*

# Performance Analysis of CRC Codes for Systematic and Nonsystematic Polar Codes with List Decoding

**Takumi Murata and Hideki Ochiai**

*Department of Electrical and Computer Engineering, Yokohama National University, 79-5 Tokiwadai, Hodogaya, Yokohama, Kanagawa 240-8501, Japan*

Correspondence should be addressed to Takumi Murata; murata-takumi-wf@ynu.jp

Received 24 November 2017; Revised 22 February 2018; Accepted 20 March 2018; Published 8 May 2018

Academic Editor: Qin Huang



Successive cancellation list (SCL) decoding of polar codes is an effective approach that can significantly outperform the original successive cancellation (SC) decoding, provided that proper cyclic redundancy-check (CRC) codes are employed at the stage of candidate selection. Previous studies on CRC-assisted polar codes mostly focus on improvement of the decoding algorithms as well as their implementation, and little attention has been paid to the CRC code structure itself. For the CRC-concatenated polar codes with CRC code as their outer code, the use of longer CRC code leads to reduction of information rate, whereas the use of shorter CRC code may reduce the error detection probability, thus degrading the frame error rate (FER) performance. Therefore, CRC codes of proper length should be employed in order to optimize the FER performance for a given signal-to-noise ratio (SNR) per information bit. In this paper, we investigate the effect of CRC codes on the FER performance of polar codes with list decoding in terms of the CRC code length as well as its generator polynomials. Both the original nonsystematic and systematic polar codes are considered, and we also demonstrate that different behaviors of CRC codes should be observed depending on whether the inner polar code is systematic or not.

## 1. Introduction

Polar codes, proposed by Arıkan [1], are known to achieve the symmetric capacity for any given binary-input memoryless channels (B-MCs) with low complexity at both encoder and decoder. Polar codes are characterized by *channel polarization* which is caused by channel splitting and channel combining, and information bits are transmitted over good channels. To identify those channels accurately, first Arıkan proposed a technique which recursively updates mutual information of the channels. However, complete estimation of mutual information can be performed over the binary erasure channel only. Therefore, several researchers proposed channel estimation techniques such as density evolution [2], Gaussian approximation [3], and channel upgrading/degrading [4]. Their performance, however, turn out to be inferior to modern capacity approaching codes such as turbo and low-density parity-check (LDPC) codes when they are compared under the constraint of the same block length.

Various types of decoding algorithms have been proposed for polar codes, such as successive cancellation list (SCL) [5, 6], belief propagation [7], and linear program [8]. In particular, it is well known that the SCL decoder can significantly improve the performance of the polar codes. Furthermore, based on the observation that the correct codeword is in the list but not necessarily the most likely one, Tal and Vardy proposed the use of error detecting codes, such as cyclic redundancy-check (CRC) codes, in order to identify the correct codeword from the list. The effectiveness of such CRC-assisted SCL decoder has been subsequently recognized by several other researchers [9, 10]. Software implementation of its fast decoder has been also developed in [11].

The performance of the polar codes concatenated with CRC code as its outer code, together with the list decoding and CRC code detection (or list-CRC decoding [11]), depends on the length of the CRC code and its generator polynomial. To date, however, little attention has been paid to the structure of CRC codes themselves. Most of the preceding studies

employ the CRC codes with 16 bits or longer and their length has been determined empirically. Otherwise, perfect error detection capability is assumed for simulation purpose, assuming that *ideal* CRC codes are employed. Nevertheless, if the block length of the code is short, the relative redundancy associated with CRC codes becomes dominant and thus reduces the overall efficiency of the concatenated code. The CRC code design for convolutional codes has been recently studied in [12], where the optimal CRC codes are identified by developing the equivalent composite convolutional codes based on combination of the original convolutional codes and CRC codes. More recently, an optimal CRC search algorithm for polar codes by using SCL decoder is developed in [13]. However, identifying the best CRC code length for the concatenated polar coding system has remained unknown. In this work, therefore, we attempt to analyze the effect of the CRC code length on the resulting frame error rate (FER) performance of the concatenated polar codes with list-CRC decoding. We first show that the miss-detection error probability of the CRC codes that depends on their *length* directly leads to degradation in terms of the FER performance. We also demonstrate that even if the length of CRC codes is identical, the performance of the list-CRC decoding may be affected by the generator polynomials of CRC codes employed.

Systematic polar coding, also proposed by Arıkan [14], is an effective approach to improve the performance of the conventional polar codes in terms of *bit* error rate (BER). However, the price of the systematic polar codes is their additional processing complexity that requires the SC decoding at the encoder side. Later, a simpler encoding scheme has been proposed in [15], where the conventional polar encoding is employed twice, instead of the SC decoder. Other efficient encoding schemes are proposed more recently in [16], where the three encoding implementations are described based on the trade-off between time and space complexity. In this work, we also address the CRC code property requirement for systematic polar codes. Based on the fact that the distance spectrum property of the systematic polar codes is different from that of the nonsystematic ones [17], we demonstrate that the CRC code structure suitable for systematic polar codes is different from that for the original nonsystematic codes.

This paper is organized as follows. In Section 2, the system model considered throughout the paper is described. Section 3 develops the relationship between the CRC code length and the resulting FER performance of the concatenated polar codes with the list-CRC decoder. In Section 4, the analytical results developed are compared with the corresponding simulations. Finally, Section 5 concludes the paper. We note that our initial results on the conventional nonsystematic polar codes are reported in the conference paper [18]. This paper is its considerable extension including the systematic polar codes and their distance spectrum properties.

## 2. System Model

We consider the binary polar codes concatenated with a binary CRC code at the transmitter, together with their SCL decoding at the receiver. Information bits of length $k$ are first encoded by CRC encoder to generate $K = k + r$ bits, where the parity bits of length $r$ are appended by the cyclic encoder for the purpose of correct information identification at the decoder. This output sequence (i.e., CRC codeword) is encoded by the inner polar encoder to generate the binary codeword of length $N$. This is modulated by binary phase-shift keying (BPSK) and transmitted over an additive white Gaussian noise (AWGN) channel. At the receiver, the likelihood is calculated from the received signal and then passed to the list-CRC decoder, where $L$ survived paths (codeword candidates) from the list decoder are tested by CRC detector starting from the most likely path. When the candidate codeword is determined to be correct by the CRC detector, the corresponding $k$ information bits are considered as the transmitted information bits.

*2.1. Polar Codes.* Following the notation of [1], let $W : \mathcal{X} \to \mathcal{Y}$ denote a binary-input memoryless channel (B-MC) with input alphabet $\mathcal{X} = \{0, 1\}$ and output alphabet $\mathcal{Y}$. We denote the corresponding channel transition probability as $W(y \mid x)$, where $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

The polar encoding is denoted by $x_1^N = u_1^N \mathbf{G}_N$, where $x_1^N = (x_1, x_2, \ldots, x_N) \in \{0, 1\}^N$ is the codeword, $u_1^N = (u_1, u_2, \ldots, u_N) \in \{0, 1\}^N$ is the information block, and $\mathbf{G}_N$ is the generator matrix of the polar codes, which is the mapping function of $\mathcal{X}^N \to \mathcal{X}^N$ [1]. The generator matrix is represented by the following recursive form:

$$\mathbf{G}_N = \mathbf{B}_N \mathbf{F}_2^{\otimes n}, \quad \mathbf{F}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \tag{1}$$

where $n = \log_2 N$, $\mathbf{F}^{\otimes n}$ denotes the $n$th Kronecker power of $\mathbf{F}$, and $\mathbf{B}_N$ is the bit-reversal permutation matrix.

Polar codes combine the $N$ input binary channels at the encoder and then split them at the decoder. After the channel combining and splitting processes, the set of $N$ binary-input coordinate channels, referred to as *bit channels* and commonly denoted by $W_N^{(i)} : \mathcal{X} \to \mathcal{Y}^N \times \mathcal{X}^{i-1}$ for $1 \le i \le N$, can be expressed as

$$\begin{aligned} W_N^{(i)}\left(y_1^N, u_1^{i-1} \mid u_i\right) &\triangleq \sum_{u_{i+1}^N \in \mathcal{X}^{N-i}} \frac{1}{2^{N-1}} W_N\left(y_1^N \mid u_1^N\right), \\ W_N\left(y_1^N \mid u_1^N\right) &= \prod_{i=1}^{N} W\left(y_i \mid x_i\right), \end{aligned} \tag{2}$$

where the set of the vectors $(y_1^N, u_1^{i-1})$ corresponds to the output of the channel $W_N^{(i)}$ and $u_i$ is its input.

Due to the channel combining and splitting operation, the mutual information of the bit channel, $I(W_N^{(i)})$, polarizes to either 0 or 1. Polar codes select $K$ out of $N$ bit positions as information bits based on their channel reliabilities. We denote the set of the selected channels for information transmission as $\mathcal{A} \subset \{1, 2, \ldots, N\}$. Its complement $\mathcal{A}^c$ contains the *frozen bits* that are known to the receiver a priori and thus are not transmitted. The code rate of the polar codes is thus given by $K/N$.

*2.2. Concatenated Encoder.* As an outer code, the CRC sequence of length $r$ bits is added to the information sequence. This CRC code is of rate $k/(k+r)$ and thus the effective rate of the polar code is $K/N = (k+r)/N$ but only $k$ bits represent information. Therefore, increasing $r$ for better error detection performance may lead to increased code rate and thus reduction of the error correcting capability of the inner polar code. Hence, the redundancy introduced by the CRC code should be designed carefully, especially when the block length $N$ is relatively short.

*2.3. List-CRC Decoding.* Polar codes with SC decoding are proved to achieve channel capacity for any binary-input memoryless channels. For a given channel index $i$ with $1 \le i \le N$, the estimated bit $\widehat{u}_i$ that corresponds to $u_i$ is expressed as

$$\widehat{u}_i = \begin{cases} \arg\max_{u_i} W_N^{(i)}\left(y_1^N, \widehat{u}_1^{i-1} \mid u_i\right), & \text{for } i \in \mathscr{A}, \\ 0, & \text{for } i \in \mathscr{A}^c, \end{cases} \tag{3}$$

where $y_1^N = (y_1, y_2, \ldots, y_N)$ is the received symbol vector and $\widehat{u}_1^{i-1} = (\widehat{u}_1, \widehat{u}_2, \ldots, \widehat{u}_{i-1})$ is the previously estimated bit vector with respect to the $i$th bit.

SCL decoding searches a limited number of paths that correspond to the input bits in a tree diagram; that is, it retains at most $L$ candidates in parallel [5, 6] for some positive integer $L$ that represents the list size.

When the $L$ most likely paths are determined at the final stage, the path that holds the highest likelihood is chosen as the most reliable path and its information sequence is tested by CRC detector (which is referred to as a *CRC test* in what follows). If the CRC test indicates that the most likely candidate is incorrect, the candidate with the second highest likelihood will be tested. This process is iterated until any of the candidates that pass the CRC test is found, or all the $L$ candidates are tested.

Although the use of CRC codes should improve the performance, increasing the length of CRC bits should reduce the efficiency as mentioned in the previous subsection. On the other hand, if the CRC code length is not sufficient, it may erroneously detect the incorrect candidate as correct one. Therefore, there should be a trade-off in the length of CRC code, which will be elaborated in the next section.

*2.4. Systematic Encoding.* For a systematic code, each codeword can be explicitly formed by the information bits and parity bits. In many conventional block codes, a systematic encoding structure is adopted for its practical advantage upon retrieval of information bits from the decoded codeword. On the other hand, the original polar codes described in Section 2.1 have a nonsystematic structure. However, it has been shown that systematic polar coding [14] outperforms the original polar codes in terms of BER even though the FER performance remains identical.

For systematic polar encoding, there are largely two encoding approaches. The first approach is the basic technique for general linear codes where the parity bits are identified from the generator matrix, and the second one is to employ the SC decoder as a part of encoding process, which is briefly described as follows: let $x_1^N = \{\mathbf{x}_{\mathscr{A}}, \mathbf{x}_{\mathscr{A}^c}\}$ be a codeword vector, where $\mathbf{x}_{\mathscr{A}}$ denotes a part of the codeword corresponding to information index and $\mathbf{x}_{\mathscr{A}^c}$ denotes the corresponding parity bits. First, information sequence $u_1^K$ is set as $\mathbf{x}_{\mathscr{A}}$ and all the bits in $\mathbf{x}_{\mathscr{A}^c}$ are set as erased bits, and then decoding of the codeword is performed by the SC decoder. Consequently, we obtain a temporary information sequence $u'^N_1 = \{\mathbf{u}'_{\mathscr{A}}, \mathbf{u}'_{\mathscr{A}^c}\}$, and then $\mathbf{x}_{\mathscr{A}^c}$ is determined by encoding the temporary sequence $u'^N_1$.

## 3. Performance Analysis of List-CRC Decoding

The exact performance analysis of SC decoding is challenging due to correlation between codeword bits. For a binary erasure channel (BEC), Parizi and Telatar have shown that the correlations between the erasure events decay fast and thus the union bound on the frame error probability becomes tight as the codeword length increases [19]. More recently, Shuval and Tal have derived an improved lower bound for a binary memoryless symmetric channel based on the correlation between the codeword bits [20].

In this section, we analyze the performance of the list-CRC decoding for a given length of CRC bits $r$, provided that the statistical distribution of the correct candidate in the list is available. We note that the distribution itself is difficult to analyze, and thus we simply assume that it is obtained by resorting to Monte-Carlo simulation based on the conventional list decoding without concatenation of error detecting codes.

*3.1. Ideal Decoding Error Probability.* Let $\mathscr{C}_p \subset \mathscr{X}^N$ denote the $(N, k+r)$ polar code as our inner code and let $\mathscr{C}_c \subset \mathscr{X}^{k+r}$ denote a set of the codewords encoded by $r$-bit CRC encoder as our outer code.

For a given information sequence $u_1^k \in \mathscr{X}^k$, $c_1^{k+r} \in \mathscr{C}_c$ represents the corresponding CRC codeword. Likewise, we denote the codeword of polar codes by $x_1^N \in \mathscr{C}_p$.

For a given received sequence $y_1^N \in \mathscr{Y}^N$, the estimated codeword of list decoder $\widehat{c}_1^{k+r}$ is expressed by

$$\widehat{c}_1^{k+r} = \arg\max_{c_{\mathscr{A}} \in \mathscr{C}_c} W_N\left(y_1^N \mid c_{\mathscr{A}}, c_{\mathscr{A}^c}\right), \tag{4}$$

where $c_{\mathscr{A}}$ and $c_{\mathscr{A}^c}$ are the vectors corresponding to information and frozen bits, respectively, and the resulting estimated information sequence $\widehat{u}_1^k$ is denoted by

$$\widehat{u}_1^k = \mathbb{I}_c\left(\widehat{c}_1^{k+r}\right), \tag{5}$$

where $\mathbb{I}_c(\cdot)$ is the decoding operation of $\mathscr{C}_c$ which maps $\mathscr{X}^{k+r} \to \mathscr{X}^k$.

Let $L$ denote the number of the candidate codewords generated by the list decoder, let $\mathscr{L} = \{1, 2, \ldots, L\}$ be a set of indices in the list, and let $\ell^\star$ denote the list index which corresponds to the correct codeword. Here, without loss of generality we assume that smaller indices correspond to more likely candidates. If the correct codeword is not in the list, that

is, $\ell^\star \notin \mathscr{L}$, we assume that $\ell^\star > L$. Moreover, let $p_{\ell^\star}(\gamma, l) =$ the probability conditioned on $\gamma$, that is, $\Pr\{\ell^\star = l \mid \gamma\}$ denote the distribution of $\ell^\star$, that is, the probability that the index $\ell^\star$ agrees with $l$ at a given SNR $\gamma = E_b/N_0$ over an AWGN channel.

If we assume the ideal case where the correct codeword in the list can be identified without error, the ideal decoding error probability $P_{e,\mathrm{id}}(\gamma, L)$ is expressed as

$$P_{e,\mathrm{id}}(\gamma, L) = 1 - \sum_{l=1}^{L} p_{\ell^\star}(\gamma, l). \tag{6}$$

*3.2. Undetected Error Probability of CRC Codeword.* The CRC test may fail with a certain probability $P_{\mathrm{ud}}$ (i.e., undetected error probability of the CRC codeword). As $P_{\mathrm{ud}}$ increases, the error correcting performance of the concatenated polar codes may be degraded.

In general, the undetected error probability of the code $\mathscr{C}_c$ over a binary symmetric channel (BSC) with crossover probability $p \in [0, 1/2]$ is given by [21]

$$P_{\mathrm{ud}}(\mathscr{C}_c, p) = \sum_{i=1}^{k+r} A_i p^i (1-p)^{k+r-i}, \tag{7}$$

where $A_i$ $(1 \le i \le k+r)$ is the weight distribution of $\mathscr{C}_c$. We note that the CRC code $\mathscr{C}_c$ may be called *good* if the relationship

$$P_{\mathrm{ud}}(\mathscr{C}_c, p) \le P_{\mathrm{ud}}\left(\mathscr{C}_c, \frac{1}{2}\right) = \frac{2^k - 1}{2^{k+r}} \approx \frac{1}{2^r} \tag{8}$$

holds for all $p$ [21]. The upper bound of (8) may be also achieved if all the binary sequences of length $k + r$; that is, all the elements of $\mathscr{X}^{k+r}$ may appear in the list with equal probability [22] (in other words, the candidates in the list are modeled as *completely random strings* [22]).

The major challenge in analyzing the probability $P_{\mathrm{ud}}$ in the case of polar codes with list-CRC decoding is that, due to the correlation among the codewords, the binary sequences in the list may be also correlated. Nevertheless, as mentioned in [22], the upper bound of $P_{\mathrm{ud}}$ in (8) would be a good estimate of the probability of the undetected error, provided that the minimum distance of CRC code is much lower than that of the candidates in the list.

*3.3. Approximate Upper Bound on Decoding Error Probability.* Given the above undetected error probability model of CRC code, the probability of correct detection by the CRC test is lower bounded as

$$P_{\mathrm{cd}} = 1 - P_{\mathrm{ud}} \ge 1 - \frac{2^k - 1}{2^{k+r}}. \tag{9}$$

On the $L$-candidate list decoding, if the $l$th estimate codeword is correct, all the codewords up to the $(l-1)$th list should be incorrect. In other words, all the $(l-1)$ codewords must be correctly detected as invalid codewords by the CRC test and each probability is bounded by (9). Hence, the correct decoding probability $P_c(\gamma, L)$ is expressed as

$$\begin{aligned} P_c(\gamma, L) &= p_{\ell^\star}(\gamma, 1) \\ &\quad + \sum_{l=2}^{L} p_{\ell^\star}(\gamma, l) \prod_{l'=1}^{l-1} \Pr\left\{c_1^{k+r}\left(l'\right) \notin \mathscr{C}_c \mid \ell^\star = l\right\} \\ &\ge p_{\ell^\star}(\gamma, 1) \\ &\quad + \sum_{l=2}^{L} p_{\ell^\star}(\gamma, l) \prod_{l'=1}^{l-1} \left\{1 - \frac{2^k - 1}{2^{k+r} - (l'-1)}\right\}, \end{aligned} \tag{10}$$

where $c_1^{k+r}(l)$ is the $l$th estimate vector of polar codes in the list and the term $2^{k+r} - (l'-1)$ corresponds to the number of remaining binary sequences of length $k + r$ prior to the $l'$th CRC test. Since $2^{k+r} \gg L$, we may express

$$P_c(\gamma, L) \gtrsim \sum_{l=1}^{L} p_{\ell^\star}(\gamma, l) \left(1 - 2^{-r}\right)^{l-1}. \tag{11}$$

Thus, the approximate upper bound of the decoding error probability for the list CRC-concatenated system $P_e(\gamma, L)$ with the candidates of *completely random strings* is expressed as

$$\begin{aligned} P_e(\gamma, L) = 1 - P_c(\gamma, L) &\lesssim 1 - \sum_{l=1}^{L} p_{\ell^\star}(\gamma, l) \left(1 - 2^{-r}\right)^{l-1} \\ &\approx 1 - \sum_{l=1}^{L} p_{\ell^\star}(\gamma, l) \left\{1 - (l-1) 2^{-r}\right\} \end{aligned} \tag{12}$$

$$= P_{e,\mathrm{id}}(\gamma, L) + \underbrace{2^{-r} \sum_{l=1}^{L} p_{\ell^\star}(\gamma, l)(l-1)}_{\triangleq P(\gamma, r, L)}. \tag{13}$$

We observe that the first term in (13) corresponds to the ideal error performance when CRC test is perfect and is derived in (6), whereas the second term $P(\gamma, r, L)$ in (13) is the performance degradation associated with the imperfect CRC code, which can be reduced by increasing the redundant bit length $r$. Therefore, we should select $r$ such that $P(\gamma, r, L)$ is negligible compared to $P_{e,\mathrm{id}}(\gamma, L)$.

## 4. Simulation Results

In this section, we investigate the FER performance of the CRC-concatenated nonsystematic and systematic polar codes with list-CRC decoding through various simulations with emphasis on the difference in the CRC code length as well as its generator polynomials. Throughout simulations, we employ BPSK for modulation and the channel model is AWGN. The polar codes simulated are designed according to [2] with design-SNR $R$ $E_b/N_0 = -1.5917$ dB (as proposed in [23]), where $R$ is the rate of the inner polar codes. The list size

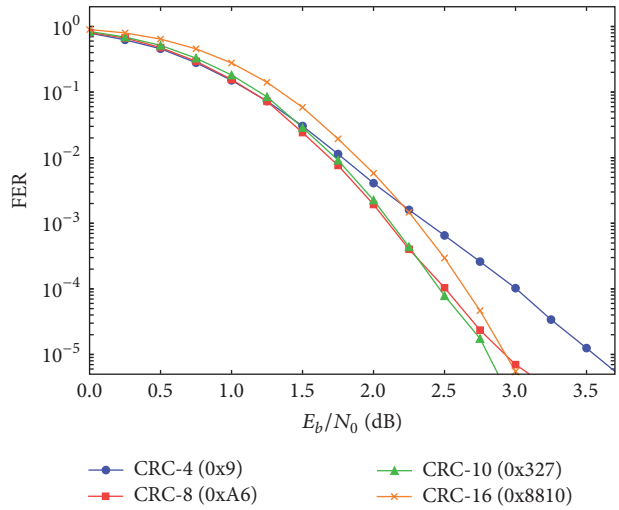



Figure 1: FER performance of list-CRC decoding with various lengths of CRC code with the code length $N = 512$.

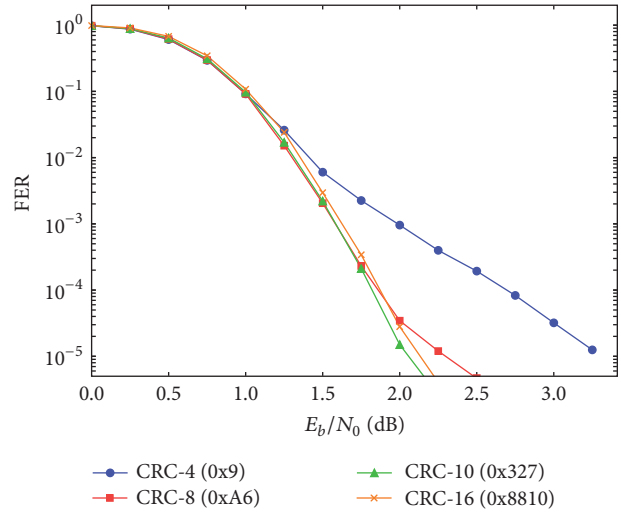



Figure 2: FER performance of list-CRC decoding with various lengths of CRC with the code length $N = 2048$.

of the SCL decoding is chosen as $L = 32$. We also consider the fixed spectral efficiency scenario where the code rate of the entire system is 1/2. This is achieved by setting the parameters of polar codes as $(N, N/2 + r)$ with $N$ and $r$ representing the code length of the polar code and the redundant bits imposed by the CRC code, respectively. Therefore, increasing CRC code length (for better list-CRC decoding performance) leads to an increment of the code rate of polar codes with no effective increase of information bits, which in turn reduces the error correction capability. Therefore, it should reveal the trade-off relationship.

*4.1. Comparison of CRC Code Length.* We first examine the trade-off between the CRC code length and the code rate of polar codes. In what follows, we describe the generator polynomial of CRC encoders by the hexadecimal representation; for example, the notation 0x8810 for 16-bit CRC encoder indicates 1000 1000 0001 0000 in binary representation, which corresponds to $x^{16} + x^{12} + x^5$, and addition of the implicit "+1" term specifies the generator polynomial of $g(x) = x^{16} + x^{12} + x^5 + 1$ [24].

Figures 1 and 2 compare the FER performance of the conventional nonsystematic polar codes with code lengths $N = 512$ and 2048, respectively, where the CRC code length $r$ is chosen from 4, 8, 10, and 16. In this example, only the results with the CRC polynomials that are found to be best among those compared for each $r$ are shown.

From the results with $N = 512$ shown in Figure 1, because of the loss in terms of code rate, the case of CRC-16, which is a commonly adopted scenario in the study of CRC-concatenated polar codes, turns out to be inferior to that of CRC-10. Therefore, the negative effect of the rate loss associated with the CRC code length becomes dominant compared to the improvement achieved by the reduction of the undetected error probability $P_{ud}$ for the short polar codes. On the other hand, when $N = 2048$, the performance gap between CRC-16 and CRC-10 becomes smaller as observed in Figure 2. This stems from the fact that as $N$ increases the effect of the relative loss in the code rate becomes negligibly small.

For both cases, the performance loss due to the miss-detection of CRC codes becomes noticeable when $r$ is small.

*4.2. Performance Comparison of Analytical and Simulation Results.* We next examine the validity of the FER upper bound expression given by (12). We first obtain the statistical distribution of the correct paths $\ell^\star$ by Monte-Carlo simulation (without concatenating CRC code). This has been done by counting the locations of the correct paths through the simulation of $10^7$ trials. We note that the precise FER performance depends on the particular realization of the polar codes, and hence its theoretical characterization should be challenging. It is thus left as future work.

*4.2.1. Comparison with Generator Polynomials for Nonsystematic Polar Codes.* We first consider the case of $N = 2048$ and $r = 8$ for nonsystematic polar codes. We have selected two specific groups of generator polynomials of CRC codes: (1) all the eighth-order primitive polynomials and (2) all the seventh-order primitive polynomials multiplied by $(x + 1)$. The multiplication of the factor $(x + 1)$ makes the CRC detector capable of detecting all the odd Hamming weight errors and thus is found to be preferable in some applications. Also, the cases where we insert a random interleaver after CRC encoder are considered where the block diagram of this particular scenario is depicted in Figure 3.

The results for the generator polynomials consisting of all the eighth-order primitive polynomials are compared in Figure 4. We observe that without interleaving, some specific CRC polynomials may poorly perform compared to the analytical results. This may be caused by the fact that bit error patterns due to the SC decoder, namely, the sequences in the list, and the distance property of the CRC codes do not match well for this system. However, with interleaving, the resulting performance becomes similar and comparable with

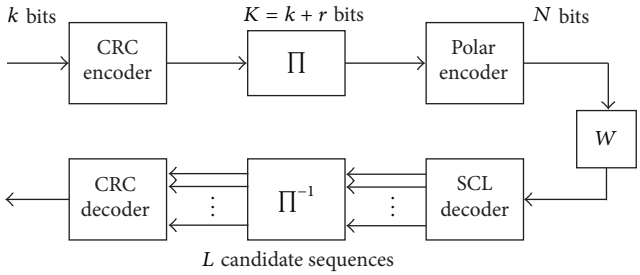



Figure 3: A block diagram of the CRC-concatenated polar codes with an interleaver.

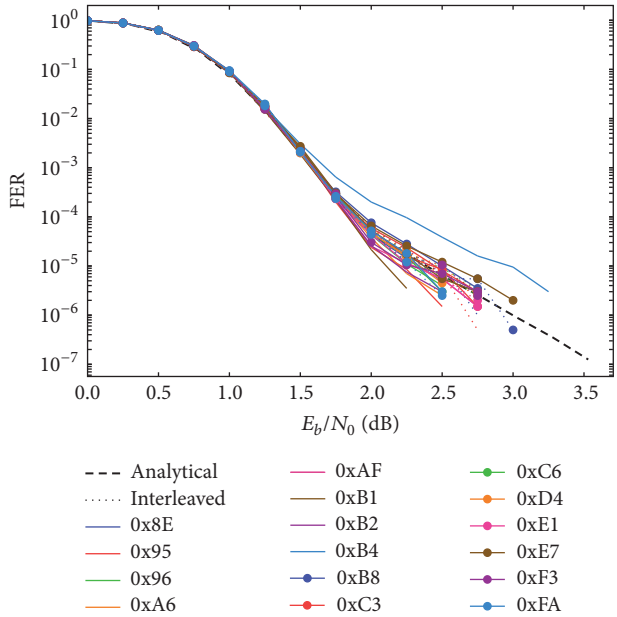



Figure 4: Comparison of FER performances between the analytical bound and full simulations for list-CRC decoding with the CRC codes based on the eighth-order primitive polynomials. The dashed line represents the analytical result. The dotted and solid lines correspond to the simulation results with and without interleaving, respectively.

the analytical result (derived based on the assumption of the list with *completely random strings*).

The results for the seventh-order primitive polynomials multiplied by $(x + 1)$ are compared in Figure 5. In this case, all the simulation results without insertion of interleaver are inferior to the analytical bound as opposed to the results in Figure 4. Insertion of interleavers may mitigate this gap, but we still observe that the performance achieved by the CRC codes designed in this manner will be poor compared to the cases shown in Figure 4. Therefore, the introduction of the term $(x+1)$ of the CRC polynomial, which makes all the odd-weight errors detectable, may not be effective in the case of the investigated nonsystematic polar codes. The reason for this behavior will be elucidated through the comparison with the systematic polar codes in what follows.

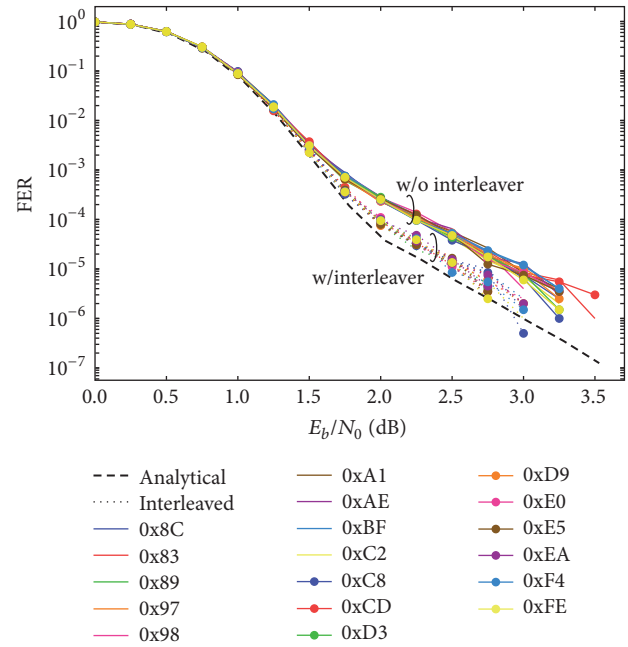



Figure 5: Comparison of FER performances between the analytical bounds and full simulations for list-CRC decoding with CRC codes based on the seventh-order primitive polynomials multiplied by $(x + 1)$. The dashed line represents the analytical result. The dotted and solid lines correspond to the simulation results with and without interleaving, respectively.

Table 1: The number of even and odd weight inputs with code length $N = 512$, code rate $R = 1/2$, and list size $L = 2^{20}$.

| Output weight | $d_{\min} = 8$ | | $d = 16$ | |
|---|---|---|---|---|
| Input weight | Odd | Even | Odd | Even |
| Nonsystematic | 1 | 63 | 3470 | 114297 |
| Systematic | 32 | 32 | 55840 | 62114 |

*4.2.2. Comparison with Generator Polynomials for Systematic Polar Codes.* We next consider the case of systematic polar coding, where all the parameters are chosen identically to the previous results with nonsystematic encoding. The results for the eighth-order primitive polynomials are compared in Figure 6. These results show that the concatenated polar coding systems, both with and without interleaver, agree with the analytical result.

The results for the seventh-order primitive polynomials multiplied by $(x + 1)$ are compared in the case of systematic polar codes in Figure 7. In contrast to the results of nonsystematic polar codes shown in Figure 5, the systematic polar codes are not affected by whether the factor $(x + 1)$ is present or not in the generator polynomial of the CRC codes. This performance difference between two encoding approaches can be explained by the input-output weight enumerator function (IOWEF) of polar codes.

In Table 1, for the codeword of the polar codes with the first and second lowest Hamming weights (i.e., $d = 8$ and 16 in our case), the corresponding numbers of the odd and even Hamming weight information sequences are compared. The

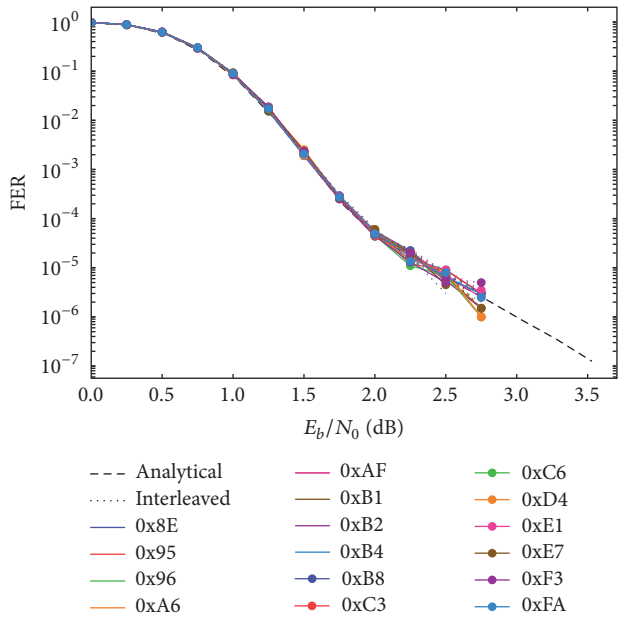



Figure 6: Comparison of FER performances between the analytical bound and full simulations for systematic polar codes and list-CRC decoding with the CRC codes based on the eighth-order primitive polynomials. The dotted and solid lines correspond to the simulation results with and without interleaving, respectively.

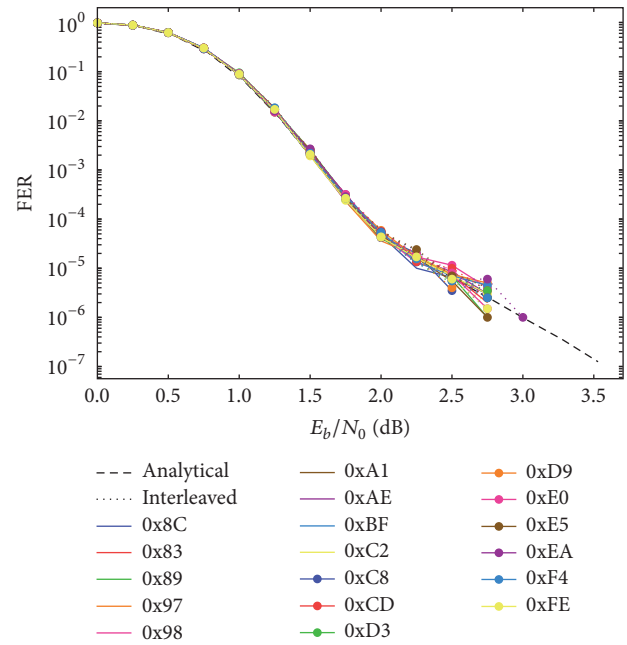



Figure 7: Comparison of FER performances between the analytical bounds and full simulations for systematic polar codes and list-CRC decoding with the CRC codes based on the seventh-order primitive polynomials multiplied by $(x + 1)$. The dotted and solid lines correspond to the simulation results with and without interleaving, respectively.

IOWEF is calculated by the method based on SCL decoding with generation of huge list size as proposed in [17], and the associated parameters here are set as $N = 512$, $R = 1/2$, and $L = 2^{20}$. The result of the conventional nonsystematic polar codes in the table indicates that the even-weight input sequence is dominant when the minimum Hamming weight codeword is generated, whereas the numbers of odd and even-weight input sequences are equal in the case of systematic polar codes. As the SNR increases, the candidate codewords are dominated by those with the minimum Hamming weight. Therefore, for the nonsystematic polar codes where the even-weight errors occur frequently, the use of the CRC polynomials with the factor $(x + 1)$, which are capable of detecting all the odd-weight errors, becomes ineffective. On the other hand, since the systematic polar codes have balanced minimum Hamming weights of even and odd, the inclusion of the factor $(x + 1)$ may not affect the undetected error probability.

## 5. Conclusion

In this work, we have investigated the performance of list-CRC decoder for CRC-concatenated polar codes. In particular, we have focused on the relationship between the FER performance and CRC code length by deriving the analytical bound of the FER. Furthermore, the effect of the generator polynomials of CRC codes on the polar code structure has been analyzed. Specifically, it has been firstly shown that there is a trade-off relationship between the FER performance and CRC code length when the overall code rate of the concatenated system is identical. In general, if the CRC is longer than necessary for a given target FER, the performance degradation will be expected due to the relative increase in the code rate of the inner polar codes. Secondly, it has been shown that the performance of the nonsystematic polar codes depends on the generator polynomials of CRC codes, and it is not effective in employing the term $(x + 1)$ that makes all the odd-weight errors detectable from the viewpoint of their distance spectrum properties. This is in contrast to the systematic polar codes, where the performance is robust against the structure of the CRC codes.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported in part by MIC/SCOPE (Grant no. 155003003) and by JSPS KAKENHI (Grant no. JP16H02345).

## References

[1] E. Arıkan, "Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[2] R. Mori and T. Tanaka, "Performance of polar codes with the construction using density evolution," *IEEE Communications Letters*, vol. 13, no. 7, pp. 519–521, 2009.

[3] P. Trifonov, "Efficient design and decoding of polar codes," *IEEE Transactions on Communications*, vol. 60, no. 11, pp. 3221–3227, 2012.

[4] I. Tal and A. Vardy, "How to construct polar codes," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 59, no. 10, pp. 6562–6582, 2013.

[5] I. Tal and A. Vardy, "List decoding of polar codes," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pp. 1–5, 2011.

[6] I. Tal and A. Vardy, "List decoding of polar codes," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 61, no. 5, pp. 2213–2226, 2015.

[7] E. Arıkan, "A performance comparison of polar codes and reed-muller codes," *IEEE Communications Letters*, vol. 12, no. 6, pp. 447–449, 2008.

[8] N. Goela, S. B. Korada, and M. Gastpar, "On LP decoding of polar codes," in *Proceedings of the IEEE Information Theory Workshop (ITW '10)*, pp. 1–5, Dublin, Ireland, September 2010.

[9] K. Niu and K. Chen, "CRC-aided decoding of polar codes," *IEEE Communications Letters*, vol. 16, no. 10, pp. 1668–1671, 2012.

[10] B. Li, H. Shen, and D. Tse, "An adaptive successive cancellation list decoder for polar codes with cyclic redundancy check," *IEEE Communications Letters*, vol. 16, no. 12, pp. 2044–2047, 2012.

[11] G. Sarkis, P. Giard, A. Vardy, C. Thibeault, and W. J. Gross, "Fast List Decoders for Polar Codes," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 2, pp. 318–328, 2016.

[12] C.-Y. Lou, B. Daneshrad, and R. D. Wesel, "Convolutional-code-specific CRC code design," *IEEE Transactions on Communications*, vol. 63, no. 10, pp. 3459–3470, 2015.

[13] Q. Zhang, A. Liu, X. Pan, and K. Pan, "CRC Code Design for List Decoding of Polar Codes," *IEEE Communications Letters*, vol. 21, no. 6, pp. 1229–1232, 2017.

[14] E. Arıkan, "Systematic polar coding," *IEEE Communications Letters*, vol. 15, no. 8, pp. 860–862, 2011.

[15] G. Sarkis, P. Giard, A. Vardy, C. Thibeault, and W. J. Gross, "Fast polar decoders: algorithm and implementation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 5, pp. 946–957, 2014.

[16] H. Vangala, Y. Hong, and E. Viterbo, "Efficient algorithms for systematic polar encoding," *IEEE Communications Letters*, vol. 20, no. 1, pp. 17–20, 2016.

[17] Z. Liu, K. Chen, K. Niu, and Z. He, "Distance spectrum analysis of polar codes," in *Proceedings of the 2014 IEEE Wireless Communications and Networking Conference, WCNC 2014*, pp. 490–495, Turkey, April 2014.

[18] T. Murata and H. Ochiai, "On design of CRC codes for polar codes with successive cancellation list decoding," in *Proceedings of the 2017 IEEE International Symposium on Information Theory, ISIT 2017*, pp. 1868–1872, Germany, June 2017.

[19] M. B. Parizi and E. Telatar, "On the correlation between polarized BECs," in *Proceedings of the 2013 IEEE International Symposium on Information Theory, ISIT 2013*, pp. 784–788, Turkey, July 2013.

[20] B. Shuval and I. Tal, "A lower bound on the probability of error of polar codes over BMS channels," https://arxiv.org/abs/1701.01628v2.

[21] T. Kløve and V. I. Korzhik, *Error Detecting Codes*, Kluwer Academic Publishers, Boston, MA, USA, 1995.

[22] D. P. Bertsekas and R. G. Gallager, *Data Networks, Prentice Hall*, 2nd edition, 1992.

[23] H. Vangala, E. Viterbo, and Y. Hong, "A comparative study of polar code constructions for the AWGN channel," https://arxiv.org/abs/1501.02473.

[24] P. Koopman and T. Chakravarty, "Cyclic Redundancy Code (CRC) polynomial selection for embedded networks," in *Proceedings of the 2004 International Conference on Dependable Systems and Networks*, pp. 145–154, July 2004.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 8928761, 6 pages
https://doi.org/10.1155/2018/8928761

*Research Article*

# Adding a Rate-1 Third Dimension to Parallel Concatenated Systematic Polar Code: 3D Polar Code

**Zhenzhen Liu, Kai Niu, Chao Dong, and Jiaru Lin**

*Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China*

Correspondence should be addressed to Zhenzhen Liu; zzliu@bupt.edu.cn

Received 24 November 2017; Revised 8 March 2018; Accepted 27 March 2018; Published 3 May 2018

Academic Editor: Zesong Fei



In this paper, a three-dimensional polar code (3D-PC) scheme is proposed to improve the error floor performance of parallel concatenated systematic polar code (PCSPC). The proposed 3D-PC is constructed by serially concatenating the PCSPC with a rate-1 third dimension, where only a fraction $\lambda$ of parity bits of PCSPC are extracted to participate in the subsequent encoding. It takes full advantage of the characteristics of parallel concatenation and serial concatenation. In addition, the convergence behavior of 3D-PC is analyzed by the extrinsic information transfer (EXIT) chart. The convergence loss between PCSPC ($\lambda = 0$) and different $\lambda$ provides the reference for choosing the value of $\lambda$ for 3D-PC. Finally, the simulation results confirm that the proposed 3D-PC scheme lowers the error floor.

## 1. Introduction

The novel concept of parallel concatenated systematic polar code (PCSPC) was first put forward in [1]. PCSPC scheme consists of two systematic polar codes (SPCs) [2]. It has performance advantage with respect to original SPC. In [3], the extrinsic information transfer (EXIT) charts of different length SPC have been given. As a promotion of the above EXIT chart results, the convergence behavior of PCSPC can be analyzed. It can be observed that SPC with larger code length leads to narrower opening. Therefore, it is difficult for PCSPC with large code length SPCs to converge at low error rate. The motivation of our work is to solve this problem.

As we know, there is error floor for turbo code (TC) at block error rate (BLER) around $10^{-5}$ [4]. In order to improve the performance of TC in the error floor region, three-dimensional turbo code (3D-TC) has been studied in [5–7]. 3D-TC scheme was proposed by serially concatenating a rate-1 cyclic recursive systematic convolutional (CRSC) code to conventional TC. It is important to note that only a fraction $\lambda$ of parity bits from TC are extracted to participate in the encoding again. Compared with conventional TC, 3D-TC scheme has larger minimum distance. Therefore, 3D-TC improves the error floor performance greatly. In addition, the influence of $\lambda$ of 3D-TC on convergence threshold and minimum distance has been researched in [6, 7].

It is known from the literature that serial concatenated code has larger minimum distance with respect to parallel concatenated code; however, its convergence threshold is worse than that of parallel concatenation [8]. Meanwhile, inspired by the idea in [7], 3D polar code (3D-PC) scheme is proposed to improve the error floor performance of PCSPC in this paper. It makes full use of the features of parallel concatenation and serial concatenation. 3D-PC is constituted by adding a rate-1 CRSC code to PCSPC. And only a fraction $\lambda$ of parity bits of PCSPC are sent to the third encoder. Moreover, the convergence behavior of 3D-PC is analyzed by EXIT chart method [9]. It can be utilized to guide the choice of $\lambda$ which is an important parameter that affects the performance of 3D-PC. Simulation results corroborate the effectiveness of 3D-PC scheme to improve the low error rate performance.

The paper is organized as follows. Section 2 reviews systematic polar code and EXIT chart. 3D-PC scheme is proposed in Section 3. In Section 4, convergence analysis of 3D-PC is presented. The simulation results are shown in Section 5. Section 6 concludes this paper.

## 2. Preliminaries

*2.1. Systematic Polar Code.* Polar code is a capacity-achieving channel code which was proposed by Arıkan in [10]. Given code length $N$ and code rate $R = K/N$, the reliabilities of $N$ subchannels can be obtained by Gaussian approximation method [11] or other construction algorithms. Then the $K$ subchannels with high reliability are used to transmit information bits, and other $N-K$ subchannels are utilized to deliver frozen bits. Let set $\mathscr{A} \subset \{1, \ldots, N\}$ denote the indexes of those $K$ high reliability subchannels. Supposing that the input sequence $\mathbf{v} = (v_1, \ldots, v_N)$ is given, the codeword $\mathbf{x}$ of polar code can be obtained by

$$\mathbf{x} = \mathbf{v}\mathbf{G}_N = \mathbf{v}\left(\mathbf{B}_N\mathbf{F}_2^{\otimes n}\right), \tag{1}$$

where $\mathbf{G}_N$ is the generator matrix, $\mathbf{B}_N$ denotes the bit-reversal permutation matrix, $\otimes n$ denotes the $n$-th Kronecker product, and $\mathbf{F}_2 = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]$.

Since the input source sequence $\mathbf{v}$ can be decomposed into two parts $\mathbf{v}_{\mathscr{A}}$ and $\mathbf{v}_{\mathscr{A}^c}$, the codeword $\mathbf{x}$ in (1) can be written as

$$\mathbf{x} = \mathbf{v}\mathbf{G}_N = \mathbf{v}_{\mathscr{A}}\mathbf{G}_{\mathscr{A}} + \mathbf{v}_{\mathscr{A}^c}\mathbf{G}_{\mathscr{A}^c}, \tag{2}$$

where $\mathbf{v}_{\mathscr{A}} \subset \mathbf{v}$ is the information bits, $\mathscr{A}^c = \{1, \ldots, N\} \backslash \mathscr{A}$ denotes the complement of $\mathscr{A}$, and $\mathbf{G}_{\mathscr{A}}$ consists of the rows of $\mathbf{G}_N$ with indices in $\mathscr{A}$.

Systematic polar code is constructed based on polar code [2]. Assume that $K$-elements set $\mathscr{B}$ denotes the indexes of system bits; then $\mathbf{x}_{\mathscr{B}}$ denotes system bits and $\mathbf{x}_{\mathscr{B}^c}$ is the check bits. Equation (2) can be rewritten as

$$\mathbf{x}_{\mathscr{B}} = \mathbf{v}_{\mathscr{A}}\mathbf{G}_{\mathscr{A}\mathscr{B}} + \mathbf{v}_{\mathscr{A}^c}\mathbf{G}_{\mathscr{A}^c\mathscr{B}}, \tag{3}$$

$$\mathbf{x}_{\mathscr{B}^c} = \mathbf{v}_{\mathscr{A}}\mathbf{G}_{\mathscr{A}\mathscr{B}^c} + \mathbf{v}_{\mathscr{A}^c}\mathbf{G}_{\mathscr{A}^c\mathscr{B}^c}, \tag{4}$$

where $\mathbf{G}_{\mathscr{A}\mathscr{B}}$ denotes the submatrix of $\mathbf{G}_N$ with row indexes in $\mathscr{A}$ and column indexes belonging to $\mathscr{B}$.

As to SPC, the systematic bits $\mathbf{x}_{\mathscr{B}}$ are known and $\mathbf{v}_{\mathscr{A}^c}$ are also known and set to zero; thus $\mathbf{v}_{\mathscr{A}}$ can be calculated according to (3):

$$\mathbf{v}_{\mathscr{A}} = \left(\mathbf{x}_{\mathscr{B}} - \mathbf{v}_{\mathscr{A}^c}\mathbf{G}_{\mathscr{A}^c\mathscr{B}}\right)\left(\mathbf{G}_{\mathscr{A}\mathscr{B}}\right)^{-1} = \mathbf{x}_{\mathscr{B}}\left(\mathbf{G}_{\mathscr{A}\mathscr{B}}\right)^{-1}. \tag{5}$$

Further, the check bits $\mathbf{x}_{\mathscr{B}^c}$ can be computed by (4):

$$\mathbf{x}_{\mathscr{B}^c} = \mathbf{x}_{\mathscr{B}}\left[\left(\mathbf{G}_{\mathscr{A}\mathscr{B}}\right)^{-1}\mathbf{G}_{\mathscr{A}\mathscr{B}^c}\right]. \tag{6}$$

Here, the codeword $\mathbf{x}$ of SPC is achieved.

*2.2. EXIT Chart.* EXIT chart [9] is an efficient convergence analysis tool for the iterative decoding structure. It tracks the average mutual information of constituent decoders.

We use $X$ and $A$ to denote the transmitted bits and the corresponding a priori information, respectively. And $A$ is modeled as an independent Gaussian random variable with the following expression:

$$A = \mu_A x + n_A \tag{7}$$

with

$$u_A = \frac{\sigma_A^2}{2}, \tag{8}$$

where $n_A$ is a Gaussian random variable with mean zero and variance $\sigma_A^2$. Under the above assumption, the mutual information between transmitted bits $X$ and a priori information $A$ can be written as

$$\begin{aligned} I_A &= I(X; A) \\ &= 1 \\ &\quad - \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_A} e^{-\left(\xi - \frac{\sigma_A^2}{2}\right)^2 / 2\sigma_A^2} \log_2\left(1 + e^{-\xi}\right) d\xi. \end{aligned} \tag{9}$$

Assume that extrinsic information is denoted by $E$. The mutual information between $X$ and $E$ is calculated as

$$\begin{aligned} I_E = I(X; E) = \frac{1}{2} \cdot \sum_{x=-1,1} \int_{-\infty}^{+\infty} p_E(\xi \mid X = x) \\ \times \log_2 \frac{2 \cdot p_E(\xi \mid X = x)}{p_E(\xi \mid X = -1) + p_E(\xi \mid X = +1)} d\xi, \end{aligned} \tag{10}$$

where $p_E(\xi \mid X = x)$ is the probability distribution function given condition $X = x$. It can be obtained by Monte Carlo simulation.

## 3. Proposed 3D Polar Code Scheme

*3.1. Encoding Structure.* In short, 3D-PC scheme can be regarded as a concatenation of the inner code and outer code, PCSPC. The encoding structure of 3D-PC is illustrated in Figure 1. First of all, the input information sequence $\mathbf{u}$ with length $K$ is encoded by parallel concatenated systematic polar encoder. The component encoders of PCSPC are written as $C_a$ and $C_b$, respectively. Both of them are systematic polar encoders. We use $\mathbf{x}_a$ and $\mathbf{x}_b$ to denote the parity bits sequence of $C_a$ and $C_b$, respectively. Further, the codeword $\mathbf{x}_{\mathrm{PC}}$ can be obtained by taking the bits from $\mathbf{x}_a$ and $\mathbf{x}_b$ alternatively. The fraction $\lambda$ of $\mathbf{x}_{\mathrm{PC}}$ is interleaved by the interleaver $\Pi_c$ and sent to the postencoder $C_c$ for encoding, where $\lambda$ is named as permeability rate. And codeword $\mathbf{x}_c$ is output by the postencoder $C_c$. The parity bits chosen for encoding follow a certain puncturing pattern $\mathbf{p}$ with length $2/\lambda$. The fraction $1-\lambda$ of $\mathbf{x}_{\mathrm{PC}}$ is passed to the channel straightly, denoted by $\mathbf{x}_{\mathrm{ch}}$. The patterns $\overline{\mathbf{p}}$ and $\mathbf{p}$ are complementary. Furthermore, the last codeword $\mathbf{x}$ of 3D-PC with code length $N_T$ is obtained by combining the input sequence $\mathbf{u}$, the parity sequence $\mathbf{x}_{\mathrm{ch}}$, and the parity sequence $\mathbf{x}_c$. Here the code rate of 3D-PC is calculated by $R = K/N_T = 1/3$. In order to achieve higher code rate, it is need to puncture some parity bits from $\mathbf{x}_{\mathrm{ch}}$ or $\mathbf{x}_c$. Since $\mathbf{x}_c$ contains more information, $\mathbf{x}_{\mathrm{ch}}$ is first taken into consideration.

For complexity and performance reasons, the selected $C_c$ encoder should meet some requirements: its decoder is as simple as possible, its decoder inputs soft information and outputs soft information, and its decoder should not

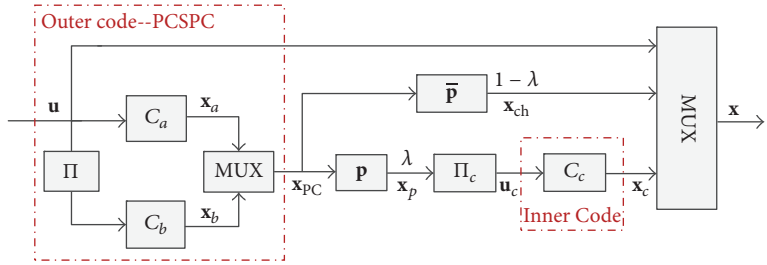



FIGURE 1: 3D polar encoder.

introduce too much error [5]. As a result, a rate-1 cyclic recursive systematic convolutional encoder with generator polynomial $g(D) = 1/(1 + D^2)$ is selected as the encoder $C_c$ [6].

In literature [5, 6], the interleavers Π and $\Pi_c$ have been well designed to increase the minimum distance. Because the design of interleaver has a great influence on the performance of TC. While the effect of interleaver on polar code is not so obvious, random interleaver is considered in this 3D polar encoding structure for convenience.

In this paper, regular puncturing pattern is applied to $\mathbf{p}$. If $\mathbf{p}$ is adopted to $\mathbf{x}_{\mathrm{PC}}$ with length $N_c$, there are altogether $N_c\lambda$ ones in the period $N_c$. The bits of $\mathbf{x}_{\mathrm{PC}}$ corresponding to the positions of $N_c\lambda$ ones are not punctured. For example, assume that $\lambda = 1/4$ and $\mathbf{p} = [11000000]$; then every fourth bit of $\mathbf{x}_a$ and $\mathbf{x}_b$ is extracted and sent to $C_c$ for encoding again. According to the relationship between $\mathbf{p}$ and $\overline{\mathbf{p}}$, it is easy to obtain $\overline{\mathbf{p}} = [00111111]$. If we apply $\overline{\mathbf{p}}$ to $\mathbf{x}_{\mathrm{PC}}$, then the bits which are reserved are sent to the channel.

*3.2. Decoding Structure.* In general, a concatenated code can be decoded by the iterative decoding structure. The decoding diagram of 3D-PC is shown in Figure 2. The sequence $\mathbf{y}$ is received from channel $W$ and is demultiplexed into three parts, $\mathbf{y}_u$, $\mathbf{y}_{\mathrm{ch}}$, and $\mathbf{y}_c$. The corresponding channel logarithm likelihood ratios (LLRs) are denoted by $\mathbf{\Psi}_u$, $\mathbf{\Psi}_{\mathrm{ch}}$, and $\mathbf{\Psi}_c$. Later they participate in the subsequent decoding. The decoders ${C_a}^{-1}$, ${C_b}^{-1}$, and ${C_c}^{-1}$ are corresponding to encoders $C_a$, $C_b$, and $C_c$, respectively.

First, $\mathbf{\Psi}_c$ from channel and $\mathbf{\Gamma}_c$ from ${C_a}^{-1}$ and ${C_b}^{-1}$ are fed to ${C_c}^{-1}$ for decoding. Then the extrinsic information $\mathbf{\Lambda}_c$ is deinterleaved, combined with $\mathbf{\Psi}_{\mathrm{ch}}$ and demultiplexed into two parts, $\mathbf{\Psi}_a$ and $\mathbf{\Psi}_b$. The obtained $\mathbf{\Psi}_{\mathrm{a}}$ and $\mathbf{\Psi}_b$ are regarded as channel LLRs of parity bits and assist ${C_a}^{-1}$ and ${C_b}^{-1}$ in decoding, respectively. For outer decoder, the extrinsic information related to $\mathbf{u}$ is exchanged between ${C_a}^{-1}$ and ${C_b}^{-1}$ because both the input information of $C_a$ and that of $C_b$ are from $\mathbf{u}$. Additionally, the extrinsic information, $\mathbf{\Xi}_a$ and $\mathbf{\Xi}_b$, of parity bits which is output by ${C_a}^{-1}$ and ${C_b}^{-1}$ goes through the following operations: multiplex, puncture, and interleave. Then extrinsic LLR information $\mathbf{\Gamma}_c$ is obtained and delivered to ${C_c}^{-1}$ as a priori information at next iteration. The extrinsic LLR information of part parity bits $\mathbf{x}_p$ is exchanged between inner decoder ${C_c}^{-1}$ and outer decoder as framed in Figure 2. The exchange procedure is terminated when the given out-loop iteration number is reached and the decision is made by the LLR information of ${C_b}^{-1}$.

Since it is needed to exchange extrinsic information between ${C_a}^{-1}$ and ${C_b}^{-1}$, the decoder adopted should meet the soft-in-soft-out (SISO) requirement. As to the decoding of SPC, there are two SISO decoding algorithms, belief propagation (BP) decoding [12] and soft cancellation (SCAN) decoding [13]. Therefore, BP decoder and SCAN decoder can be considered for the decoders ${C_a}^{-1}$ and ${C_b}^{-1}$.

As to the decoding of tail-biting convolutional code, the optimal algorithm is maximum a posteriori probability (MAP) decoding algorithm, but its complexity is very high. Two suboptimal MAP decoding algorithms have been proposed for tail-biting convolutional code, tail-biting BCJR (TB-BCJR), and A3 [14]. Afterwards, a less complexity MAP algorithm has been presented to decode tail-biting convolutional code [15]. Therefore, the TB-BCJR, A3 algorithms and the low complexity MAP algorithm can be chosen as the candidate schemes for ${C_c}^{-1}$ decoder.

## 4. Convergence Behavior Analysis

In this part, EXIT chart is utilized to analyze the convergence threshold of 3D-PC. In Figure 3, the simplified decoding structure for the calculation of EXIT chart is given. In Figure 3, $I_{A,\mathrm{inner}}$ denotes the average mutual information between $A(\mathbf{u}_c)$ and $\mathbf{u}_c$, $I_{E,\mathrm{inner}}$ denotes the average mutual information between $E(\mathbf{u}_c)$ and $\mathbf{u}_c$, $I_{A,\mathrm{outer}}$ denotes the average mutual information between $A(\mathbf{x}_p)$ and $\mathbf{x}_p$, and $I_{E,\mathrm{outer}}$ denotes the average mutual information between $E(\mathbf{x}_p)$ and $\mathbf{x}_p$. The detailed calculation processes of EXIT chart curve are presented as follows:

(1) Given signal to noise ratio (SNR), $\mathbf{u}_c$ and $0 \le I_{A,\mathrm{inner}} \le 1$; then the a priori information $A(\mathbf{u}_c)$ can be obtained by the assumed model [9] and is sent for the inner decoder ${C_c}^{-1}$.

(2) Monte Carlo simulation based on ${C_c}^{-1}$ is performed to get the distributions of $p_E$ of (10).

(3) Then $I_{E,\mathrm{inner}}$ is calculated by substituting $p_E$ into (10).

(4) Traverse $I_{A,\mathrm{inner}}$ at a certain step size in a certain internal $[0, 1]$ and calculate the corresponding $I_{E,\mathrm{inner}}$. Then the curve which depicts the relation between $I_{E,\mathrm{inner}}$ and $I_{A,\mathrm{inner}}$ is obtained.

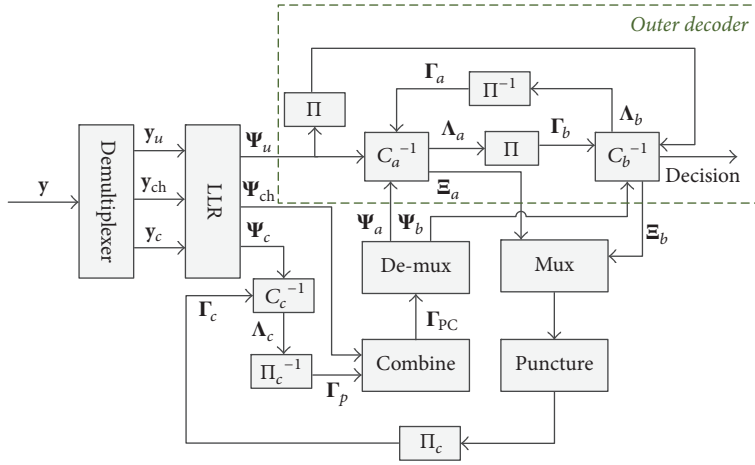



FIGURE 2: The iterative decoding structure of 3D-PC.

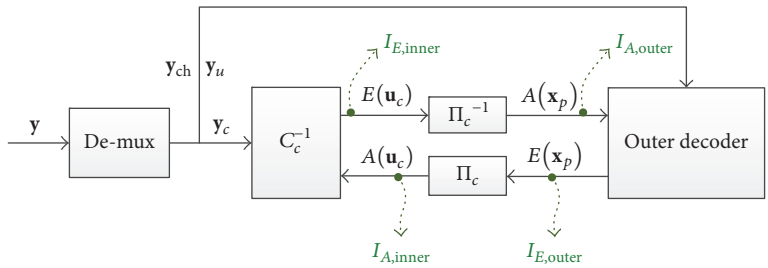



FIGURE 3: Simplified decoding structure for EXIT chart analysis.

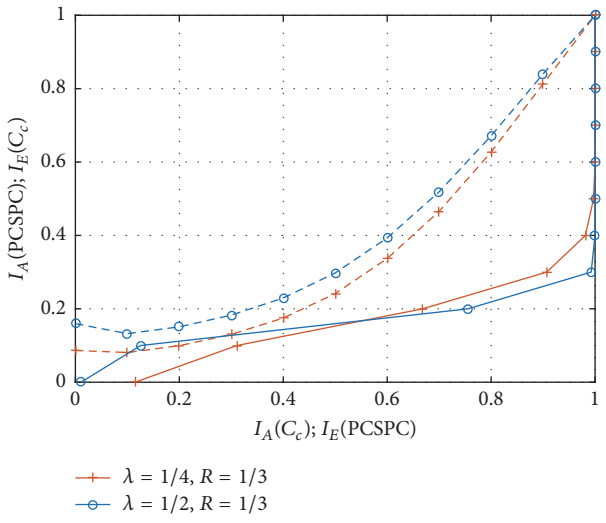



FIGURE 4: EXIT chart of $\lambda = 1/4$, $R = 1/3$, 3D-PC, SNR = 3.41 and of $\lambda = 1/2$, $R = 1/3$, 3D-PC, SNR = 4.8.

Likewise, $I_{E,\text{outer}}$ can be got by the above processes. The differences are that the decoder for Monte Carlo simulation is outer decoder other than $C_c^{-1}$, $I_{A,\text{outer}}$ is given, and the transmitted bits are $\mathbf{x}_p$ instead of $\mathbf{u}_c$.

Figure 4 gives the EXIT chart of 3D-PC with two configurations, $\{\lambda = 1/4, R = 1/3, \text{SNR} = 3.41\}$ and $\{\lambda = 1/2, R = 1/3, \text{SNR} = 4.8\}$. The EXIT chart curves of the outer code and inner code are denoted by solid curves and dash curves, respectively. From Figure 4, it can be seen that there is an opening between the EXIT chart curves of inner code and outer code for both configurations. Since there is no disjoint for each pair of EXIT chart curves, the decoding of 3D-PC can reach convergence. In general, the EXIT chart curves can be depicted with the variety of SNR. The convergence threshold is the SNR at which the tunnel between EXIT chart curves pairs is very narrow. As to 3D-PC with $\lambda = 1/4$ and $R = 1/3$, the convergence threshold is 3.4 dB. Table 1 lists the convergence thresholds of 3D-PC under different $\lambda$. The simulation frames for Monte Carlo simulation are $1.0 \times 10^4$.

TABLE 1: Convergence thresholds of 3D-PC.

| $\lambda$ | $\lambda = 1$ | $\lambda = 1/2$ | $\lambda = 1/4$ | $\lambda = 1/8$ | $\lambda = 0$ |
|---|---|---|---|---|---|
| Thresholds | 6.60 dB | 4.60 dB | 3.40 dB | 3.16 dB | 2.60 dB |

From Table 1, it can be observed that the convergence threshold increases with the increase of $\lambda$. Compared with the best convergence threshold when $\lambda$ is 0, the convergence loss under $\lambda = 1/8$ and $\lambda = 1/4$ is relatively small. Therefore, those two $\lambda$ configurations are set to 3D-PC.

## 5. Simulation Results

In Figure 5, the BLER performance of 3D-PC is given. The underlying channel is additive white Gaussian noise (AWGN)

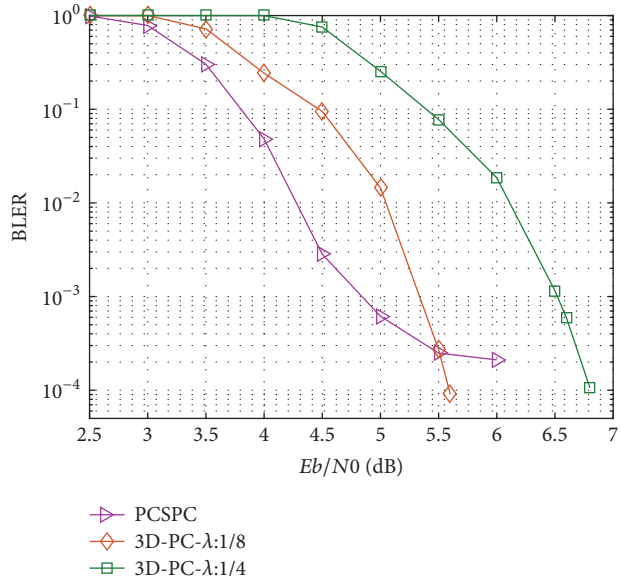



Figure 5: Performance comparison between PCSPC and 3D-PC. For both PCSPC and 3D-PC, the input information length and code rate of them are $K = 4096$ and $R = 1/3$, respectively.

channel. The input block size is set to $K = 4096$. The code rate of the component SPC is 1/2. However, it is noteworthy that the output of the component SPC is $K$ parity bits. And the total code rate of 3D-PC is $R = 1/3$. The interleavers $\Pi$ and $\Pi_c$ used for simulation are random interleavers. The internal iteration number of outer decoder is 1 and the iteration number between the inner decoder and the outer decoder is equal to 6. In addition, SCAN decoding algorithm is utilized for the decoding of SPC and the CRSC code is decoded by low complexity MAP decoding [15]. Different permeability rates $\lambda$ are set to 3D-PC scheme, such as 1/4 and 1/8.

As a comparison scheme, the performance of PCSPC is also given in Figure 5. The constituent codes are SPCs with code rate 1/2 and code length 8192. Under this configuration, the total code rate of PCSPC is 1/3 which is the same as that of 3D-PC. The SCAN decoding algorithm is applied to decode the component codes. For fair comparison, total iteration numbers between the PCSPC component decoders are required to set the same for both the conventional PCSPC and the proposed 3D-PC scheme. Thereafter the outer loop number between the two constituent decoders is equal to 6.

By observing Figure 5, it can be found that the performance in water region is lost for 3D-PC with respect to PCSPC. This phenomenon is accordant with the analysis in Section 4. That is, the convergence threshold becomes larger with the increase of $\lambda$. In addition, 3D-PC has better BLER performance than PCSPC in low error rate. For PCSPC, error floor phenomenon begins at about BLER $2 \times 10^{-4}$. However, the error floor does not appear around BLER $2 \times 10^{-4}$ for 3D-PC. In other words, the error floor is lowered by the proposed 3D-PC scheme. The reason may be that 3D-PC has lager minimum distance compared with PCSPC.

In addition to performance, complexity is also important. As to the conventional PCSPC [1], the computation complexity is written as

$$\Theta_P = O\left(2tN \log N\right), \tag{11}$$

where $t$ is iteration number between the component decoders and $N$ is the code length of component systematic polar code. For the proposed 3D scheme, it includes not only the complexity of PCSPC decoder, but also the complexity of tail-biting convolutional code decoder [15]. Comprehensively, the complexity is about

$$\Theta_{3D} = O\left(T\left(2N \log N + 4 \times 2^m \times \lambda N\right)\right), \tag{12}$$

where $T$ is the out-loop iteration number, $m$ is the memory element of tail-biting convolutional code, $N$ is the code length of component polar code, and $\lambda$ is the permeability rate. In (12), $2N \log N$ and $4 \times 2^m \times \lambda N$ denote the complexity of outer decoder and inner decoder in one outer iteration, respectively. Since the inner iteration number between the PCSPC component decoders is 1, the complexity of outer decoder is $2N \log N$ according to (11). As to Log-MAP algorithm, the complexity can be regarded as the metric updates in the trellis nodes. Corresponding to (12), 4 denotes the metric updates per trellis node, $2^m$ is the state numbers, and $\lambda N$ denotes the input information length of tail-biting convolutional code which can be known from 3D polar encoder (refer to Section 3).

In this paper, $t$ and $T$ are set the same to ensure that the total iteration number between the PCSPC component decoders is the same. Moreover, the increased complexity is $4 \times 2^m \times T\lambda N$ which is brought by inner decoder. Since the memory of the tail-biting convolutional code we use is small and $0 \leq \lambda \leq 1$, the additional complexity of the proposed scheme is less compared with the complexity of the conventional PCSPC decoder. Here, we adopt the parameter configurations in this paper to give a specific example. Assume that $N = 8192$, $t = T = 6$, $m = 2$, and $\lambda = 1/4$; then $\Theta_P = 1277952$ and $\Theta_{3D} = 1474560$ are obtained by (11) and (12). Hence, compared to the complexity of the original PCSPC, the additional complexity of 3D polar code is about 15%.

## 6. Conclusion

In this paper, 3D-PC is presented to lower the error floor of PCSPC. It makes the best use of the characteristics of parallel concatenation and serial concatenation. The simulation results verify the effectiveness of 3D-PC. In addition, EXIT chart is utilized to analyze the convergence threshold of 3D-PC under different permeability rate configurations. The obtained convergence thresholds can guide the choice of permeability rate of 3D-PC.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 61771066), the National Natural Science Foundation of China (no. 61671080), and the National Science and Technology Major Project (no. 2017ZX03001004).

## References

[1] D. Wu, A. Liu, Y. Zhang, and Q. Zhang, "Parallel concatenated systematic polar codes," *IEEE Electronics Letters*, vol. 52, no. 1, pp. 43–45, 2016.

[2] E. Arikan, "Systematic polar coding," *IEEE Communications Letters*, vol. 15, no. 8, pp. 860–862, 2011.

[3] Q. Zhang, A. Liu, Y. Zhang, and X. Liang, "Practical Design and Decoding of Parallel Concatenated Structure for Systematic Polar Codes," *IEEE Transactions on Communications*, vol. 64, no. 2, pp. 456–466, 2016.

[4] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and encoding: turbo-codes," in *Proceedings of the IEEE International Conference on Communications*, pp. 1064–1070, Geneve, Switzerland, May 1993.

[5] C. Berrou, A. Graell i Amat, Y. O. C. Mouhamedou, C. Douillard, and Y. Saouter, "Adding a rate-1 third dimension to turbo codes," in *Proceedings of the 2007 IEEE Information Theory Workshop, ITW 2007*, pp. 156–161, USA, September 2007.

[6] C. Berrou, A. Graell i Amat, Y. Ould-Cheikh-Mouhamedou, and Y. Saouter, "Improving the distance properties of turbo codes using a third component code: 3D turbo codes," *IEEE Transactions on Communications*, vol. 57, no. 9, pp. 2505–2509, 2009.

[7] E. Rosnes and A. Graell i Amat, "Performance analysis of 3-D turbo codes," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 57, no. 6, pp. 3707–3720, 2011.

[8] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: performance analysis, design, and iterative decoding," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 44, no. 3, pp. 909–926, 1998.

[9] S. T. Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Transactions on Communications*, vol. 49, no. 10, pp. 1727–1737, 2001.

[10] E. Arıkan, "Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[11] P. Trifonov, "Efficient design and decoding of polar codes," *IEEE Transactions on Communications*, vol. 60, no. 11, pp. 3221–3227, 2012.

[12] E. Arikan, "A performance comparison of polar codes and reed-muller codes," *IEEE Communications Letters*, vol. 12, no. 6, pp. 447–449, 2008.

[13] U. U. Fayyaz and J. R. Barry, "Low-complexity soft-output decoding of polar codes," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 5, pp. 958–966, 2014.

[14] J. B. Anderson and S. M. Hladik, "Tailbiting MAP decoders," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 2, pp. 297–302, 1998.

[15] P. Wijesinghe, U. Gunawardana, and R. Liyanapathirana, "Low complexity MAP decoding of tailbiting convolutional codes," in *Proceedings of the 2010 International Conference on Signal Processing and Communications, SPCOM 2010*, India, July 2010.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 4397671, 14 pages
https://doi.org/10.1155/2018/4397671

*Research Article*

# Construction and Decoding of Rate-Compatible Globally Coupled LDPC Codes

**Ji Zhang**[1,2], **Baoming Bai**[1], **Xijin Mu**[1], **Hengzhou Xu**[3], **Zhen Liu**[1] and **Huaan Li**[1]

[1] *State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an, China*
[2] *School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang, China*
[3] *School of Network Engineering, Zhoukou Normal University, Zhoukou, China*

Correspondence should be addressed to Baoming Bai; bmbai@mail.xidian.edu.cn

Received 24 November 2017; Revised 2 February 2018; Accepted 28 February 2018; Published 2 May 2018

Academic Editor: Zesong Fei



This paper presents a family of rate-compatible (RC) globally coupled low-density parity-check (GC-LDPC) codes, which is constructed by combining algebraic construction method and graph extension. Specifically, the highest rate code is constructed using the algebraic method and the codes of lower rates are formed by successively extending the graph of the higher rate codes. The proposed rate-compatible codes provide more flexibility in code rate and guarantee the structural property of algebraic construction. It is confirmed, by numerical simulations over the AWGN channel, that the proposed codes have better performances than their counterpart GC-LDPC codes formed by the classical method and exhibit an approximately uniform gap to the capacity over a wide range of rates. Furthermore, a *modified two-phase local/global iterative decoding scheme* for GC-LDPC codes is proposed. Numerical results show that the proposed decoding scheme can reduce the unnecessary cost of local decoder at low and moderate SNRs, without any increase in the number of decoding iterations in the global decoder at high SNRs.

## 1. Introduction

Globally coupled low-density parity-check (GC-LDPC) codes, which were proposed by Li et al. in [1–4], are a special type of LDPC codes designed for correcting random symbol errors and bursts of errors or erasures. They have a different structure from the conventional LDPC block codes and the spatially coupled LDPC (SC-LDPC) codes [5–12]. From the perspective of the Tanner graph, a GC-LDPC code using a group of *global* check nodes (CNs) couples (or connects) a set of disjoint Tanner graphs called *local* graphs. We refer to such codes as CN-based globally coupled LDPC (CN-GC-LDPC) codes. Due to the special structure, CN-GC-LDPC codes not only perform well on both the additive white Gaussian noise (AWGN) channel and the binary erasure channel (BEC) but also are effective for correcting burst erasures.

For time-varying channels, from the wireless communication theory, we need to adapt the rate according to the available channel state information (CSI); such an error control strategy is referred to as rate adaptability [13]. Rate-compatible (RC) channel codes with incremental redundancy are often used in conjunction with the HARQ strategy [14–20]. Most recently, RC-LDPC codes have been adopted by the 3rd Generation Partnership Project (3GPP) as the channel coding scheme for 5G enhanced mobile broadband (eMBB) data channel [21]. Such codes with a wide range of rates and block lengths are a family of nested codes which can be interpreted as a graph extension of high-rate codes [17, 22, 23].

Unlike the SC-LDPC codes which have some conventional design methods for RC-LDPC codes, such as puncturing variable nodes from the codes with low rate and extending variable nodes to the codes with high rate, the classical GC-LDPC codes are mostly restricted to invariant code rate [11, 20, 22–25]. And the existing construction methods of GC-LDPC codes are not flexible enough for RC code design. In this paper, we present a family of RC CN-GC-LDPC codes. The proposed construction is based on combining algebraic construction method and graph extension. The highest rate code, which can be called the mother code, is constructed using the algebraic method. And the codes of lower rates are formed by successively extending the graph

of the higher rate. Compared to the original CN-GC-LDPC codes, the proposed family of RC CN-GC-LDPC codes not only provides more flexibility in code rate and allows more efficient encoding but also guarantees the structural property of algebraic construction. Examples show that the proposed codes perform well and exhibit an approximately uniform gap to the capacity over a wide range of rates.

Furthermore, localization is a significant advantage of CN-GC-LDPC codes which allows us to process a consecutive sequence by some local sequences. If a decoder uses *two-phase local/global iterative scheme*, its decoding latency and memory requirements are much less than those of the classical decoding scheme without performance degradation [1, 2, 26, 27]. However, the local decoder performs a plenty of useless operations at low and moderate SNRs, which causes the unnecessary cost of the decoder. In light of this case, we present a *modified two-phase local/global iterative decoding scheme* for CN-GC-LDPC codes which reduces the unnecessary cost of the local decoder at low and moderate SNRs.

The remainder of this paper is organized as follows. In Section 2, we give a brief introduction to CN-GC-LDPC codes. In Section 3, we introduce the concept of graph extension through a four-edge-type LDPC code and propose a simple construction method of four-edge-type LDPC code family. In Section 4, we present a *modified two-phase local/global iterative decoding scheme* for CN-GC-LDPC codes. Numerical results for RC GC-LDPC codes are presented in Section 5. Finally, Section 6 concludes this paper.

## 2. Background

*2.1. CN-Based QC-GC-LDPC Codes.* A CN-based QC-GC-LDPC code $\mathscr{C}_{gc}$ is constructed using a base matrix $\mathbf{B}_{gc}$ which consists of two subarrays as shown:

$$\mathbf{B}_{gc} = \left[b_{j,k}\right]_{0\le j<mt+s,0\le k<nt} = \left[\begin{array}{ccccc} \mathbf{B}_0 & & & & \\ & \mathbf{B}_1 & & & \\ & & \ddots & & \\ & & & \mathbf{B}_i & \\ & & & & \ddots \\ & & & & \mathbf{B}_{t-1} \\ \hline \multicolumn{5}{c}{\mathbf{X}_{gc}} \end{array}\right]. \quad (1)$$

The upper subarray of $\mathbf{B}_{gc}$ is a $t \times t$ diagonal array with $m \times n$ matrices $\mathbf{B}_i$ on its main diagonal and the lower part is an $s \times nt$ matrix $\mathbf{X}_{gc}$, where $0 \le i \le t-1$. So, $\mathbf{B}_{gc}$ is an $(mt+s) \times nt$ matrix.

Consider a finite field GF($q$) and let $\alpha$ be a primitive element of GF($q$). Let $\mathbf{H}_{gc,qc}$ denote the parity-check matrix of CN-based QC-GC-LDPC code. We can use the base matrix $\mathbf{B}_{gc}$ and $(q-1)$-fold dispersion to form $\mathbf{H}_{gc,qc}$ as follows: for each entry $b_{j,k}$ in $\mathbf{B}_{gc}$, if $b_{j,k} = 0$, then replace $b_{j,k} = 0$ by the $(q-1) \times (q-1)$ zero matrix (ZM); otherwise, assume $b_{j,k} = \alpha^l$ with $0 \le l < q-1$, and then replace $b_{j,k}$ by a $(q-1) \times (q-1)$ circulant permutation matrix (CPM) whose first row has a single 1-component at the $l$th element. The resultant $\mathbf{H}_{gc,qc}$ is an $(mt+s)(q-1) \times nt(q-1)$ matrix over GF(2). The null space of $\mathbf{H}_{gc,qc}$ gives a CN-based QC-GC-LDPC code.

From the perspective of graph, let $\mathscr{G}_i$ be the Tanner graph associated with $\mathbf{B}_i$ for $0 \le i \le t-1$. We call $\mathscr{G}_l$ a local Tanner graph. The different $\mathscr{G}_i$ are pairwise disjoint (for $0 \le j, k \le t-1$, and $j \ne k$, $\mathscr{G}_j$ and $\mathscr{G}_k$ are disjoint). The $s$ global CNs connect all these local graphs, composing the Tanner graph associated with $\mathbf{B}_{gc}$, denoted as $\mathscr{G}_{gc}$, as shown in Figure 1 (in this figure, the edge labels are ignored). We see that the $s$ global CNs provide the only connections between any two disjoint local Tanner graphs.

*2.2. The Basic Method of Code Construction.* Li et al. presented two types of CN-based QC-GC-LDPC codes and their construction methods in [1, 2, 4]. In this subsection, we briefly review the construction method for the first type of QC-GC-LDPC codes in [1]. Let GF($q$) be a nonbinary (NB) field with $q$ elements, where $q$ is a power of a prime. Let $\alpha$ be a primitive element of GF($q$). Then, the powers of $\alpha$, namely, $\alpha^0, \alpha^1, \alpha^2, \ldots, \alpha^{q-2}$, give all the nonzero elements of GF($q$) and $\alpha^{q-1} = \alpha^0 = 1$. We construct the following matrix $\mathbf{B}_{rs}$ over GF($q$) using the method in [2]:

$$\mathbf{B}_{rs} = \begin{bmatrix} \alpha^0 - 1 & \alpha - 1 & \cdots & \alpha^{q-2} - 1 \\ \alpha^{q-2} - 1 & \alpha^0 - 1 & \cdots & \alpha^{q-3} - 1 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha - 1 & \alpha^2 - 1 & \cdots & \alpha^0 - 1 \end{bmatrix}. \quad (2)$$

Suppose $q-1$ can be factored as the product of two positive integers $l$ and $r$; that is, $q - 1 = lr$. The matrix $\mathbf{B}_{rs}$ which has a cyclic structure is then an $lr \times lr$ square matrix. The first $l$ rows in $\mathbf{B}_{rs}$ comprise an $l \times lr$ matrix, denoted by $\mathbf{W}_0$. Then, $\mathbf{W}_0$ can be partitioned into $r$ submatrices of size $l \times l$, denoted by $\mathbf{W}_{0,0}, \mathbf{W}_{0,1}, \ldots, \mathbf{W}_{0,r-1}$. Based on the cyclic structure of $\mathbf{B}_{rs}$, we can obtain the next $l$ rows in $\mathbf{B}_{rs}$ by cyclically shifting all the rows of $\mathbf{W}_0 = [\mathbf{W}_{0,0}, \mathbf{W}_{0,1}, \ldots, \mathbf{W}_{0,r-1}]$ $l$ positions to right, denoted by $\mathbf{W}_1 = [\mathbf{W}_{0,r-1}, \mathbf{W}_{0,0}, \ldots, \mathbf{W}_{0,r-3}, \mathbf{W}_{0,r-2}]$. And so, the matrix $\mathbf{B}_{rs}$ can be partitioned into $r$ submatrices of size $l \times lr$, denoted by $\mathbf{W}_0, \mathbf{W}_1, \ldots, \mathbf{W}_i, \ldots, \mathbf{W}_{r-1}$, where $\mathbf{W}_i = [\mathbf{W}_{0,r-i}, \ldots, \mathbf{W}_{0,r-1}, \mathbf{W}_{0,0}, \ldots, \mathbf{W}_{0,r-i-1}]$, $0 < i \le r-1$. Hence, we can form the following block-cyclic structured matrix $\mathbf{B}_W$ from $\mathbf{B}_{rs}$:

$$\mathbf{B}_W = \begin{bmatrix} \mathbf{W}_{0,0} & \mathbf{W}_{0,1} & \cdots & \mathbf{W}_{0,r-1} \\ \mathbf{W}_{0,r-1} & \mathbf{W}_{0,0} & \cdots & \mathbf{W}_{0,r-2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{W}_{0,1} & \mathbf{W}_{0,2} & \cdots & \mathbf{W}_{0,0} \end{bmatrix}. \quad (3)$$

For $0 \le j < r$ and $1 \le m, n \le l$, we take an $m \times n$ submatrix $\mathbf{R}_{0,j}$ from $\mathbf{W}_{0,j}$. For the submatrix $\mathbf{R}_{0,j}$, we restrict the locations of its entries in $\mathbf{W}_{0,j}$ to be identical (i.e., for $0 \le j_1, j_2 < r$, the entries of $\mathbf{R}_{0,j_1}$ and $\mathbf{R}_{0,j_2}$ are taken from

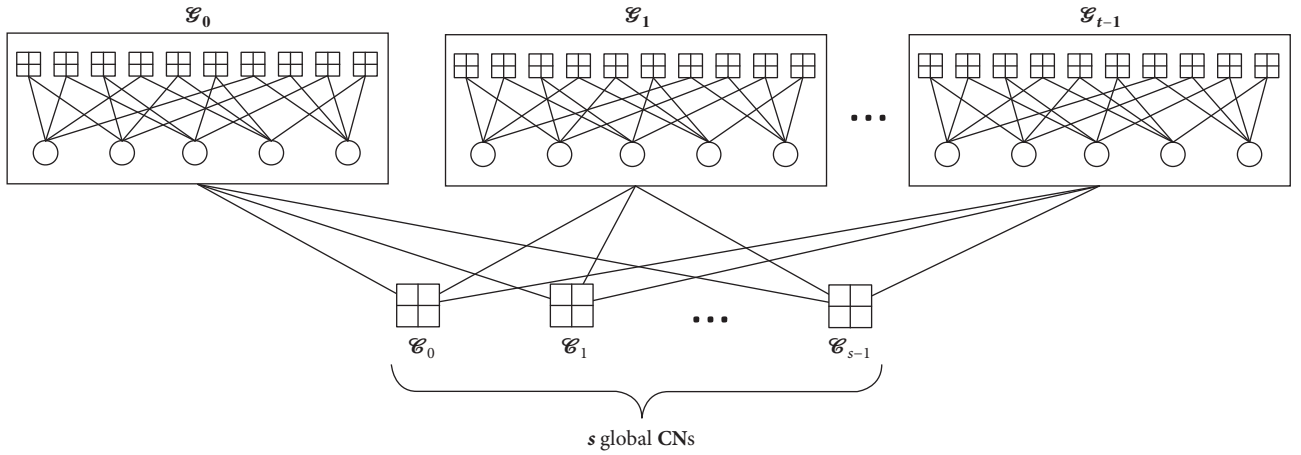



Figure 1: The Tanner graph of CN-based QC-GC-LDPC codes.

identical locations in $\mathbf{W}_{0,j_1}$ and $\mathbf{W}_{0,j_2}$). Then, we form the following $r \times r$ array $\mathbf{B}_R$ of $m \times n$ submatrices over GF($q$) with a block-cyclic structure:

$$\mathbf{B}_R = \begin{bmatrix} \mathbf{R}_{0,0} & \mathbf{R}_{0,1} & \cdots & \mathbf{R}_{0,r-1} \\ \mathbf{R}_{0,r-1} & \mathbf{R}_{0,0} & \cdots & \mathbf{R}_{0,r-2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{0,1} & \mathbf{R}_{0,2} & \cdots & \mathbf{R}_{0,0} \end{bmatrix}. \quad (4)$$

We denote the set of rows in the first row blocks of $\mathbf{B}_R$ by $\Gamma = [\mathbf{R}_{0,0}, \mathbf{R}_{0,1}, \ldots, \mathbf{R}_{0,r-1}]$. So, $\mathbf{B}_R$ is a submatrix of $\mathbf{B}_W$. In forming the $r \times r$ array $\mathbf{B}_R$ of $m \times n$ submatrices over GF($q$) given by (4), there are $l-m$ rows in each row-block and $(l-n)$ columns in each column-block of $\mathbf{B}_W$ which are unused. We denote the set of $r(l-m)$ unused rows of $\mathbf{B}_W$ in each row-block by $\mathbf{\Pi_1}$. So, there are $r$ sections which have a total of $l$ components in each row of $\mathbf{\Pi_1}$. For each section in row, we remove $l-n$ components which are not used in forming the array $\mathbf{B}_R$ from $\mathbf{B}_W$. We denote the set of $r(l-m)$ rows by $\mathbf{\Pi_1^*}$. The rows in set $\mathbf{\Pi_1^*}$ and the rows in set $\Gamma$ are disjoint. Let $s$ and $t$ be two positive integers with $1 \le s \le r(l-m)$ and $1 \le t \le r$. We remove the last $r-t$ sections from $\mathbf{\Pi_1^*}$ and take $s$ rows from the remaining sections of $\mathbf{\Pi_1^*}$ to form an $s \times nt$ matrix $\mathbf{X}_{gc}$ over GF($q$). By taking the $t \times t$ diagonal array from the main diagonal of $\mathbf{B}_R$ and appending the matrix $\mathbf{X}_{gc}$ to the bottom of them, we form the following base matrix of a QC-GC-LDPC code:

$$\mathbf{B}_{gc,rs} = \begin{bmatrix} \mathbf{R}_{0,0} & & & \\ & \mathbf{R}_{0,0} & & \\ & & \ddots & \\ & & & \mathbf{R}_{0,0} \\ \hline \multicolumn{4}{c}{\mathbf{X}_{gc}} \end{bmatrix}. \quad (5)$$

Each entry in the upper subarray of $\mathbf{B}_{gc,rs}$ is equal to $\mathbf{R}_{0,0}$ which is an $m \times n$ matrix, and such subarray is called *local part*. Note that we can use $t$ different member matrices in the set $\{\mathbf{R}_{0,0}, \mathbf{R}_{0,1}, \ldots, \mathbf{R}_{0,r-1}\}$ as the matrices on the main diagonal of $t \times t$ upper subarray of $\mathbf{B}_{gc,rs}$ as well. And we call the lower subarray of $\mathbf{B}_{gc,rs}$ *global part*. The $(q-1)$-fold dispersion of $\mathbf{B}_{gc,rs}$ results in an $(mt+s) \times nt$ array $\mathbf{H}_{gc}$ of $(q-1) \times (q-1)$ CPMs and/or ZMs. The null space of $\mathbf{H}_{gc}$ gives a CN-based QC-GC-LDPC code whose Tanner graph has a girth of at least 6.

Let $R_{gc}$ be the design rate of a $(\overline{d}_v, \underline{d}_v; \overline{d}_c, \underline{d}_c)$ binary CN-GC-LDPC code, where $\overline{d}_v$ and $\underline{d}_v$ denote the VNs (variable nodes) degrees of *local part* and *global part*, respectively, and $\overline{d}_c$ and $\underline{d}_c$ denote the CNs degrees of *local part* and *global part*, respectively. Then, $R_{gc} = 1 - (\overline{d}_v + \underline{d}_v)(tm + s)/(tm\overline{d}_c + s\underline{d}_c)$, $1 \le \overline{d}_v \le m$, $1 \le \underline{d}_v \le s$, $1 \le \overline{d}_c \le n$, and $1 \le \underline{d}_c \le nt$. For $\overline{d}_v = m$, $\underline{d}_v = s$, $\overline{d}_c = n$, and $\underline{d}_c = nt$, $R_{gc}$ is equal to $1 - m/n - s/nt$.

*Example 1.* Consider the prime field GF(127) for the code construction. Suppose we choose two sets of parameters, $l = 42$, $r = 3$, $m = 5$, $n = 40$, $t = 3$, $s = 1$ and $l = 21$, $r = 6$, $m = 3$, $n = 21$, $t = 6$, $s = 1$. Based on the construction method described above, we form two binary matrices of size $2016 \times 15120$ and $2394 \times 15876$, respectively. We denote those two matrices as $\mathbf{H}_{gc,1}(126, 126)$ and $\mathbf{H}_{gc,2}(126, 126)$. For $\mathbf{H}_{gc,1}(126, 126)$, it has two column weights, 5 and 6, and two row weights, 39 and 120. The null space of $\mathbf{H}_{gc,1}(126, 126)$ gives (15120, 13104) CN-based QC-GC-LDPC code $\mathscr{C}_1$ with a rate of 0.8667. The Tanner graph of $\mathscr{C}_1$ contains 3,165,498 cycles of length 6 and 545,198,472 cycles of length 8. For $\mathbf{H}_{gc,2}(126, 126)$, it has two column weights, 3 and 4, and two row weights, 21 and 125. The null space of $\mathbf{H}_{gc,2}(126, 126)$ gives (15876, 13494) CN-based QC-GC-LDPC code $\mathscr{C}_2$ with a rate of 0.85. The Tanner graph of $\mathscr{C}_2$ contains 198,828 cycles of length 6 and 21,715,639 cycles of length 8.

## 3. Rate-Compatible GC-LDPC Codes

In this section, we first introduce the concept of graph extension through a four-edge-type LDPC code. Then, we present a construction method of RC GC-LDPC codes.

*3.1. Four-Edge-Type LDPC Codes.* The Tanner graph of a four-edge-type LDPC code is illustrated in Figure 2. We partition

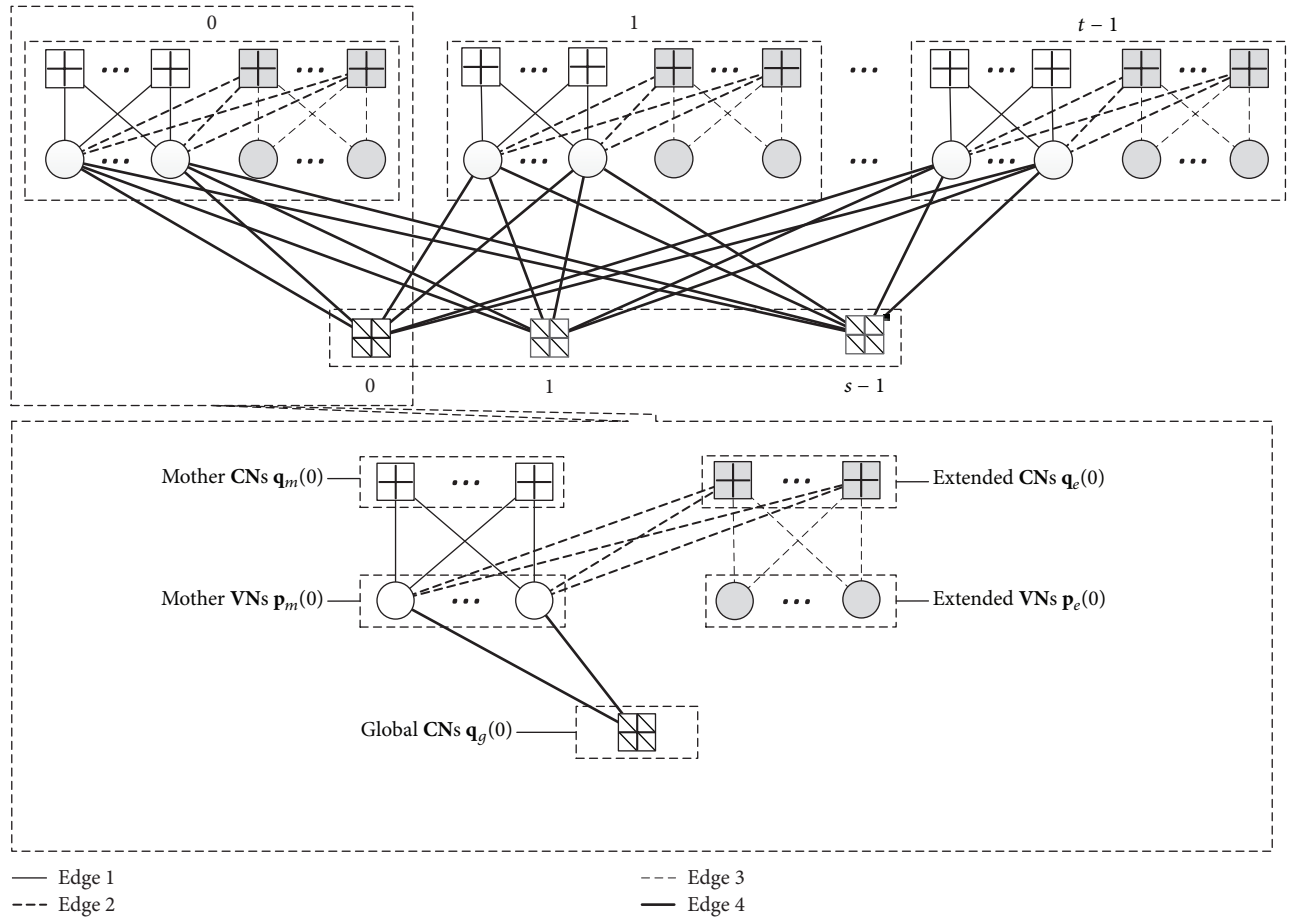



Figure 2: The Tanner graph of four-edge-type LDPC codes.

the VNs into $t$ sections, denoted by $\mathbf{p}(0), \mathbf{p}(1), \ldots, \mathbf{p}(t-1)$, each containing $N_s$ VNs. Every $\mathbf{p}(i)$ is then split into two parts: the mother VNs $\mathbf{p}_m(i)$ of length $N_m(i)$ and the extended VNs $\mathbf{p}_e(i)$ of length $N_e(i)$; that is, $\mathbf{p}(i) = [\mathbf{p}_m(i), \mathbf{p}_e(i)]$. We denote the corresponding coded symbols of VNs as a vector $\mathbf{v}$ of size $N$. At the $i$th section, we denote the corresponding symbol vectors of $\mathbf{p}_m(i)$ and $\mathbf{p}_e(i)$ as $\mathbf{v}_m(i) = [v_{m,j}(i)]_{0 \le j < N_m(i)}$ and $\mathbf{v}_e(i) = [v_{e,j}(i)]_{0 \le j < N_e(i)}$, respectively. Then, we denote by $\mathbf{q}_g$ the $M_g$ CNs in the lower part which are called global CNs. In the upper part, we partition the CNs into $t$ sections, each containing $M_s$ CNs, denoted by $\mathbf{q}(0), \mathbf{q}(1), \ldots, \mathbf{q}(t-1)$. We split each section of such CNs into two parts: the mother CNs $\mathbf{q}_m(i)$ of length $M_m(i)$ and the extended CNs $\mathbf{q}_e(i)$ of length $M_e(i)$, where $0 \le i < t$. We refer to the edges connecting $\mathbf{p}_m(i)$ and $\mathbf{q}_m(i)$ as the type 1 edges, the edges connecting $\mathbf{p}_m(i)$ and $\mathbf{q}_e(i)$ as the type 2 edges, the edges connecting $\mathbf{p}_e(i)$ and $\mathbf{q}_e(i)$ as the type 3 edges, and the edges connecting $\mathbf{p}_m(i)$ and $\mathbf{q}_g$ as the type 4 edges.

From Figure 2, we can see that the mother VN not only is connected to the global CNs by the type 4 edges but also is connected to the mother CNs and the extended CNs by the type 1 edges and the type 2 edges, respectively. However, the extended VNs which have the same number of the extended CNs only are connected to the extended CNs by the type 3 edges. This means that we can form GC-LDPC codes by connecting the mother CNs and the global CNs to the mother VNs in the graph. By extending the new nodes and then increasing new edges of type 2 and type 3, we can continuously form the new GC-LDPC codes. Note that the type 2 edges are the only edges connecting the mother nodes and the extended nodes, so they are quite different from the type 1 edges. Since there are four different types of edges in the Tanner graph, we refer to such LDPC codes as *four-edge-type LDPC codes*.

For $0 \le i < t$, the submatrices which correspond to the four types of edges are denoted by

$$\begin{aligned}
\mathbf{H}_m(i) &= \left[h_{m,j,k}(i)\right]_{0 \le j < M_m(i), 0 \le k < N_m(i)}, \\
\mathbf{H}_{m \to e}(i) &= \left[h_{m \to e,j,k}(i)\right]_{0 \le j < M_e(i), 0 \le k < N_m(i)}, \\
\mathbf{H}_e(i) &= \left[h_{e,j,k}(i)\right]_{0 \le j < M_e(i), 0 \le k < N_e(i)}, \\
\mathbf{H}_{m \to g}(i) &= \left[h_{m \to g,j,k}(i)\right]_{0 \le j < M_g, 0 \le k < N_m(i)}.
\end{aligned} \tag{6}$$

Then, the parity-check matrix $\mathbf{H}_{\mathrm{FET}}$ of four-edge-type LDPC codes can be represented by

$$\mathbf{H}_{\mathrm{FET}} = \left[\begin{array}{cccccccc} \mathbf{H}_m(0) & \mathbf{0} & & & & & & \\ \mathbf{H}_{m\to e}(0) & \mathbf{H}_e(0) & & & & & & \\ & & \mathbf{H}_m(1) & \mathbf{0} & & & & \\ & & \mathbf{H}_{m\to e}(1) & \mathbf{H}_e(1) & & & & \\ & & & & \ddots & & & \\ & & & & & \mathbf{H}_m(t-1) & \mathbf{0} & \\ & & & & & \mathbf{H}_{m\to e}(t-1) & \mathbf{H}_e(t-1) & \\ \hline \mathbf{H}_{m\to g}(0) & \mathbf{0} & \mathbf{H}_{m\to g}(1) & \mathbf{0} & \cdots & \mathbf{H}_{m\to g}(t-1) & \mathbf{0} & \end{array}\right]. \tag{7}$$

Analogous to GC-LDPC codes, we call the upper subarray *local part* and the lower subarray *global part*. We extract $\mathbf{H}_m(i)$, $\mathbf{H}_{m\to e}(i)$, and $\mathbf{H}_e(i)$ from $\mathbf{H}_{\mathrm{FET}}$ to form an $M_{\mathrm{local}}(i) \times N_{\mathrm{local}}(i)$ matrix over GF($q$) as follows:

$$\mathbf{H}_{\mathrm{local}}(i) = \begin{bmatrix} \mathbf{H}_m(i) & \mathbf{0} \\ \mathbf{H}_{m\to e}(i) & \mathbf{H}_e(i) \end{bmatrix}, \tag{8}$$

where $\mathbf{0}$ is an all-zero matrix, $M_{\mathrm{local}}(i) = M_m(i) + M_e(i)$, and $N_{\mathrm{local}}(i) = N_m(i) + N_e(i)$ for $0 \le i < t$. So, the corresponding base matrix of $\mathbf{H}_{\mathrm{local}}(i)$ and $\mathbf{H}_{\mathrm{FET}}$ is $\mathbf{B}_i$ and $\mathbf{B}_{gc}$, respectively.

Then, the four-edge-type LDPC code satisfies the following parity-check equation:

$$\begin{aligned} \left[\mathbf{v}_m(i) \;\; \mathbf{v}_e(i)\right] \mathbf{H}_{\mathrm{local}}^T(i) &= \mathbf{0}, \\ \mathbf{v}_m(i)\, \mathbf{H}_{m\to g}^T(i) &= \mathbf{0}. \end{aligned} \tag{9}$$

We denote the LDPC code given by the null space of $\mathbf{H}_{\mathrm{FET}}$ as $\mathscr{C}_{\mathrm{FET}}$ and the LDPC code given by the null space of $\mathbf{H}_{\mathrm{local}}(i)$ as $\mathscr{C}_{\mathrm{local}}(i)$, respectively (where $0 \le i < t$). Notice that $\mathbf{H}_e(i)$ is square and has a full rank for $0 \le i < t$, and the entries above $\mathbf{H}_e(i)$ have to be all zero to achieve rate compatibility; thus, $\mathscr{C}_{\mathrm{local}}(i)$ is a RC-LDPC code [16]. We extract $\mathbf{H}_m(i)$ and $\mathbf{H}_{m\to g}(i)$ from $\mathbf{H}_{\mathrm{FET}}$ to form an $(M_g + \sum_i M_m(i)) \times \sum_i N_m(i)$ matrix $\mathbf{A}$ over GF($q$) as follows:

$$\mathbf{A} = \left[\begin{array}{cccc} \mathbf{H}_m(0) & & & \\ & \mathbf{H}_m(1) & & \\ & & \ddots & \\ & & & \mathbf{H}_m(t-1) \\ \hline \mathbf{H}_{m\to g}(0) & \mathbf{H}_{m\to g}(1) & \cdots & \mathbf{H}_{m\to g}(t-1) \end{array}\right]. \tag{10}$$

Then, based on (8) and (9), we have

$$\left[\mathbf{v}_m(0) \;\; \mathbf{v}_m(1) \;\; \cdots \;\; \mathbf{v}_m(t-1)\right] \mathbf{A}^T = \mathbf{0}. \tag{11}$$

This means that $\mathbf{H}_{\mathrm{FET}}$ is a RC parity-check matrix obtained progressively by extension, in order of decreasing rate [17]. Furthermore, all the mother VNs are connected to $\mathbf{q}_g$ and there are no edges connecting $\mathbf{p}(i)$ and $\mathbf{q}(j)$ for $i \ne j$; thus, the four-edge-type LDPC codes can be viewed as a type of RC CN-based QC-GC-LDPC codes.

*3.2. Construction of Rate-Compatible GC-LDPC Codes.* We now present a method to construct RC GC-LDPC codes. Because the extended VNs are not connected to $\mathbf{q}_g$, the construction of RC GC-LDPC codes can be realized by successively extending the graph of *local part*. The successive extension of the parity-check matrices of *local part* in the $j$th section is shown in Figure 3.

Assume that we intend to construct a family of four-edge-type LDPC codes containing $f$ member codes $\boldsymbol{\mathscr{C}} = \{\mathscr{C}_{\mathrm{FET},0}, \mathscr{C}_{\mathrm{FET},1}, \ldots, \mathscr{C}_{\mathrm{FET},f-1}\}$. The corresponding target rate set $\boldsymbol{\mathscr{R}} = \{R_{\mathrm{FET},0}, R_{\mathrm{FET},1}, \ldots, R_{\mathrm{FET},f-1}\}$ is given. We denote $\mathbf{H}_{\mathrm{FET},i}$ as the $M_{\mathrm{FET},i} \times N_{\mathrm{FET},i}$ parity-check matrix of $\mathscr{C}_{\mathrm{FET},i}$, where $0 \le i < f$. At the *local part* of $\mathbf{H}_{\mathrm{FET},i}$, there are $t$ submatrices on the main diagonal of size $M_{\mathrm{local},i}(j) \times N_{\mathrm{local},i}(j)$ and we denote each submatrix as $\mathbf{H}_{\mathrm{local},i}(j)$, where $0 \le i < f$ and $0 \le j < t$. So, $\mathbf{H}_{\mathrm{local},0}(j)$ is the mother matrix of $\mathbf{H}_{\mathrm{local},1}(j)$. And the code rate of $\mathscr{C}_{\mathrm{FET},0}$ is

$$R_{\mathrm{FET},0} = 1 - \frac{\mathrm{Rank}\left(\mathbf{H}_{\mathrm{FET},0}\right)}{\sum_j N_{\mathrm{local},0}(j)}. \tag{12}$$

Suppose $\mathbf{H}_{\mathrm{FET},0}$ has a full rank:

$$R_{\mathrm{FET},0} = 1 - \frac{M_g + \sum_j M_{\mathrm{local},0}(j)}{\sum_j N_{\mathrm{local},0}(j)}. \tag{13}$$

By adding $M_{e,1}(j)$ new rows and $N_{e,1}(j)$ new columns to mother matrix $\mathbf{H}_{\mathrm{local},0}(j)$, we obtain an extended matrix $\mathbf{H}_{\mathrm{local},1}(j)$. The upper part of $\mathbf{H}_{\mathrm{local},1}(j)$ is composed of a mother matrix $\mathbf{H}_{\mathrm{local},0}(j)$ and an all-zero matrix of size $M_{\mathrm{local},0}(j) \times N_{e,1}(j)$. The lower part of $\mathbf{H}_{\mathrm{local},1}(j)$ is composed of a matrix $\mathbf{H}_{m\to e,1}(j)$ of size $M_{e,1}(j) \times N_{\mathrm{local},0}(j)$ and a matrix $\mathbf{H}_{e,1}(j)$ of size $M_{e,1}(j) \times N_{e,1}(j)$. Thus, the $t$ updated submatrices on the main diagonal of $\mathbf{H}_{\mathrm{FET},0}$ form the *local part* of $\mathbf{H}_{\mathrm{FET},1}$. Then, we partition the *global part* of $\mathbf{H}_{\mathrm{FET},0}$ into $t$ submatrices of size $M_g \times N_{\mathrm{local},0}(j)$ and we denote each submatrix as $\mathbf{H}_{m\to g,0}(j)$, where $0 \le j < t$. By adding $M_g \times N_{e,1}(j)$ all-zero matrix to the right of $\mathbf{H}_{m\to g,0}(j)$, we form an extended matrix $\mathbf{H}_{m\to g,1}(j)$, where $0 \le j < t$. We

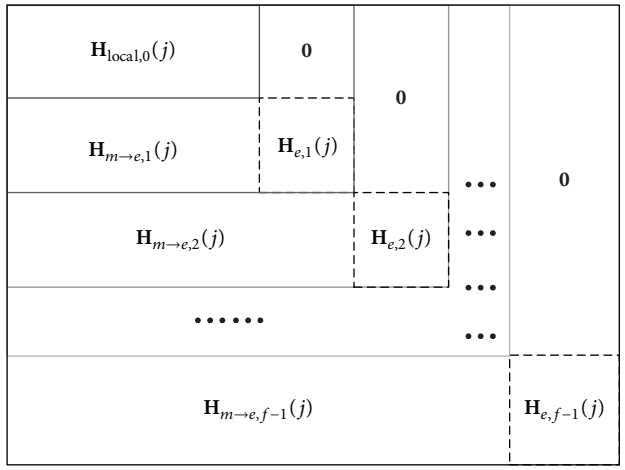



Figure 3: Parity-check matrix extension in the $j$th section of *local part*.

combine the $t$ extended submatrices to compose the *global part* of $\mathbf{H}_{\text{FET},1}$. Suppose $\mathbf{H}_{\text{FET},1}$ has a full rank; we have

$$\begin{aligned} R_{\text{FET},1} &= 1 - \frac{M_g + \sum_j M_{\text{local},1}(j)}{\sum_j N_{\text{local},1}(j)} \\ &= 1 - \frac{M_g + \sum_j M_{\text{local},0}(j) + \sum_j M_{e,1}(j)}{\sum_j N_{\text{local},0}(j) + \sum_j N_{e,1}(j)}. \end{aligned} \tag{14}$$

Recursively, for the code $\mathscr{C}_{\text{FET},i+1}$, we can obtain its parity-check matrix $\mathbf{H}_{\text{FET},i+1}$ from the previously generated parity-check matrix $\mathbf{H}_{\text{FET},i}$ for $\mathscr{C}_{\text{FET},i}$, $0 \le i \le f-2$. Suppose $\mathbf{H}_{\text{FET},i+1}$ has a full rank; the rate of the code $\mathscr{C}_{\text{FET},i+1}$ satisfies

$$\begin{aligned} R_{\text{FET},i+1} &= 1 - \frac{M_g + \sum_j M_{\text{local},i+1}(j)}{\sum_j N_{\text{local},i+1}(j)} \\ &= R_{\text{FET},i} - \frac{\sum_j N_{e,i+1}(j)}{\sum_j N_{\text{local},i}(j)} \cdot \frac{R_{\text{FET},i}}{2 - R_{\text{FET},i}}. \end{aligned} \tag{15}$$

In particular, it is not necessary for $\mathbf{H}_{\text{FET},1}$ to be a full-rank matrix for constructing the RC GC-LDPC codes. Using elementary row and column operations of $\mathbf{H}_{\text{FET},i}$, we have

$$\left[\begin{array}{c|c} \mathbf{H}_{\text{FET},i-1} & 0 \\ \hline \mathbf{H}_{\text{ME},i} & \begin{array}{ccc} \mathbf{H}_{e,i}(0) & & \\ & \mathbf{H}_{e,i}(1) & \\ & & \ddots \\ & & \mathbf{H}_{e,i}(t-1) \end{array} \end{array}\right], \tag{16}$$

where

$$\mathbf{H}_{\text{ME},i} = \begin{bmatrix} \mathbf{H}_{m\to e,i}(0) & & & \\ & \mathbf{H}_{m\to e,i}(1) & & \\ & & \ddots & \\ & & & \mathbf{H}_{m\to e,i}(t-1) \end{bmatrix} \tag{17}$$

and $0 < i \le f-1$. Consider $\mathbf{H}_{e,i}(j)$ as a nonsingular matrix, where $0 \le j \le t-1$; the rank $\text{Rank}(\mathbf{H}_{\text{FET},i})$ of $\mathbf{H}_{\text{FET},i}$ can be written as

$$\begin{aligned} \text{Rank}(\mathbf{H}_{\text{FET},i}) &= \text{Rank}(\mathbf{H}_{\text{FET},i-1}) + \sum_{j=0}^{t-1} M_{e,i}(j) \\ &= \text{Rank}(\mathbf{H}_{\text{FET},0}) + \sum_{k=1}^{i}\sum_{j=0}^{t-1} M_{e,k}(j), \end{aligned} \tag{18}$$

where $0 < i \le f-1$. And it is clear to see that $\mathbf{H}_{\text{FET},i}$ can be a full-rank matrix if and only if $\mathbf{H}_{\text{FET},0}$ is a full-rank matrix. The construction for the parity-check matrix of the RC GC-LDPC codes is described in detail in Algorithm 1.

Furthermore, the RC CN-based QC-GC-LDPC codes with extension structure allow more efficient encoding. Especially for $0 \le j \le t-1$ and $0 \le i \le f-1$, if $\mathbf{H}_{e,i}(j)$ is an identity matrix, the encoding of such RC CN-based QC-GC-LDPC codes is more efficient: after encoding the mother VNs, the encoding of the extended VNs only involves XOR operations on the precode output symbols.

*Example 2.* Consider the prime field GF(193) for code construction. Let $l = 64$, $r = 3$, $m = 5$, $n = 27$, $t = 3$, and $s = 1$. Based on the construction method described in Section 2, we form a $3072 \times 15552$ binary matrix $\mathbf{H}_{gc,3}(192, 192)$. It has one column weight 6, two row weights, 27 and 81. Based on base graph 1 in 5G standard, we construct a masking matrix $\mathbf{D}_{\text{local}}$ of size $960 \times 5184$ for each submatrix on the main diagonal at the local part of $\mathbf{H}_{gc,3}(192, 192)$ [28]. Particularly, the first $2 \times 192$ columns of $\mathbf{D}_{\text{local}}$ are punctured columns. The degree distribution for $\mathbf{D}_{\text{local}}$ is $\lambda(x) = 0.0127 + 0.0759x + 0.7975x^2 + 0.0506x^3 + 0.0633x^4$, $\rho(x) = 0.0380x^2 + 0.962x^{18}$. $\lambda$ and $\rho$ are the variable and check degree distributions from the edge perspective. We can construct a $3072 \times 15552$ matrix $\mathbf{H}_{\text{FET},0}$ using $\mathbf{D}_{\text{local}}$ to mask each submatrix on the main diagonal at the local part of $\mathbf{H}_{gc,3}(192, 192)$. Suppose $f = 20$. Based on the construction method described in Algorithm 1, we construct a family of four-edge-type LDPC codes. Particularly, by applying graph extension, we obtain the protomatrix of $\mathbf{H}_{m\to e,i}(j)$ and $\mathbf{H}_{e,i}(j)$, where $1 \le i \le 19$ and $0 \le j \le 2$. Particularly, $\mathbf{p}_e(j)$ and $\mathbf{q}_e(k)$ are connected by one edge if and only if $k = j$. In the terminologies of protograph construction, lifting the protograph of $\mathbf{H}_{m\to e,i}(j)$ and $\mathbf{H}_{e,i}(j)$ is equivalent to dispersing the base matrix of them [11]. Then, we can use the method in [29] to find circulants for the protomatrix of $\mathbf{H}_{m\to e,i}(j)$ and $\mathbf{H}_{e,i}(j)$. Based on the construction method described in Algorithm 1, we construct a family of four-edge-type LDPC codes. And we refer to its

**Input**: $q, t, s, f$, and $\mathscr{R}$
**Output**: $\mathbf{H}_{\mathrm{FET}}$
(1) Using the construction method described in Section 2 to form an $M_{\mathrm{FET},0} \times N_{\mathrm{FET},0}$ matirx $\mathbf{H}_{\mathrm{FET},0}$ which specify a GC-LDPC code $\mathscr{C}_{\mathrm{FET},0}$ with rate $R_{\mathrm{FET},0}$.
(2) **for** $i = 0 : f - 2$ **do**
(3) Use $\mathbf{H}_{\mathrm{FET},i}$ as the mother matrix.
(4) **for** $j = 0 : t - 1$ **do**
(5) Generate a matrix $\mathbf{H}_{m\to e,i+1}(j)$ of size $M_{e,i+1}(i) \times N_{\mathrm{local},i}(j)$.
(6) Generate a matrix $\mathbf{H}_{e,i+1}(j)$ of size $M_{e,i+1}(j) \times N_{e,i+1}(j)$ which is square and has a full rank.
(7) **for** $j = 0 : t - 1$ **do**
(8) Compose the parity-check matrix $\mathbf{H}_{\mathrm{local},i+1}(j)$ of the form

$$\mathbf{H}_{\mathrm{local},i+1}(j) = \begin{bmatrix} \mathbf{H}_{\mathrm{local},i}(j) & \mathbf{0} \\ \mathbf{H}_{m\to e,i+1}(j) & \mathbf{H}_{e,i+1}(j) \end{bmatrix}.$$

(9) **for** $j = 0 : t - 1$ **do**
(10) Compose the matrix $\mathbf{H}_{m\to g,i+1}(j)$ of the form
$\mathbf{H}_{m\to g,i+1}(j) = [\mathbf{H}_{m\to g,i}(j) \quad \mathbf{0}]$.
where $\mathbf{0}$ is an all-zero matrix of size $M_g \times N_{e,i+1}(j)$.
(11) Compose the matrix $\mathbf{H}_{\mathrm{FET},i+1}$ of the form

$$\mathbf{H}_{\mathrm{FET},i+1} = \begin{bmatrix} \mathbf{H}_{\mathrm{local},i+1}(0) & & \\ & \ddots & \\ & & \mathbf{H}_{\mathrm{local},i+1}(t-1) \\ \hline \mathbf{H}_{m\to g,i+1}(0) & \cdots & \mathbf{H}_{m\to g,i+1}(t-1) \end{bmatrix}.$$

update $i : i = i + 1$.

Algorithm 1: Algorithm for constructing RC GC-LDPC codes.

$i$th member as $\mathscr{C}_{\mathrm{FET},i}$, where $0 \le i \le 19$. Consider that the four-edge-type LDPC code is constructed based on finite field which guarantees the structural property of algebraic construction, and the Tanner graphs of such codes have a girth of at least 6. The parameters of such family of four-edge-type LDPC codes are summarized in Table 1 and the diagram of its matrix is shown in Figure 4.

*Example 3.* In order to improve flexibility of code rate, we can puncture the parity bits from a family of four-edge-type LDPC codes. Based on $\mathbf{H}_{\mathrm{FET},1}$ presented in Example 2, for instance, we set the last 98 columns from $\mathbf{H}_{e,1}(j)$ as punctured columns, where $0 \le j \le 2$. Then, we obtain (14682,12480) GC-LDPC code with a rate of 0.85 and denote it as $\mathscr{C}_3$.

## 4. Local/Global Two-Phase Decoding Scheme

For classical iterative decoding scheme, the decoder includes all the VNs in a block and performs total decoding operations in one phase [30]. We refer to such an iterative decoding scheme as *one-phase iterative scheme*. In contrast to classical iterative decoding scheme, Li et al. devised a *two-phase local/global iterative scheme* for CN-GC-LDPC codes [1, 2]. Taking advantage of the cascading structure of *local part*, we can split *local part* into $t$ independent sections. And each section can use an independent decoder at the local phase. If all sections of *local part* are successfully decoded and the locally decoded codeword satisfies the parity-check constraints in *global part*, the locally decoded codeword would be delivered to the user. If it does not, the global decoder starts to process the received codeword from the local decoder. In a good channel environment, we only need to use a group of ($t$ or less) local decoders to process a group of ($t$ or less) consecutive received sections in parallel. This means that we can process a consecutive sequence by some local sequences for a GC-LDPC code. The advantage of this decoding scheme is that lower latency and less memory requirements are required by the decoder. We refer to such a scheme as *normal two-phase local/global iterative scheme*. However, in a bad channel environment, we find that the local decoder performs a plenty of useless operations, which causes the unnecessary cost of the decoder. Therefore, we present a *modified two-phase local/global iterative decoding scheme* for CN-GC-LDPC codes.

*4.1. Modified Local/Global Two-Phase Iterative Decoding Scheme.* Let $\mathbf{z} = (\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_{t-1})$ be the received sequences. Firstly, in local phase of decoding, $t$ received sequences are decoded by $t$ independent decoders with the maximum iteration number $I_{\max_1}$. If all the sections are successfully decoded and the locally decoded codeword satisfies the parity-check constraints in *global part*, the locally decoded codeword could be delivered to the user. If one of the decoders fails to decode a received section, we switch the decoding from the local phase to the global phase. If all sections are successfully decoded, but the locally decoded codeword does not satisfy the parity-check constraints in *global part*, save the decoded information (*LLR*s) and return to the local decoder. We set the maximum iteration number of local decoders to $I_{\max_2}$. Then, if one of the decoders fails to decode a received section or the locally decoded codeword does not satisfy the parity-check constraints in

Table 1: Parameters of a RC GC-LDPC code.

| Member | Size | Code rate | $N$ (bits) | Degree distribution pair |
|---|---|---|---|---|
| $\mathbf{H}_{\mathrm{FET},0}$ | $3072 \times 15552$ | 0.8667 | 14400 | $\lambda(x) = 0.0189x + 0.0849x^2 + 0.7925x^3 + 0.0472x^4 + 0.0566x^5$<br>$\rho(x) = 0.0283x^2 + 0.7170x^{18} + 0.2547x^{80}$ |
| $\mathbf{H}_{\mathrm{FET},1}$ | $3648 \times 16128$ | 0.8333 | 14976 | $\lambda(x) = 0.0088 + 0.0175x + 0.0789x^2 + 0.5614x^3 + 0.2193x^4 + 0.0526x^5 + 0.0614x^6$<br>$\rho(x) = 0.0263x^2 + 0.0702x^7 + 0.6667x^{18} + 0.2368x^{80}$ |
| $\mathbf{H}_{\mathrm{FET},2}$ | $4224 \times 16704$ | 0.8025 | 15552 | $\lambda(x) = 0.0163 + 0.0163x + 0.0732x^2 + 0.2927x^3 + 0.4878x^4 + 0.0488x^5 + 0.065x^7$<br>$\rho(x) = 0.0244x^2 + 0.065x^7 + 0.0732x^8 + 0.6179x^{18} + 0.2195x^{80}$ |
| $\mathbf{H}_{\mathrm{FET},4}$ | $5376 \times 17856$ | 0.7471 | 16704 | $\lambda(x) = 0.0286 + 0.0143x + 0.0429x^2 + 0.1429x^3 + 0.4286x^4 + 0.2143x^5 + 0.0571x^7 + 0.0714x^9$<br>$\rho(x) = 0.0214x^2 + 0.05x^6 + 0.0571x^7 + 0.0643x^8 + 0.0714x^9 + 0.5429x^{18} + 0.1929x^{80}$ |
| $\mathbf{H}_{\mathrm{FET},8}$ | $7680 \times 20160$ | 0.6566 | 19008 | $\lambda(x) = 0.0468 + 0.0117x + 0.0175x^2 + 0.1170x^3 + 0.0877x^4 + 0.2456x^5 + 0.3275x^6 + 0.0702x^{11} + 0.076x^{12}$<br>$\rho(x) = 0.0175x^2 + 0.1228x^6 + 0.0936x^7 + 0.1053x^8 + 0.0585x^9 + 0.4444x^{18} + 0.1579x^{80}$ |
| $\mathbf{H}_{\mathrm{FET},19}$ | $14016 \times 26496$ | 0.4924 | 25344 | $\lambda(x) = 0.0802 + 0.0084x + 0.0338x^3 + 0.1055x^4 + 0.0759x^5 + 0.0591x^6 + 0.0338x^7 + 0.1899x^8 + 0.2532x^9 + 0.0759x^{17} + 0.0844x^{19}$<br>$\rho(x) = 0.0127x^2 + 0.0422x^4 + 0.1772x^5 + 0.1477x^6 + 0.0675x^7 + 0.0759x^8 + 0.0422x^9 + 0.3207x^{18} + 0.1139x^{80}$ |

The first $2 \times 192$ columns of $\mathbf{H}_{\mathrm{local},i}(j)$ and $\mathbf{H}_{m \to g}(j)$ are punctured columns, where $0 \le i \le 19$ and $0 \le j \le 2$.

$\lambda^{(i)}$ and $\rho^{(i)}$ are the variable and check degree distributions from the edge with $\mathbf{H}_{\mathrm{FET},i}$, respectively, where $i = 0, 1, 2, 4, 8, 19$.

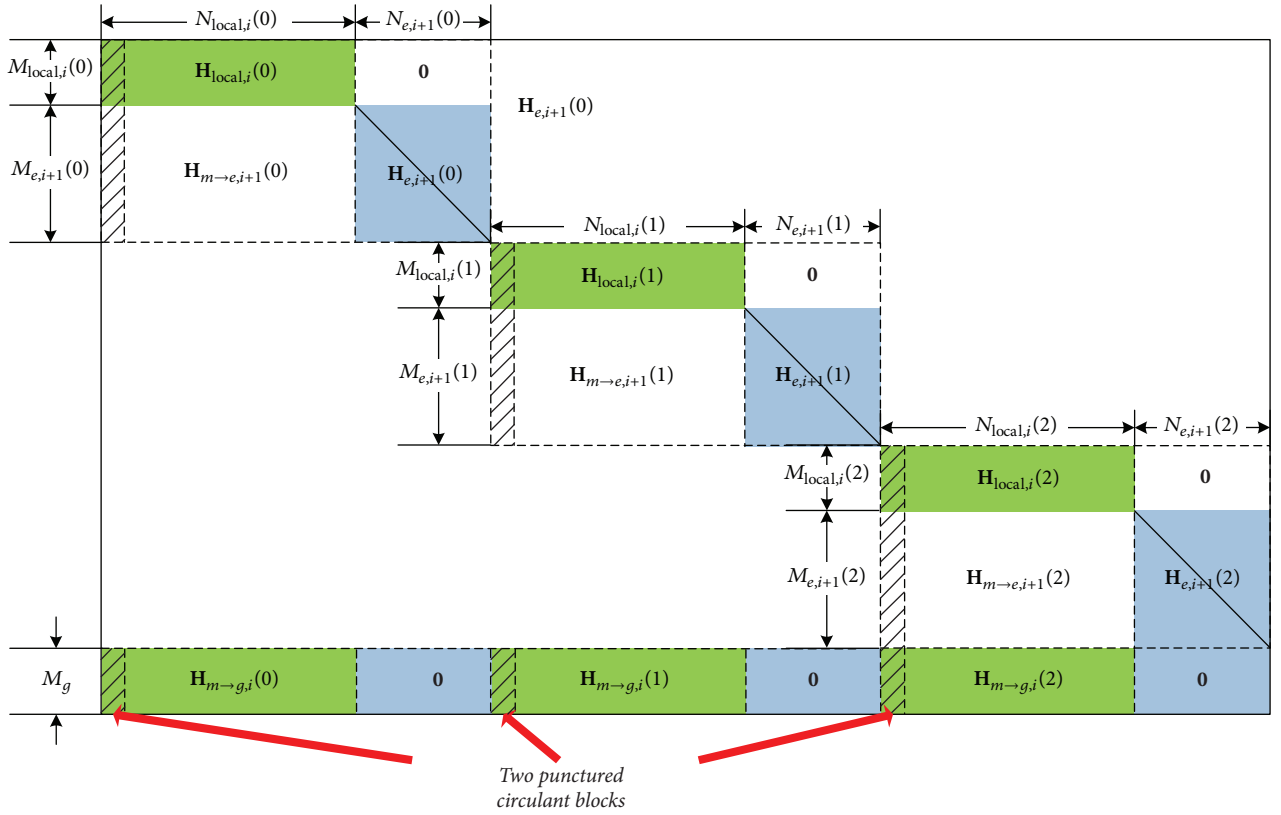



FIGURE 4: Extension structure of parity-check matrix of the RC GC-LDPC codes in Example 2.

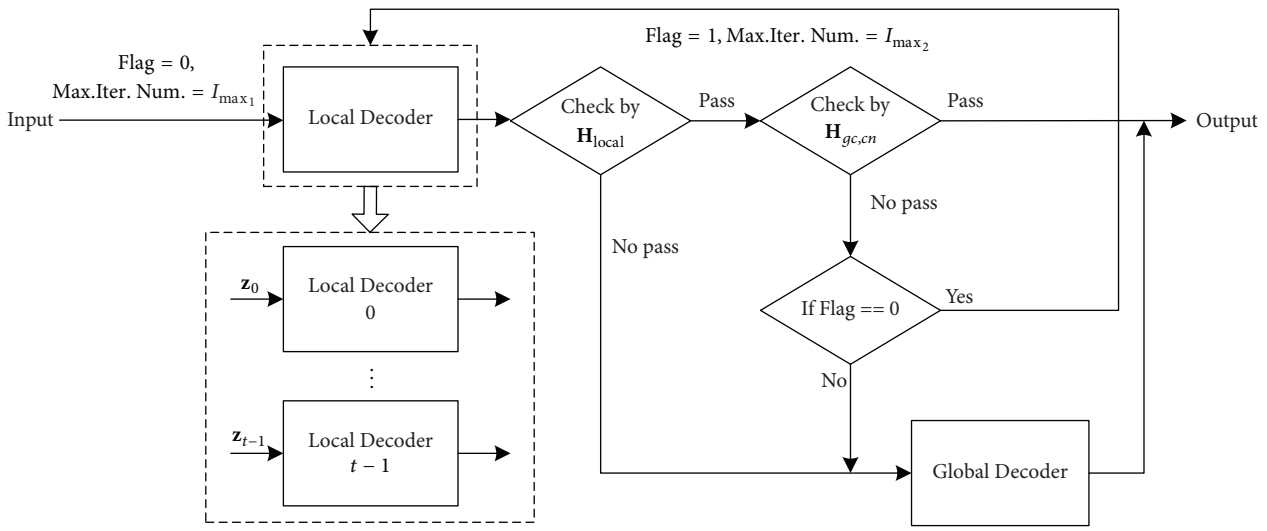



FIGURE 5: Modified local/global two-phase iterative decoding scheme.

*global part*, we switch the decoding from local phase to global phase.

In global phase of decoding, a global decoder is activated. It processes the received vector $\mathbf{z}$ with the channel information and the combined decoded information (*LLR*s) of successfully decoded sections as inputs. And the diagram of *modified two-phase local/global decoding iterative scheme* is illustrated in Figure 5.

We define the order of decoding complexity as the number of operations required per information bit and denote the order of decoding complexity for the normal two-phase decoding algorithm as $\mathcal{O}_{\text{normal}}$. Suppose the number of operations for one iteration of global part is $O_G$ and the number of operations for one iteration in the $j$th section of local part is $O_L(j)$, where $0 < j \leq t-1$. As stated in [1–3], we have

$$\mathcal{O}_{\text{normal}} = \frac{1}{K}\left(O_G I_G + \sum_{j=0}^{t-1} O_L(j)\, I_L(j)\right), \tag{19}$$

where $K$ is the length of the information bits, $I_L(j)$ $(0 \leq I_L(j) \leq I_{\max})$ is the number of iterations involving updates of variables in the $j$th section of the local part, and $I_G$ is the number of iterations involving updates of variables in the global part. Then, the order of decoding complexity of the modified two-phase decoding algorithm, which is denoted as $\mathcal{O}_{\text{modified}}$, can be summarized as

$$\mathcal{O}_{\text{modified}} = \frac{1}{K}\left(\sum_{j=0}^{t-1} O_L(j)\left(I_{L_1}(j) + I_{L_2}(j)\right) + O_G I_G\right), \tag{20}$$

where $I_{L_1}(j)$ $(0 \leq I_{L_1}(j) \leq I_{\max_1})$ and $I_{L_2}(j)$ $(0 \leq I_{L_2}(j) \leq I_{\max_2})$ are the number of iterations involving updates of variables in the $j$th section of the local part in which maximum iterations number is $I_{\max_1}$ and $I_{\max_2}$, respectively. As can be seen, in a bad channel environment, few sections are successfully decoded at first local phase. And each successfully decoded section performs not more than $I_{\max_1}$ iteration operations. So, we have

$$\mathcal{O}_{\text{normal}} \approx \frac{1}{K}\left(O_G I_G + \sum_{j=0}^{t-1} O_L(j)\, I_{\max}\right),$$
$$\mathcal{O}_{\text{modified}} \approx \frac{1}{K}\left(O_G I_G + O_L(0)\, I_{\max_1}(0)\right). \tag{21}$$

Considering $I_{\max_1}(0) \ll I_{\max}$, we have $\mathcal{O}_{\text{modified}} \ll \mathcal{O}_{\text{normal}}$. Moreover, in a good channel environment, most successfully decoded sections satisfy the parity-check constraints in *global part*. Not more than $(I_{\max_1} + I_{\max_2})$ iteration operations are required in each successfully decoded section. For $I_{\max} = I_{\max_1} + I_{\max_2}$, we have $\mathcal{O}_{\text{modified}} \approx \mathcal{O}_{\text{normal}}$.

## 5. Numerical Results

In this section, we first present the simulation performance for RC GC-LDPC codes over the AWGN channel. Then, we compare the decoding complexity of different decoding schemes presented in Section 4. The decoding latency with different decoding schemes is also discussed.

*5.1. Error-Correcting Performance.* We now provide the simulated BER and BLER performances for RC GC-LDPC codes over the AWGN channel with QPSK signaling. It is assumed that all the simulations are performed using the belief propagation (BP) algorithm with the maximum iteration number 50, if not specified. The BER and BLER performances for different code rates are plotted in Figure 6 together with the corresponding Shannon limits. The iterative decoding thresholds achieved by the proposed RC GC-LDPC codes are summarized in Table 2. It can be seen that the gaps between the iterative decoding thresholds and the Shannon limits are very small. Figure 7 depicts the BER and BLER performances

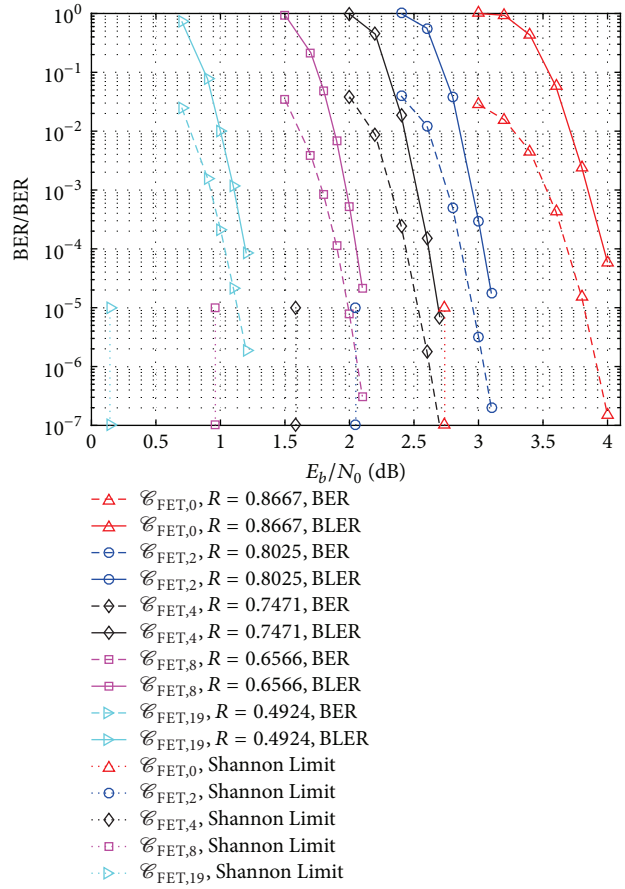



FIGURE 6: The BER and BLER performances for different code rates of $\mathscr{C}_{\text{FET},0}$, $\mathscr{C}_{\text{FET},2}$, $\mathscr{C}_{\text{FET},4}$, $\mathscr{C}_{\text{FET},8}$, and $\mathscr{C}_{\text{FET},19}$ over the AWGN channel with QPSK signaling.

of Examples 1, 2, and 3. We then see that the proposed RC GC-LDPC codes are better than the QC-GC-LDPC codes formed by the classical method in Section 2.

*5.2. Decoding Complexity.* Figures 8 and 9 depict the average iteration number of $\mathscr{C}_1$ and $\mathscr{C}_{\text{FET},0}$ with *one-phase*, *normal two-phase local/global*, and *modified two-phase local/global* iterative schemes based on the BP decoding algorithm, respectively. For both $\mathscr{C}_1$ and $\mathscr{C}_{\text{FET},0}$, the maximum iteration number of *one-phase iterative scheme* is set to 50. The maximum iteration numbers in local decoder and global decoder of *normal two-phase local/global iterative scheme* with them are 50 and 100, respectively. For *modified two-phase local/global iterative scheme*, $I_{\max_1}$, $I_{\max_2}$, and the maximum iteration number in global decoder of $\mathscr{C}_1$ are 30, 20, and 50, respectively. And $I_{\max_1}$, $I_{\max_2}$, and the maximum iteration number in global decoder of $\mathscr{C}_{\text{FET},0}$ are 60, 40, and 100, respectively. Based on Figures 8 and 9, we conclude that the *normal two-phase local/global iterative scheme* requires a significantly higher number of iterations than *modified two-phase local/global iterative scheme* at local phase and needs approximately the same iteration number as the modified scheme at global phase, especially at low and moderate SNRs. At high SNRs, *two-phase local/global iterative scheme* requires

Table 2: Parameters of a RC GC-LDPC code.

| Member | Code rate | Protograph threshold (dB) | Shannon limit (dB) | Gap to capacity (dB) |
|---|---|---|---|---|
| $\mathbf{H}_{\text{FET},0}$ | 0.8667 | 3.1919 | 2.7381 | 0.4538 |
| $\mathbf{H}_{\text{FET},1}$ | 0.8333 | 2.7292 | 2.3167 | 0.4125 |
| $\mathbf{H}_{\text{FET},2}$ | 0.8025 | 2.4967 | 2.0491 | 0.4476 |
| $\mathbf{H}_{\text{FET},4}$ | 0.7471 | 2.0461 | 1.5853 | 0.4608 |
| $\mathbf{H}_{\text{FET},8}$ | 0.6566 | 1.4321 | 0.9605 | 0.4716 |
| $\mathbf{H}_{\text{FET},19}$ | 0.4924 | 0.4879 | 0.1435 | 0.3444 |

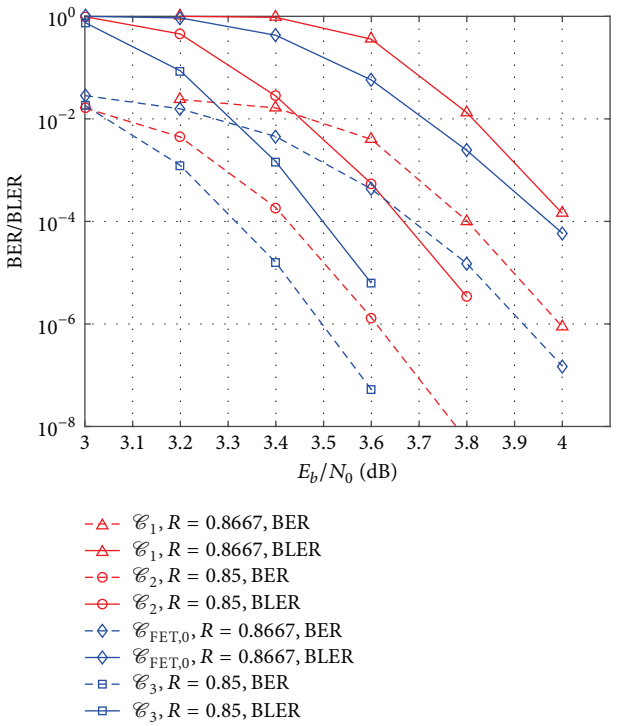



Figure 7: The BER and BLER performances of QC-GC-LDPC codes of $\mathscr{C}_1$, $\mathscr{C}_2$, $\mathscr{C}_{\text{FET},0}$, and $\mathscr{C}_3$ over the AWGN channel with QPSK signaling.

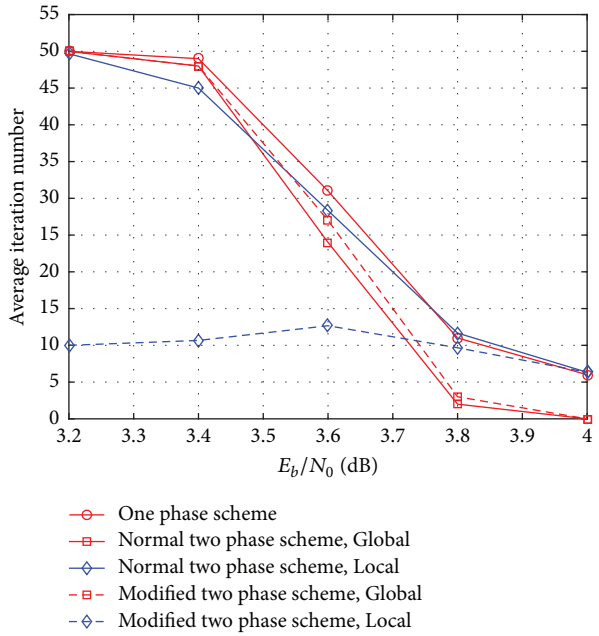



Figure 8: Average iteration number of *one-phase*, *normal two-phase local/global*, and *modified two-phase local/global* iterative scheme based on BP decoding algorithm with $\mathscr{C}_1$.

a smaller number of iterations than *one-phase iterative scheme* without performance hit.

Figures 10 and 11 depict the decoding complexity of $\mathscr{C}_1$ and $\mathscr{C}_{\text{FET},0}$ with *one-phase*, *normal two-phase local/global*, and *modified two-phase local/global* iterative schemes based on BP decoding algorithm, respectively. All operations associated with modulo-2 arithmetic have been neglected as conventionally done. The decoding complexity associated with BP algorithm is evaluated based on the forward and backward recursions proposed in [7]. For $\mathscr{C}_1$, the total complexity associated with one iteration of BP consists of 877,338 *real multiplications*, 104,328 *real divisions*, and 282,366 *real additions* at global phase. At local phase, it consists of 237,258 *real multiplications*, 29,736 *real divisions*, and 79,002 *real additions* in each local decoder. For $\mathscr{C}_{\text{FET},0}$, the total complexity associated with one iteration of BP consists of 559,872 *real multiplications*, 76,608 *real divisions*, and 198,720 *real additions* at global phase. At local phase, it consists of 129,984 *real multiplications*, 20,352 *real divisions*, and 20,352 *real additions* in each local decoder. Based on Figures 10 and 11, we conclude that the *normal two-phase local/global iterative scheme* requires significantly more operations than *modified two-phase local/global iterative scheme* at low and moderate SNRs.

*5.3. Decoding Latency.* The decoding delay in a data transmission system is defined as the delay incurred in receiving the coded bits before decoding takes place and the ensuing decoder processing delay in [31]. In this paper, we assume that all schemes being compared have approximately the same decoding complexity, but the decoder processing time is negligible. For *one-phase iterative scheme*, no information symbols are decoded until an entire block is received. Thus, the maximum decoding delay experienced by an information bit when LDPC code is used for *one-phase iterative scheme* is the arrival time of one incoming block. Suppose its number of iterations is $I_{\text{total}}$. Then, the total decoding latency in received symbols and the total number of soft received values that must be stored in the decoder memory at any given time (decoding latency for short) can be represented by

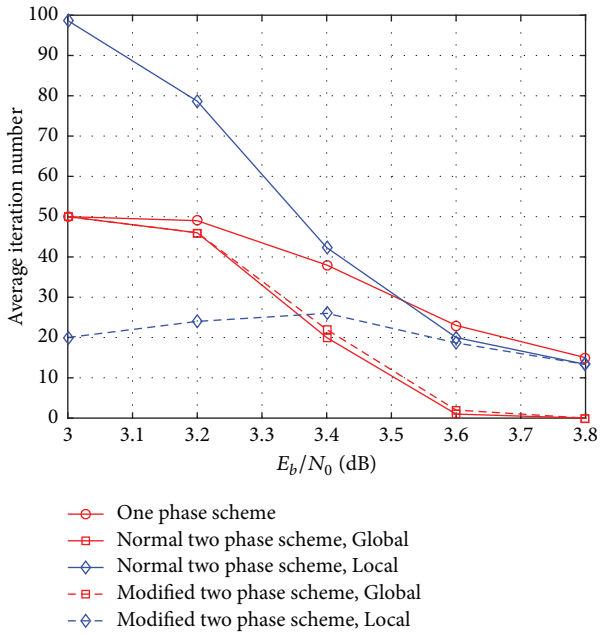



Figure 9: Average iteration number of *one-phase*, *normal two-phase local/global*, and *modified two-phase local/global* iterative scheme based on BP decoding algorithm with $\mathscr{C}_{\mathrm{FET},0}$.

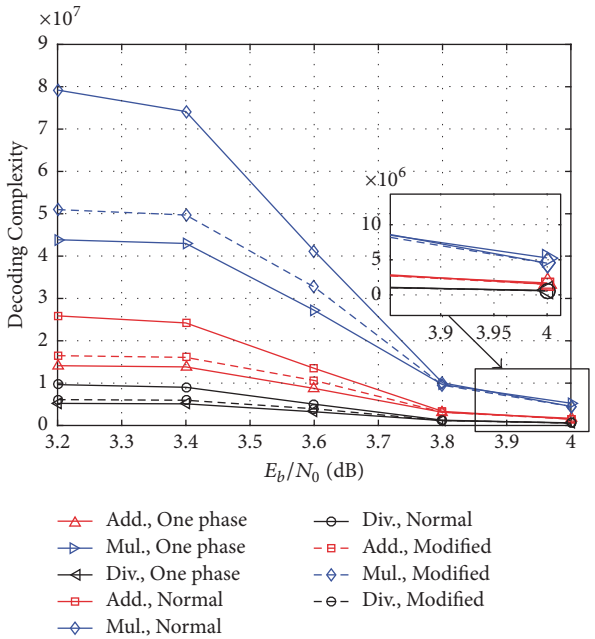



Figure 10: Decoding complexity of $\mathscr{C}_1$ with *one-phase*, *normal two-phase local/global*, and *modified two-phase local/global* iterative scheme based on BP decoding algorithm.

$I_{\mathrm{total}}N$, where $N$ is the total number of symbols. For *two-phase local/global iterative scheme*, all information symbols are assigned to $t$ local decoders. Suppose the number of iterations in global phase is $I_{\mathrm{global}}$. For the $j$th local decoder in *two-phase local/global iterative scheme*, suppose the number of iterations is $I_j$. So, $I_{\mathrm{global}}N + \sum_{j=0}^{t-1} I_j N_{\mathrm{local}}$ represents the

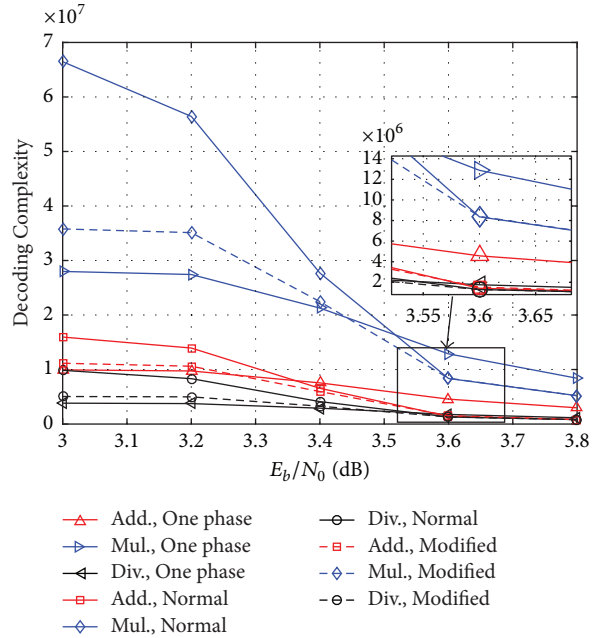



Figure 11: Decoding complexity of $\mathscr{C}_{\mathrm{FET},0}$ with *one-phase*, *normal two-phase local/global*, and *modified two-phase local/global* iterative scheme based on BP decoding algorithm.

decoding latency of *two-phase local/global iterative scheme*. By using $t$ local decoders in fully parallel local phase decoding, the maximum decoding delay experienced by an information bit is the arrival time of each incoming block in local decoder. Then, the decoding latency reduces to $I_{\mathrm{global}}N + \max\{I_j N_{\mathrm{local}},\ 0 \le j \le t-1\}$. Note that $I_{\mathrm{global}}$ decreases with the increase of SNR. At moderate and high SNRs, the decoding latency approaches $\max\{I_j N_{\mathrm{local}},\ 0 \le j \le t-1\}$ when $I_{\mathrm{global}} \to 0$. Since $N_{\mathrm{local}} \ll N$ and $I_{\mathrm{total}} \approx \max\{I_j,\ 0 \le j \le t-1\}$, then $\max\{I_j N_{\mathrm{local}},\ 0 \le j \le t-1\} \ll I_{\mathrm{total}}N$. This means that latency and memory requirements of *two-phase local/global iterative scheme* are much less than for the *one-phase iterative scheme* when the channel environment is better.

## 6. Conclusion

In this paper, we introduced the graph extension through a four-edge-type LDPC code and presented a family of RC GC-LDPC codes; they are constructed by combining algebraic method and graph extension. It was shown that the proposed family of RC CN-GC-LDPC codes can provide more flexibility in code rate and guarantee the structural property of algebraic construction. It is confirmed, by numerical simulations over the AWGN channel, that the proposed RC GC-LDPC codes outperform their counterpart QC-GC-LDPC codes formed by the method in [1, 2] in terms of waterfall performance and exhibit an approximately uniform gap to the capacity over a wide range of rates. Moreover, we presented a *modified two-phase local/global iterative scheme* which can reduce unnecessary cost of local decoders at low and moderate SNRs.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 61771364 and 61701368, the Joint Funds of the National Natural Science Foundation of China under Grant U1504601, and Youth Foundation of the Henan University of Science and Technology (2014QN030).

## References

[1] J. Li, S. Lin, K. Abdel-Ghaffar, W. E. Ryan, and D. J. Costello, "Globally coupled LDPC codes," in *Proceedings of the 2016 Information Theory and Applications Workshop, ITA 2016*, La Jolla, CA, USA, February 2016.

[2] J. Li, S. Lin, K. Abdel-Ghaffar, W. E. Ryan, and D. J. Costello Jr., *LDPC Code Designs, Constructions, and Unification*, Cambridge University Press, 2017.

[3] J. Li, K. Liu, S. Lin, and K. Abdel-Ghaffar, "Reed-Solomon based globally coupled quasi-cyclic LDPC codes," in *Proceedings of the 2017 Information Theory and Applications Workshop, ITA 2017*, San Diego, CA, USA, February 2017.

[4] J. Li, K. Liu, S. Lin, and K. Abdel-Ghaffar, "Reed-solomon based nonbinary globally coupled LDPC codes: Correction of random errors and bursts of erasures," in *Proceedings of the 2017 IEEE International Symposium on Information Theory, ISIT 2017*, pp. 381–385, Aachen, Germany, June 2017.

[5] R. G. Gallager, "Low-Density Parity-Check Codes," *IRE Transactions on Information Theory*, vol. 8, no. 1, pp. 21–28, 1962.

[6] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *IEEE Electronics Letters*, vol. 32, no. 18, pp. 1645-1646, 1996.

[7] D. J. MacKay, "Good error-correcting codes based on very sparse matrices," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 45, no. 2, pp. 399–431, 1999.

[8] S. Kudekar, T. J. Richardson, and R. L. Urbanke, "Threshold saturation via spatial coupling: why convolutional LDPC ensembles perform so well over the BEC," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 57, no. 2, pp. 803–834, 2011.

[9] D. J. Costello Jr., L. Dolecek, T. E. Fuja, J. Kliewer, D. G. M. Mitchell, and R. Smarandache, "Spatially coupled sparse codes on graphs: Theory and practice," *IEEE Communications Magazine*, vol. 52, no. 7, pp. 168–176, 2014.

[10] D. G. Mitchell, M. Lentmaier, and J. Costello, "Spatially coupled LDPC codes constructed from protographs," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 61, no. 9, pp. 4866–4889, 2015.

[11] K. Liu, M. El-Khamy, and J. Lee, "Finite-length algebraic spatially-coupled quasi-cyclic LDPC codes," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 2, pp. 329–344, 2016.

[12] M. Zhang, Z. Wang, Q. Huang, and S. Wang, "Time-Invariant Quasi-Cyclic Spatially Coupled LDPC Codes Based on Packings," *IEEE Transactions on Communications*, vol. 64, no. 12, pp. 4936–4945, 2016.

[13] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, Cambridge, UK, 2005.

[14] J. Hagenauer, "Rate-Compatible Punctured Convolutional Codes (RCPC Codes) and their Applications," *IEEE Transactions on Communications*, vol. 36, no. 4, pp. 389–400, 1988.

[15] J. Ha, J. Kim, and S. W. McLaughlin, "Rate-compatible puncturing of low-density parity-check codes," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 50, no. 11, pp. 2824–2836, 2004.

[16] G. Yue, X. Wang, and M. Madihian, "Design of rate-compatible irregular repeat accumulate codes," *IEEE Transactions on Communications*, vol. 55, no. 6, pp. 1153–1163, 2007.

[17] N. Jacobsen and R. Soni, "Design of rate-compatible irregular LDPC codes based on edge growth and parity splitting," in *Proceedings of the 2007 IEEE 66th Vehicular Technology Conference, VTC 2007-Fall*, pp. 1052–1056, USA, October 2007.

[18] C.-H. Hsu and A. Anastasopoulos, "Capacity achieving LDPC codes through puncturing," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 54, no. 10, pp. 4698–4706, 2008.

[19] T. V. Nguyen, A. Nosratinia, and D. Divsalar, "The design of rate-compatible protograph LDPC codes," *IEEE Transactions on Communications*, vol. 60, no. 10, pp. 2841–2850, 2012.

[20] X. Mu, C. Shen, and B. Bai, "A Combined Algebraic-and Graph-Based Method for Constructing Structured RC-LDPC Codes," *IEEE Communications Letters*, vol. 20, no. 7, pp. 1273–1276, 2016.

[21] Document 3GPP R1-1613710 3GPP TSG RAN WG1 Meeting 87, 3GPP, Nov. 2016, http://www.3gpp.org/ftp/tsg_ran/WG1_RL1/TSGR1_87/Docs/R1-1613710.zip.

[22] Z. Si, R. Thobaben, and M. Skoglund, "Rate-compatible LDPC convolutional codes achieving the capacity of the BEC," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 58, no. 6, pp. 4021–4029, 2012.

[23] W. Hou, S. Lu, and J. Cheng, "Rate-compatible spatially coupled LDPC codes via repeat-accumulation extension," in *Proceedings of the 2014 8th International Symposium on Turbo Codes and Iterative Information Processing, ISTC 2014*, pp. 87–91, Bremen, Germany, August 2014.

[24] H. Zhou, D. G. M. Mitchell, N. Goertz, and D. J. Costello, "Robust rate-compatible punctured LDPC convolutional codes," *IEEE Transactions on Communications*, vol. 61, no. 11, pp. 4428–4439, 2013.

[25] W. Nitzold, G. P. Fettweis, and M. Lentmaier, "Rate-compatible spatially-coupled LDPC code ensembles with nearly-regular degree distributions," in *Proceedings of the IEEE International Symposium on Information Theory, ISIT 2015*, pp. 41–45, Hong Kong, June 2015.

[26] S. Lin and D. J. Costello Jr., *Error Control Coding: Fundamentals and Applications*, Prentice-Hall, Upper Saddle River, NJ, USA, 2nd edition, 2004.

[27] W. Ryan and S. Lin, *Channel Codes: Classical and Modern*, Cambridge University Press, New York, NY, USA, 2009.

[28] Document 3GPP R1-1711982 3GPP TSG RAN WG1 Meeting AH NR2, 3GPP, June 2017, http://www.3gpp.org/ftp/tsg_ran/WG1_RL1/TSGR1_AH/NR_AH_1706/Docs/R1-1711982.zip.

[29] H. Xu, D. Feng, R. Luo, and B. Bai, "Construction of quasi-cyclic LDPC codes via masking with successive cycle elimination,"

*IEEE Communications Letters*, vol. 20, no. 12, pp. 2370–2373, 2016.

[30] M. P. C. Fossorier, M. Mihaljevic, and H. Imai, "Reduced complexity iterative decoding of low-density parity check codes based on belief propagation," *IEEE Transactions on Communications*, vol. 47, no. 5, pp. 673–680, 1999.

[31] S. V. Maiya, D. J. Costello Jr., and T. E. Fuja, "Low latency coding: Convolutional codes vs. LDPC codes," *IEEE Transactions on Communications*, vol. 60, no. 5, pp. 1215–1225, 2012.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 5264724, 9 pages
https://doi.org/10.1155/2018/5264724

*Research Article*

# Construction of Quasi-Cyclic LDPC Codes Based on Fundamental Theorem of Arithmetic

**Hai Zhu**,[1] **Liqun Pu**,[2] **Hengzhou Xu**,[1] **and Bo Zhang**[1]

[1]*School of Network Engineering, Zhoukou Normal University, Zhoukou, China*
[2]*School of Mathematics and Statistics, Zhengzhou University, Zhengzhou, China*

Correspondence should be addressed to Hai Zhu; zhu_sea@163.com and Hengzhou Xu; hzxu@zknu.edu.cn

Received 23 November 2017; Revised 5 February 2018; Accepted 7 March 2018; Published 15 April 2018

Academic Editor: Qin Huang



Quasi-cyclic (QC) LDPC codes play an important role in 5G communications and have been chosen as the standard codes for 5G *enhanced mobile broadband* (eMBB) data channel. In this paper, we study the construction of QC LDPC codes based on an arbitrary given expansion factor (or lifting degree). First, we analyze the cycle structure of QC LDPC codes and give the necessary and sufficient condition for the existence of short cycles. Based on the fundamental theorem of arithmetic in number theory, we divide the integer factorization into three cases and present three classes of QC LDPC codes accordingly. Furthermore, a general construction method of QC LDPC codes with girth of at least 6 is proposed. Numerical results show that the constructed QC LDPC codes perform well over the AWGN channel when decoded with the iterative algorithms.

## 1. Introduction

Low-density parity-check (LDPC) codes [1] are a class of modern channel coding. Because of the advantages of approaching the Shannon capacity and the iterative decoding algorithms with lower complexity, LDPC codes have been attracting great interests of the industries and academia. For various specific communication systems [2–4], LDPC codes have been well designed and chosen as their standard codes. As an important scenario of 5G communications, the *enhanced mobile broadband* (eMBB) data channel had adopted the LDPC coding scheme [5], and LDPC codes have recently been determined after several rounds of discussions [6–12]. However, the other two scenarios of 5G communications, that is, *ultrareliable and low latency communications* (URLLC) and *massive machine-type-communication* (mMTC), have no candidate channel coding at present. The promising coding techniques for 5G communication systems are turbo codes, binary/nonbinary LDPC codes, spatially coupled (SC) LDPC codes [13], block Markov superposition transmission (BMST) [14], and polar codes. The encoding/decoding complexity, performance, spectral efficiency, and robustness comparisons among them can be found in [15]. Recently, some low-complexity decoding algorithms of these modern channel codes have been proposed [16, 17]. These significant works can facilitate and accelerate the applications of these modern coding techniques in 5G communications. According to the definition and description of URLLC and mMTC provided by ITU-R [18], these two scenarios require low latency and high reliability. That is, short data package communication which has no visible error floor down to block error rate (BLER) of $10^{-5}$ should be considered. Research results [19] show that LDPC codes have good performance in the waterfall and error-floor region. Moreover, LDPC codes have good robust property [15, 20] and then their good performance can be also obtained over various channels. Hence, LDPC coding still has a strong competitiveness in the applications of URLLC and mMTC.

LDPC codes can be divided into two major classes: (1) random-like codes constructed by means of computer search under the efficient algorithms [21, 22] and (2) structured codes constructed based on algebraic tools, combinatorial structures, and graphs, such as finite geometries [23], finite fields [24], balanced incomplete block designs (BIBDs) [20], resolvable group divisible designs (RGDDs) [25], and protographs [26, 27]. Research results show that well designed algebraic-based LDPC codes have no error floor at the bit error rate (BER) down to $10^{-15}$ [28]. To

facilitate implementation, LDPC codes usually have some special structures, such as diagonal structure and quasi-cyclic (QC) structure. In general, quasi-cyclic (QC) LDPC codes [29] have advantages of encoding and decoding with low complexity [30, 31], easy hardware implementation [32], and good iterative performance [33], and then they have attracted comprehensive attention.

In order to support lots of data packets with various lengths in the eMBB scenario of 5G communications, the designed 5G LDPC codes are chosen as rate-compatible (RC) QC LDPC codes. Notice that the number of expansion factors (or lifting degrees) of 5G QC LDPC codes is not much. On the other hand, some encoding algorithms [34] are only suitable for QC LDPC codes with certain expansion factor (or lifting degree). Furthermore, the encoding and decoding of QC LDPC codes with expansion factors (or lifting degrees) being the power of two can be easily implemented by linear shift registers. Hence, it is interesting to construct QC LDPC codes from an arbitrary given expansion factor (or lifting degree).

In this paper, we focus on the construction of QC LDPC codes from given expansion factors (or lifting degrees). We first introduce the fundamental theorem of arithmetic in number theory and divide the integer factorization into three categories. By analyzing the cycle structure of QC LDPC codes, we present three classes of QC LDPC codes based on three families of integers. Furthermore, a general construction of QC LDPC codes with girth of at least 6 based on the fundamental theorem of arithmetic is proposed. Finally, in order to show the good performance of our constructed QC LDPC codes, numerical simulation results are provided.

The rest of this paper is organized as follows. Section 2 introduces the fundamentals of number theory, the definitions, basic concepts, and cycle structure of QC LDPC codes. Section 3 presents three classes of QC LDPC codes and a general construction method. Numerical results are also provided in this section. Finally, Section 4 concludes this paper.

## 2. Preliminaries

### 2.1. Fundamentals of Number Theory

**Theorem 1.** *Every composite number, which is greater than one, factors uniquely as a product of prime numbers.*

This theorem is the well-known fundamental theorem of arithmetic in number theory, and it had been proved by Gauss and Clarke [36]. This theorem is also called the unique factorization theorem. That is, every integer greater than one is either prime itself or the product of the prime numbers and this product is unique, up to the order of the factors. For example, $1400 = 14 \times 100 = 2 \times 7 \times 2 \times 2 \times 5 \times 5 = 2^3 \times 5^2 \times 7$. This theorem is twofold: first, 1400 can be written as a product of the primes, and second, no matter how this is done, there will always be three 2s, two 5s, one 7, and no other primes in this product. Hence, for a given integer $L \geq 2$, $L$ can be represented by the unique product; that is,

$$L = p_1^{e_1} \times p_2^{e_2} \times \cdots \times p_k^{e_k}, \tag{1}$$

where $p_1, p_2, \ldots, p_k$ are prime numbers.

The following three lemmas and one theorem are useful for constructing QC LDPC codes with girth of at least 6.

**Lemma 2.** *Let $p$ be a prime, and let $e_1$ and $e_2$ be two positive integers. If $a$ and $b$ are two integers for $1 \leq a \leq p^{e_1} - 1$ and $1 \leq b \leq p^{e_2} - 1$, then $ab \neq 0 \pmod{p^{e_1+e_2}}$.*

*Proof.* Since $a$ and $b$ are two integers with $1 \leq a \leq p^{e_1} - 1$ and $1 \leq b \leq p^{e_2} - 1$, then

$$1 = 1 \times 1 \leq a \times b \leq \left(p^{e_1} - 1\right)\left(p^{e_2} - 1\right) < p^{e_1+e_2}. \tag{2}$$

That is, $ab \neq 0 \pmod{p^{e_1+e_2}}$. □

**Lemma 3.** *Let $p_1$ and $p_2$ be two different primes, and let $e_1$ and $e_2$ be two positive integers. If $a$ and $b$ are two integers for $1 \leq a \leq p_1^{e_1} - 1$ and $1 \leq b \leq p_2^{e_2} - 1$, then $ab \neq 0 \pmod{p_1^{e_1} p_2^{e_2}}$.*

*Proof.* Since $a$ and $b$ are two integers with $1 \leq a \leq p_1^{e_1} - 1$ and $1 \leq b \leq p_2^{e_2} - 1$, then

$$1 \leq a \times b \leq \left(p_1^{e_1} - 1\right)\left(p_2^{e_2} - 1\right) < p_1^{e_1} p_2^{e_2}. \tag{3}$$

That is, $ab \neq 0 \pmod{p_1^{e_1} p_2^{e_2}}$. □

**Lemma 4.** *Let $L = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$ be a positive integer where $p_1, p_2, \ldots, p_k$ are $k$ different primes and $e_1, e_2, \ldots, e_k$ are $k$ positive integers. If $a = p_i^{e_i}$ and $b = p_j^{e_j}$ with $1 \leq i, j \leq k$ and $i \neq j$, then $ab \neq 0 \pmod{L}$.*

*Proof.* Since $a = p_i^{e_i}$ and $b = p_j^{e_j}$ with $1 \leq i, j \leq k$ and $i \neq j$, then

$$1 \leq a \times b \leq p_i^{e_i} p_j^{e_j} < p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k} = L. \tag{4}$$

That is, $ab \neq 0 \pmod{L}$. □

**Theorem 5.** *Let $L$ be a positive integer. If $a$ and $b$ are two positive integers and $ab < L$, then $ab \neq 0 \pmod{L}$.*

*Proof.* Since $a$ and $b$ are two positive integers and $ab < L$, then

$$1 \leq ab < L. \tag{5}$$

That is, $ab \neq 0 \pmod{L}$. □

### 2.2. QC LDPC Codes and Their Associated Tanner Graphs.

A $(\gamma, \rho)$-regular quasi-cyclic (QC) LDPC code [29] of length $\rho L$ can be completely specified by the null space of the following matrix over GF(2):

$$\mathbf{H} = \begin{bmatrix} \mathbf{I}\left(p_{0,0}\right) & \mathbf{I}\left(p_{0,1}\right) & \cdots & \mathbf{I}\left(p_{0,\rho-1}\right) \\ \mathbf{I}\left(p_{1,0}\right) & \mathbf{I}\left(p_{1,1}\right) & \cdots & \mathbf{I}\left(p_{1,\rho-1}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{I}\left(p_{\gamma-1,0}\right) & \mathbf{I}\left(p_{\gamma-1,1}\right) & \cdots & \mathbf{I}\left(p_{\gamma-1,\rho-1}\right) \end{bmatrix}, \tag{6}$$

where, for $0 \leq i \leq \gamma - 1$ and $0 \leq j \leq \rho - 1$, $\mathbf{I}(p_{i,j})$ is an $L \times L$ circulant permutation matrix (CPM) with a one at column-$(r + p_{i,j}) \pmod{L}$ for row-$r$, $0 \leq r \leq L-1$, and zero elsewhere.

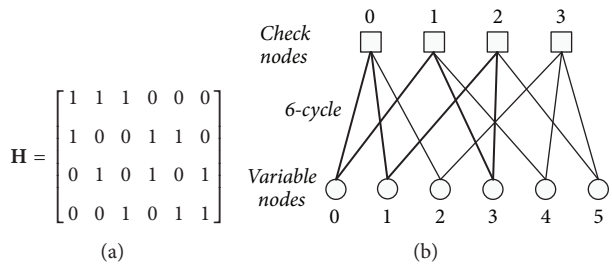



Figure 1: Tanner graph of **H**.

It is clear that $\mathbf{I}(0)$ represents the $L \times L$ identity matrix. Notice that the parameter $L$ is referred to as the *expansion factor* (or *lifting degree*) [37]. It can be easily observed that the positions of nonzero elements in **H** are uniquely determined by the following matrix, called *permutation shift matrix* or *exponent matrix*:

$$\mathbf{P} = \begin{bmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,\rho-1} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,\rho-1} \\ \vdots & \vdots & \ddots & \vdots \\ p_{\gamma-1,0} & p_{\gamma-1,1} & \cdots & p_{\gamma-1,\rho-1} \end{bmatrix}. \tag{7}$$

That is, there is a one-to-one correspondence between **P** and **H**.

An LDPC code is commonly described by a bipartite graph known as Tanner graph [38] in coding theory. Tanner graph of **H**, denoted by $\mathcal{G}(V, C)$, consists of a set $V$ of variable nodes (containing $\rho L$ code symbols of a code word) and a set $C$ of check nodes (containing $\gamma L$ local check-sum constraints on the code symbols). An edge in $\mathcal{G}(V, C)$ connects the variable node $i$ to the check node $j$ if and only if the element at column-$i$ and row-$j$ of **H** is nonzero. A cycle is formed by a sequence of vertices (or edges) in $\mathcal{G}(V, C)$ which starts and ends at the same vertex (or edge) and contains other vertices (or edges) not more than once. The cycle of length $k$ is denoted as $k$-cycle for short and the length of the shortest cycle is called the girth of $\mathcal{G}(V, C)$ (or an LDPC code). As an example, Figure 1 shows the Tanner graph of **H** and an associate 6-cycle.

In graph theory, the biadjacency matrix $\mathbf{A} = [a_{ij}]$ of a bipartite graph $\mathcal{G}(U, V)$ can be constructed as follows. The rows of **A** are labeled by the $|U|$ vertices in $U$ and the columns are labeled by $|V|$ vertices in $V$. The element $a_{ij}$ at the row labeled by the vertex $i \in U$ and the column labeled by the vertex $j \in V$ is 1 if and only if there exists an edge between the vertices $i$ and $j$ and otherwise 0. Actually, for an LDPC code given by the null space of **H**, **H** is the biadjacency matrix of its relevant Tanner graph $\mathcal{G}(V, C)$.

Moreover, isomorphism theory of QC LDPC codes was proposed in [39–41] based on the isomorphism of graphs in graph theory. According to the isomorphism of QC LDPC codes, the parity-check matrix in (6) can be simplified as the following matrix:

$$\mathbf{H} = \begin{bmatrix} \mathbf{I}(0) & \mathbf{I}(0) & \cdots & \mathbf{I}(0) \\ \mathbf{I}(0) & \mathbf{I}(p_{1,1}) & \cdots & \mathbf{I}(p_{1,\rho-1}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{I}(0) & \mathbf{I}(p_{\gamma-1,1}) & \cdots & \mathbf{I}(p_{\gamma-1,\rho-1}) \end{bmatrix}. \tag{8}$$

That is, $p_{i,0} = p_{0,j} = 0$ for $0 \le i \le \gamma - 1$, $0 \le j \le \rho - 1$. Equivalently, its exponent matrix is

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & p_{1,1} & \cdots & p_{1,\rho-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & p_{\gamma-1,1} & \cdots & p_{\gamma-1,\rho-1} \end{bmatrix}. \tag{9}$$

That is why the elements in the first row and first column of the exponent matrix **P** are usually set to 0 in the research process [29, 42]. Hence, we only consider such **H** and **P** in the following discussions.

*2.3. Cycle Structure of QC LDPC Codes.* Consider a QC LDPC code $\mathcal{C}$ given by the null space of **H** in (8). It can be seen from [29] that a cycle in the Tanner graph of $\mathcal{C}$ is associated with a family of the ordered CPMs in **H**. As shown in [29], a $2i$-cycle in the Tanner graph of the code $\mathcal{C}$ (or **H**) is represented by an ordered sequence of CPMs

$$\begin{aligned} &\mathbf{I}(p_{j_0,k_0}), \mathbf{I}(p_{j_1,k_0}), \mathbf{I}(p_{j_1,k_1}), \mathbf{I}(p_{j_2,k_1}), \mathbf{I}(p_{j_2,k_2}), \ldots, \\ &\quad \mathbf{I}(p_{j_{i-1},k_{i-1}}), \mathbf{I}(p_{j_0,k_{i-1}}), \mathbf{I}(p_{j_0,k_0}), \end{aligned} \tag{10}$$

where $j_i = j_0$, $k_i = k_0$, $0 \le j_m \le \gamma - 1$, $j_{m-1} \ne j_m$, $0 \le k_m \le \rho - 1$, and $k_{m-1} \ne k_m$ for $1 \le m \le i$. The above sequence can be simplified as

$$\mathbf{I}(p_{j_0,k_0}), \mathbf{I}(p_{j_1,k_1}), \mathbf{I}(p_{j_2,k_2}), \ldots, \mathbf{I}(p_{j_{i-1},k_{i-1}}). \tag{11}$$

It can be seen that such a $2i$-cycle corresponds to the elements $p_{j_0,k_0}, p_{j_1,k_1}, p_{j_2,k_2}, \ldots, p_{j_{i-1},k_{i-1}}$ in the exponent matrix **P**. Furthermore, short cycles of QC LDPC codes can be determined by the elements of **P** [39, 40].

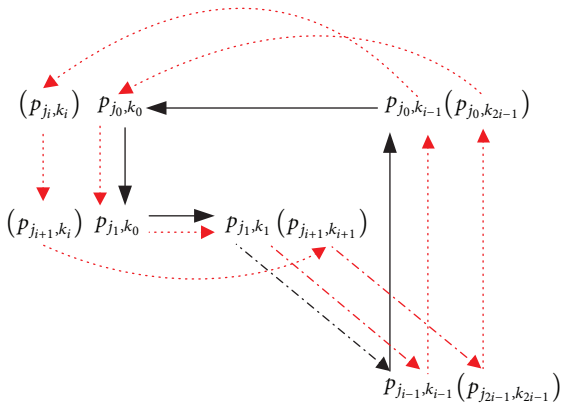



FIGURE 2: The structure of $2i$-cycle and nonexistence of the $4i$-cycle.

Let $g$ be the girth of the code $\mathscr{C}$. It can be seen from [43] that, for $g \le 2i \le 2g-2$, the necessary and sufficient condition for the existence of a $2i$-cycle in the Tanner graph of the code $\mathscr{C}$ (or $\mathbf{H}$) can be generalized as follows:

$$\sum_{m=0}^{i-1} \left(p_{j_m,k_m} - p_{j_{m+1},k_m}\right) = 0 \pmod{L} \tag{12}$$

with $j_0 = j_i$, $k_0 = k_i$, $k_m \neq k_{m+1}$, and $j_m \neq j_{m+1}$. Note that (12) is not the sufficient condition for the existence of a $2i$-cycle in the Tanner graph of the code $\mathscr{C}$ (or $\mathbf{H}$) for $2i \ge 2g$, but it is the necessary condition. Here we give a counterexample. Consider a $2i$-cycle ($i \ge g$) whose cycle structure is given in Figure 2. Clearly, (12) is satisfied. Let $p_{j_m,k_m} = p_{j_{m+i},k_{m+i}}$ for $0 \le m \le i-1$, and $j_0 = j_{2i}$, $k_0 = k_{2i}$, $k_m \neq k_{m+1}$, $j_m \neq j_{m+1}$. According to (12), we have

$$\begin{aligned}\sum_{n=0}^{2i-1} \left(p_{j_n,k_n} - p_{j_{n+1},k_n}\right) &= 2\sum_{m=0}^{i-1} \left(p_{j_m,k_m} - p_{j_{m+1},k_m}\right) \\ &= 0 \pmod{L},\end{aligned} \tag{13}$$

where $j_0 = j_i = j_{2i}$, $k_0 = k_i = k_{2i}$, $k_n \neq k_{n+1}$, $j_n \neq j_{n+1}$, for $0 \le n \le 2i-1$. That is, the ordered elements $p_{j_0,k_0}, p_{j_1,k_0}, p_{j_1,k_1}, \ldots, p_{j_0,k_{2i-1}}$, (i.e., $p_{j_0,k_0}, p_{j_1,k_0}, p_{j_1,k_1}, \ldots, p_{j_0,k_{i-1}}, p_{j_0,k_0}, p_{j_1,k_0}, p_{j_1,k_1}, \ldots, p_{j_0,k_{i-1}}$) make (12) hold, but they do not determine a $4i$-cycle. A visual representation is depicted in Figure 2. Therefore, (4) in [29] and (3) in [39] are not applicable to the cycles with lengths larger than $2g-2$.

## 3. Construction of Quasi-Cyclic LDPC Codes with Girth of at Least 6

Based on the aforementioned, there exists a one-to-one correspondence between the exponent matrix $\mathbf{P}$ and the parity-check matrix $\mathbf{H}$ of a QC LDPC code. Hence, construction of a QC LDPC code is equivalent to the design of its exponent matrix $\mathbf{P}$. In this section, we present three classes of QC LDPC codes with girth of at least 6 and then give a general construction of QC LDPC codes with girth of at least 6 based on an arbitrary integer.

First, we design the exponent matrix $\mathbf{P}$ in (9) as follows:

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & p_{1,1} & \cdots & p_{1,\rho-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & p_{\gamma-1,1} & \cdots & p_{\gamma-1,\rho-1} \end{bmatrix}, \tag{14}$$

where $p_{i,j} = i \times j \pmod{L}$ for $1 \le i \le \gamma-1$, $1 \le j \le \rho-1$. Second, we replace the 0s and $p_{i,j}$ in the designed exponent matrix $\mathbf{P}$ with CPMs $\mathbf{I}(0)$ and $\mathbf{I}(p_{i,j})$ of the same size $L \times L$, respectively, and then obtain a $\gamma \times \rho$ array $\mathbf{H}$ of $L \times L$ CPMs. This array is a $\gamma L \times \rho L$ matrix over GF(2) with column and row weights $\gamma$ and $\rho$, respectively. The null space of this matrix gives a $(\gamma, \rho)$-regular QC LDPC code.

*Remark 6.* As shown in [44], girth and short cycles play an important role in the design of LDPC codes. If the above constructed $(\gamma, \rho)$-regular QC LDPC code does not have good iterative performance, we can replace some CPMs in the above array $\mathbf{H}$ with zero matrices (ZMs) of the same size to reduce the number of short cycles and possibly enlarge the girth value. This replacement is called masking. On the other hand, if the lengths of the desired QC LDPC codes are shorter than $\rho L$ or they require much higher code rates, then we can take a $\gamma' \times \rho'$ subarray of the designed array $\mathbf{H}$, where $\gamma' \le \gamma$ and $\rho' \le \rho$. Notice that this subarray can be obtained from the following two steps: (1) Choose the first $\gamma'$ row-CPMs of the designed array $\mathbf{H}$; (2) select $\rho'$ column-CPMs from $\rho$ column-CPMs of the designed array $\mathbf{H}$. In this paper, both the masking technique and the selection method in [43] are employed to construct (or further optimize) QC LDPC codes.

*3.1. Three Classes of QC LDPC Codes with Girth of at Least 6.* Based on (12), we can see that Tanner graph of the designed array $\mathbf{H}$ contains a 4-cycle if and only if the following equation is satisfied:

$$\begin{aligned}\sum_{m=0}^{1} \left(p_{j_m,k_m} - p_{j_{m+1},k_m}\right) &= (j_0 - j_1)(k_0 - k_1) \\ &= 0 \pmod{L},\end{aligned} \tag{15}$$

where $j_0 \neq j_1$ and $k_0 \neq k_1$. It can be observed that the existence of 4-cycles in the Tanner graph of the designed array $\mathbf{H}$ is related to $L$. According to the fundamental theorem of arithmetic, the values of $L$ can be divided into three categories and three classes of QC LDPC codes with girth of at least 6 are proposed. Notice that all numerical simulations in the following examples, *binary phase shift keying* (BPSK), *additive white Gaussian noise* (AWGN) channel, and the *sum-product algorithm* (SPA), are assumed.

*3.1.1. The Case of $L = p^e$.* Let $L = p^e$, where $p$ is a prime and $e$ is a positive integer, and let $e = e_1 + e_2$, where $e_1, e_2$ are two positive integers and $e_1 \le e_2$. Consider $\gamma = p^{e_1}$ and $\rho = p^{e_2}$. Since $1 \le j_0, j_1 \le p^{e_1} - 1$, $1 \le k_0, k_1 \le p^{e_2} - 1$, $j_0 \neq j_1$, and $k_0 \neq k_1$, then $1 \le j_0 - j_1 \le p^{e_1} - 1$ and $1 \le k_0 - k_1 \le p^{e_2} - 1$, where the calculation is taken modulo $p^{e_1}$ and modulo $p^{e_2}$, respectively. Hence, (15) is not satisfied

according to Lemma 2. That is, Tanner graph of the designed array $\mathbf{H}$ has no 4-cycles and then the constructed QC LDPC codes have girth of at least 6.

*Example 7.* Consider $L = 256 = 2^8$. Let $\gamma = 2^2$ and $\rho = 2^6$. According to (14), we can obtain the exponent matrix $\mathbf{P}$ of size $4 \times 64$. By employing the method in [43], we select the first 4 rows and the 2nd, 16th, 19th, 35th, 50th, 55th, 62nd, and 63rd columns of $\mathbf{P}$ and construct a $4 \times 8$ array $\mathbf{H}$ of $256 \times 256$ CPMs by replacing the elements of the selected submatrix with the corresponding CPMs. By using the matrix

$$\mathbf{M}_1 = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \end{bmatrix} \tag{16}$$

to mask $\mathbf{H}$, a $1024 \times 2048$ matrix with column and row weights 3 and 6, respectively, is obtained. This matrix gives a $(3, 6)$-regular $(2048, 1024)$ QC LDPC code. The bit error rates (BERs) of this code decoded by the SPA (5, 10, 20, and 50 iterations) are shown in Figure 3. Also shown in Figure 3 is the performance of the $(3, 6)$-regular $(2048, 1024)$ algebraic QC LDPC code constructed based on finite field GF($p^s$) [35]. This comparable code is constructed from the prime field GF(257) and then the CPM size of its parity-check matrix is $256 \times 256$. Notice that the exponent matrix and masking matrix of this comparable code are

$$\mathbf{P}_1 = \begin{bmatrix} 179 & 75 & 202 & 52 & 116 & 24 & 15 & 176 \\ 23 & 179 & 75 & 202 & 52 & 116 & 24 & 15 \\ 25 & 23 & 179 & 75 & 202 & 52 & 116 & 24 \\ 162 & 25 & 23 & 179 & 75 & 202 & 52 & 116 \end{bmatrix},$$
$$\mathbf{M}_{4\times 8} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}, \tag{17}$$

respectively. It can be observed that these two codes have similar performance when decoded using the SPA with various iterations. It is well known that algebraic-based LDPC codes have fast decoding convergence [19, 45, 46]. That is, the SPA decoding of the proposed LDPC code also converges fast, as shown in Figure 3. We can see that the performance gap between 20 and 50 iterations is less than 0.15 dB at the BER of $10^{-6}$, and the gap is also less than 0.25 dB at the BER of $10^{-7}$; hence this code achieves a fast rate of decoding convergence.

*3.1.2. The Case of $L = p_1^{e_1} p_2^{e_2}$.* Let $L = p_1^{e_1} p_2^{e_2}$, where $p_1, p_2$ are two different prime numbers and $e_1, e_2$ are two positive integers. Assume $\gamma = \min\{p_1^{e_1}, p_2^{e_2}\}$ and $\rho = \max\{p_1^{e_1}, p_2^{e_2}\}$. Without loss of generality, $p_1^{e_1} < p_2^{e_2}$ is assumed. Since $1 \le j_0, j_1 \le p_1^{e_1} - 1$, $1 \le k_0, k_1 \le p_2^{e_2} - 1$, $j_0 \ne j_1$, and $k_0 \ne k_1$, then $1 \le j_0 - j_1 \le p_1^{e_1} - 1$ and $1 \le k_0 - k_1 \le p_2^{e_2} - 1$, where the calculation is taken modulo $p_1^{e_1}$ and modulo $p_2^{e_2}$, respectively. Hence, (15) is not satisfied according to Lemma 3. That is, Tanner graph of the designed array $\mathbf{H}$ does not contain 4-cycles and the girth of the constructed QC LDPC codes is at least 6.

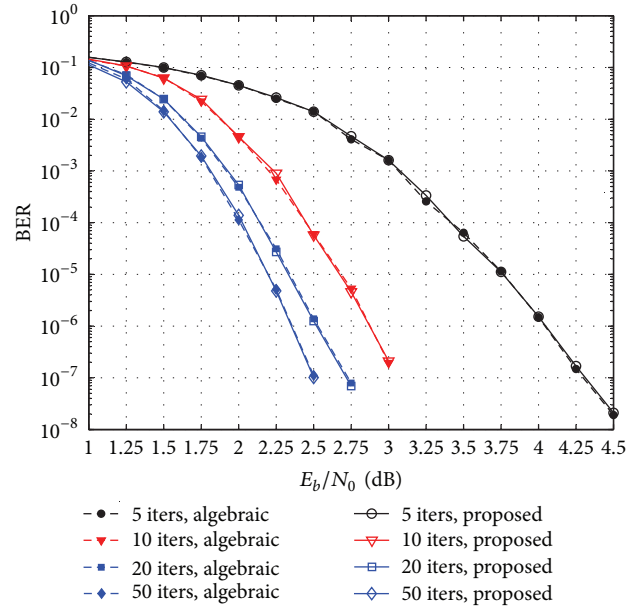



Figure 3: The bit error performance of the proposed $(3, 6)$-regular $(2048, 1024)$ QC LDPC code and the comparable $(3, 6)$-regular $(2048, 1024)$ algebraic QC LDPC code [19] in Example 7. The decoding algorithm is the SPA with 5, 10, 20, and 50 iterations.

*Example 8.* Consider $L = 72 = 2^3 \times 3^2 = 4 \times 18$. Since $12 < 18$, let $\gamma = 2^2$ and $\rho = 18$. According to (14), we can obtain the exponent matrix $\mathbf{P}$ of size $4 \times 18$. By employing the method in [43], we select the first 4 rows and the 1st, 2nd, 3rd, 4th, 5th, 6th, 12th, 13th, 15th, 16th, 17th, and 18th columns of $\mathbf{P}$ and construct a $4 \times 12$ array $\mathbf{H}$ of $72 \times 72$ CPMs by replacing the elements of the selected submatrix with the corresponding CPMs. By using the matrix

$$\mathbf{M}_2 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \end{bmatrix} \tag{18}$$

to mask $\mathbf{H}$, a $288 \times 864$ matrix with column and row weights 3 and 9, respectively, is obtained. This matrix gives a $(3, 9)$-regular $(864, 576)$ QC LDPC code. The bit/word error rates (BERs/WERs) of this code decoded by the SPA with 50 iterations are shown in Figure 4. Also shown in Figure 4 is the performance of the $(3, 9)$-regular $(864, 576)$ algebraic QC LDPC code constructed from the finite field GF(73) [35]. The exponent and masking matrices of this algebraic QC LDPC code are

$$\mathbf{P}_2 = \begin{bmatrix} 49 & 41 & 21 & 40 & 29 & 71 & 39 & 3 & 58 & 61 & 65 & 52 \\ 23 & 49 & 41 & 21 & 40 & 29 & 71 & 39 & 3 & 58 & 61 & 65 \\ 24 & 23 & 49 & 41 & 21 & 40 & 29 & 71 & 39 & 3 & 58 & 61 \\ 57 & 24 & 23 & 49 & 41 & 21 & 40 & 29 & 71 & 39 & 3 & 58 \end{bmatrix},$$

Table 1: The cycle distributions of two (864, 576) QC LDPC codes in Example 8.

| Code | 4-cycles | 6-cycles | 8-cycles | 10-cycles | 12-cycles |
|---|---|---|---|---|---|
| Proposed code | 0 | 288 | 12852 | 110736 | 1514772 |
| Algebraic code [35] | 0 | 360 | 8316 | 109800 | 1402308 |

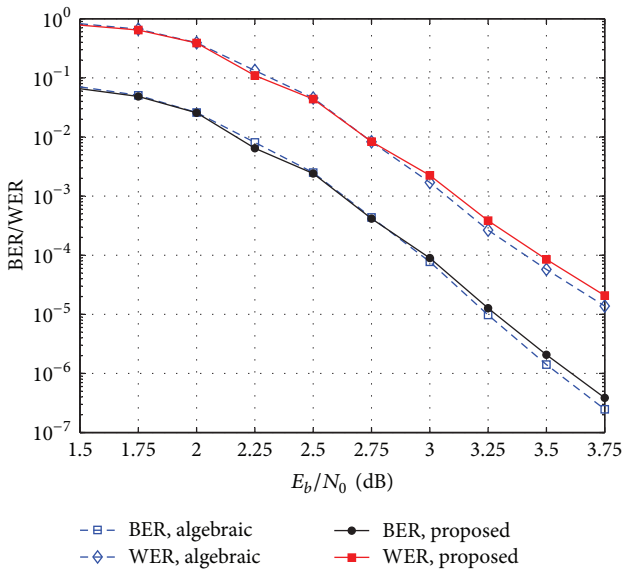



Figure 4: The error performance of the proposed (3, 9)-regular (864, 576) QC LDPC code and the comparable (3, 9)-regular (864, 576) algebraic QC LDPC code constructed based on finite field GF (73) [35] in Example 8.

$$\mathbf{M}_{4\times 12} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}, \tag{19}$$

respectively. Notice that the CPM size of this algebraic code is $72 \times 72$. It can be observed that these two codes also have similar performance. Moreover, the cycle distributions of these two codes are given in Table 1. We can see that although the proposed code has fewer shortest cycles than the algebraic QC LDPC code, the proposed code has much more cycles of length 8 than the algebraic QC LDPC code. That is why the proposed code does not perform better than the algebraic QC LDPC code in the high-SNR region.

*3.1.3. The Case of $L = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$.* Let $L = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$, where $p_1, p_2, \ldots, p_k$ are $k$ different prime numbers and $e_1, e_2, \ldots, e_k$ are $k$ positive integers. Without loss of generality, we assume $p_i^{e_i} < p_j^{e_j}$, where $1 \le i, j \le k$, and $i \ne j$. Consider $\gamma = p_i^{e_i}$ and $\rho = p_j^{e_j}$. Since $1 \le j_0, j_1 \le p_i^{e_i} - 1$, $1 \le k_0, k_1 \le p_j^{e_j} - 1$, $j_0 \ne j_1$, and $k_0 \ne k_1$, then $1 \le j_0 - j_1 \le p_i^{e_i} - 1$ and $1 \le k_0 - k_1 \le p_j^{e_j} - 1$, where the calculation is taken modulo $p_i^{e_i}$ and modulo $p_j^{e_j}$, respectively. Hence, (15) is not satisfied according to Lemma 4. That is, Tanner graph of the designed array $\mathbf{H}$ does not have 4-cycles and the constructed QC LDPC codes have girth of at least 6.

*Example 9.* Consider $L = 105 = 3 \times 5 \times 7$. Since $10 < 21$, let $\gamma = 5$ and $\rho = 21$. According to (14), we can obtain the exponent matrix $\mathbf{P}$ of size $5 \times 21$. By employing the method in [43], we select the first 5 rows and the 1st, 2nd, 3rd, 6th, 7th, 13th, 16th, 17th, 20th, and 21st columns of $\mathbf{P}$ and construct a $5 \times 10$ array $\mathbf{H}$ of $105 \times 105$ CPMs by replacing the elements of the selected submatrix with the corresponding CPMs. By using the matrix

$$\mathbf{M}_3 = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \tag{20}$$

to mask $\mathbf{H}$, a $525 \times 1050$ matrix with column and row weights 3 and 6, respectively, is obtained. This matrix gives a (3, 6)-regular (1050, 525) QC LDPC code of girth 8. For comparison, we simultaneously present the simulation for the (3, 6)-regular (1050, 525) LDPC code constructed based on the progressive edge-growth (PEG) algorithm [22]. The bit/word error rates (BERs/WERs) of these two codes decoded with the SPA (50 iterations) are shown in Figure 5. It can be seen that although these two codes have similar performance in the waterfall region, the proposed code performs better than the PEG-LDPC code in the high-SNR region.

*3.2. A General Construction of QC LDPC Codes from an Arbitrary Positive Integer.* For a given positive integer $L$, we in general find out two positive integers $a$ and $b$ such that $ab \le L$ and $a, b \ge 3$. Assume $a \le b$. Consider $\gamma = a$ and $\rho = b$, where $1 \le i, j \le k$, and $i \ne j$. Since $1 \le j_0, j_1 \le a - 1$, $1 \le k_0, k_1 \le b - 1$, $j_0 \ne j_1$, and $k_0 \ne k_1$, then $1 \le j_0 - j_1 \le a - 1$ and $1 \le k_0 - k_1 \le b - 1$, where the calculation is taken modulo $a$ and modulo $b$, respectively. Hence, (15) is not satisfied according to Theorem 5. That is, Tanner graph of the designed array $\mathbf{H}$ does not have 4-cycles and the constructed QC LDPC codes have girth of at least 6.

*Example 10.* Consider $L = 127 > 4 \times 31$, and let $\gamma = 4$, $\rho = 31$. According to (14), we can obtain the exponent matrix $\mathbf{P}$ of size $4 \times 31$. By employing the method in [43], we select the first 4 rows and the 1st, 2nd, 6th, 7th, 22nd, 26th, 29th, and 31st columns of $\mathbf{P}$ and construct a $4 \times 8$ array $\mathbf{H}$ of $127 \times 127$ CPMs by replacing the elements of the selected submatrix with the

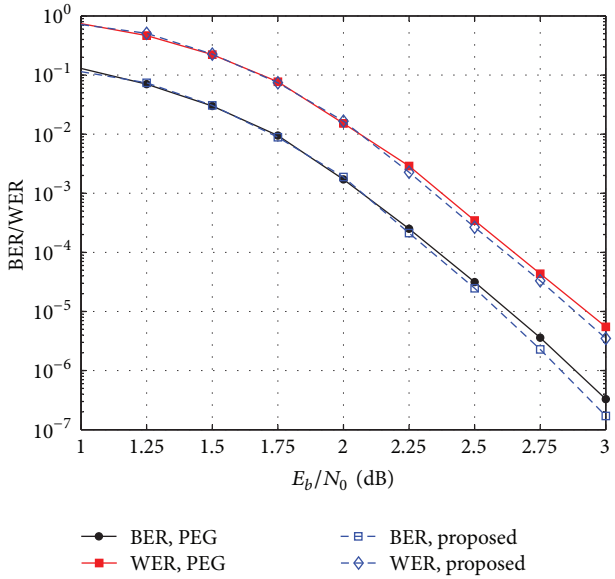



Figure 5: The error performance of the proposed (3, 6)-regular (1050, 525) QC LDPC code and the comparable (3, 6)-regular (1050, 525) QC LDPC code constructed based on the PEG algorithm [22] in Example 9.

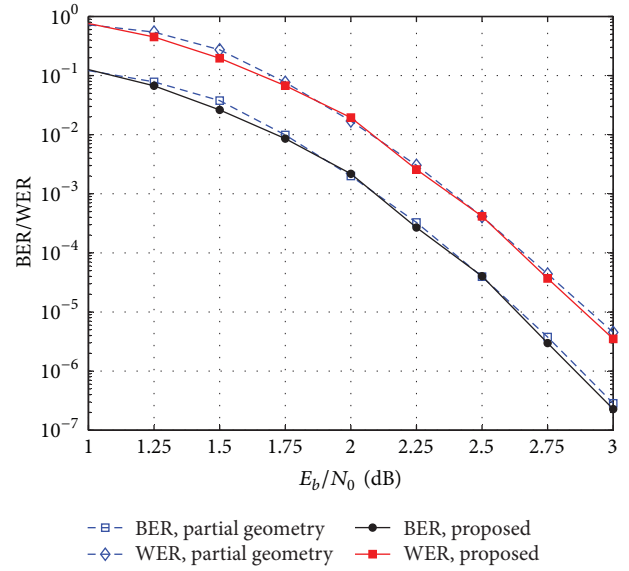



Figure 6: The bit error performance of the proposed (3, 6)-regular (1016, 508) QC LDPC code and the comparable (3, 6)-regular (1016, 508) QC LDPC code constructed from partial geometry [28] in Example 10.

corresponding CPMs. By using the method in [43], we design a masking matrix, that is,

$$\mathbf{M}_4 = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}; \quad (21)$$

to mask $\mathbf{H}$, a $508 \times 1016$ matrix with column and row weights 3 and 6, respectively, is obtained. This matrix gives a (3, 6)-regular (1016, 508) QC LDPC code of girth 8. For comparison, we also construct a (3, 6)-regular (1016, 508) QC LDPC code based on the partial geometry [28]. Note that the exponent matrix of this code is

$$\mathbf{P}_3 = \begin{bmatrix} 2 & 83 & 33 & 46 & 36 & 94 & 42 & 86 \\ 109 & 15 & 84 & 94 & 57 & 43 & 3 & 115 \\ 112 & 76 & 70 & 36 & 111 & 57 & 66 & 117 \\ 31 & 80 & 67 & 78 & 50 & 60 & 16 & 63 \end{bmatrix}, \quad (22)$$

and the masking matrix is also $\mathbf{M}_{4\times 8}$ in Example 7. The bit/word error performance of these two codes decoded by the SPA with 50 iterations is shown in Figure 6. It can be seen that these two codes have similar performance. We can also observe from Figure 6 that, for the proposed QC LDPC code, there are no error floors in the BER curves down to BER = $2.27\times10^{-7}$ and in the WER curves down to WER = $3.5\times10^{-6}$.

## 4. Conclusion

In this paper, based on the fundamental theorem of arithmetic, we presented a method for constructing QC LDPC codes with girth of at least 6 from an arbitrary integer. According to the integer factorization, we divided the integers into three categories and then constructed three classes of QC LDPC codes. Furthermore, a general construction of QC LDPC codes with girth of at least 6 was proposed. Numerical results show that the constructed QC LDPC codes have good performance over the AWGN channel and converge fast under iterative decoding. In other words, for an arbitrary integer $L(\geq 6)$, we can easily construct QC LDPC codes whose parity-check matrices consist of several CPMs and/or zero matrices of size $L \times L$, and the proposed method ensured that the resultant QC LDPC codes have girth of at least 6. Moreover, the proposed QC LDPC codes perform as well as the algebraic QC LDPC codes.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61103143, the Joint Funds of the National Natural Science Foundation of China under Grant U1504601, the Key Scientific and Technological Project of Henan under Grants 162102310589, 172102310124, and 182102310867, the Key Scientific Research Projects of Henan Educational Committee under Grant 18B510022, and the School-Based Program of Zhoukou Normal University under Grant ZKNUB2201705.

## References

[1] R. G. Gallager, "Low-Density Parity-Check Codes," *IRE Transactions on Information Theory*, vol. 8, no. 1, pp. 21–28, 1962.

[2] IEEE Standard, "Air Interface for Fixed Broadband Wireless Access Systems," IEEE Standard P802.16e/D1, 2005.

[3] European Telecommunications Standards Institute, *Digital Video Broadcasting (DVB)*, European Telecommunications Standards Institute, Sophia Antipolis, France, 2009.

[4] CCSDS, "Short Block Length LDPC Codes for TC Synchronization and Channel Coding," CCSDS 231.1-O-1, 2015.

[5] 3GPP, "Document 3GPP chairman'notes 3GPP TSG RAN WG1 meeting 87," 2016, https://www.3gpp.org.

[6] 3GPP, "Document 3GPP chairman'notes 3GPP TSG RAN WG1 meeting AH NR1," 2017, https://www.3gpp.org.

[7] 3GPP, "Document 3GPP chairman'notes 3GPP TSG RAN WG1 meeting 88," 2017, https://www.3gpp.org.

[8] 3GPP, "Document 3GPP chairman'notes 3GPP TSG RAN WG1 meeting 88bis," 2017, https://www.3gpp.org.

[9] 3GPP, "Document 3GPP chairman'notes 3GPP TSG RAN WG1 meeting 89," 2017, https://www.3gpp.org.

[10] 3GPP, "Document 3GPP chairman'notes 3GPP TSG RAN WG1 meeting AH NR2," 2017, https://www.3gpp.org.

[11] 3GPP, "Document 3GPP R1-1711982 3GPP TSG RAN WG1 meeting AH NR2," 2017, https://www.3gpp.org.

[12] 3GPP, "Document 3GPP R1-1712254 3GPP TSG RAN WG1 meeting 90," 2017, https://www.3gpp.org.

[13] M. Zhang, Z. Wang, Q. Huang, and S. Wang, "Time-Invariant Quasi-Cyclic Spatially Coupled LDPC Codes Based on Packings," *IEEE Transactions on Communications*, vol. 64, no. 12, pp. 4936–4945, 2016.

[14] X. Ma, K. Huang, and B. Bai, "Systematic block Markov superposition transmission of repetition codes," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1604–1620, 2018.

[15] B. Bai, "Nonbinary LDPC coding for 5G communication systems," in *Proceedings of the 10th International Conference on Information, Communications and Signal Processing (ICICS'15)*, pp. 2–4, Singapore, 2015.

[16] S. Wang, Q. Huang, and Z. Wang, "Symbol flipping decoding algorithms based on prediction for non-binary LDPC codes," *IEEE Transactions on Communications*, vol. 65, no. 5, pp. 1913–1924, 2017.

[17] Q. Huang, L. Song, and Z. Wang, "Set Message-Passing Decoding Algorithms for Regular Non-Binary LDPC Codes," *IEEE Transactions on Communications*, 2017.

[18] 3GPP, "Study on scenarios and requirements for next generation access technologies," Technical Report (TR) 38.913, 2016.

[19] W. Ryan and S. Lin, *Channel Codes: Classical and Modern*, Cambridge University Press, New York, NY, USA, 2009.

[20] L. Lan, Y. Y. Tai, S. Lin, B. Memari, and B. Honary, "New constructions of quasi-cyclic LDPC codes based on special classes of BIBD's for the AWGN and binary erasure channels," *IEEE Transactions on Communications*, vol. 56, no. 1, pp. 39–48, 2008.

[21] T. Tian, C. Jones, J. D. Villasenor, and R. D. Wesel, "Construction of irregular LDPC codes with low error floors," in *Proceedings of the International Conference on Communications (ICC'03)*, pp. 3125–3129, 2003.

[22] X.-Y. Hu, E. Eleftheriou, and D. M. Arnold, "Regular and irregular progressive edge-growth Tanner graphs," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 51, no. 1, pp. 386–398, 2005.

[23] Q. Diao, Y. Y. Tai, S. Lin, and K. Abdel-Ghaffar, "LDPC codes on partial geometries: construction, trapping set structure, and puncturing," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 59, no. 12, pp. 7898–7914, 2013.

[24] S. Song, B. Zhou, S. Lin, and K. Abdel-Ghaffar, "A unified approach to the construction of binary and nonbinary quasi-cyclic LDPC codes based on finite fields," *IEEE Transactions on Communications*, vol. 57, no. 1, pp. 84–93, 2009.

[25] H. Xu, D. Feng, C. Sun, and B. Bai, "Construction of LDPC codes based on resolvable group divisible designs," in *Proceedings of the International Workshop on High Mobility Wireless Communications (HMWC'15)*, pp. 111–115, 2015.

[26] D. Divsalar, S. Dolinar, C. R. Jones, and K. Andrews, "Capacity-approaching protograph codes," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 6, pp. 876–888, 2009.

[27] D. G. Mitchell, R. Smarandache, and J. Costello, "Quasi-cyclic LDPC codes based on pre-lifted protographs," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 60, no. 10, pp. 5856–5874, 2014.

[28] Q. Diao, J. Li, S. Lin, and I. F. Blake, "New classes of partial geometries and their associated LDPC codes," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 62, no. 6, pp. 2947–2965, 2016.

[29] M. P. Fossorier, "Quasi-cyclic low-density parity-check codes from circulant permutation matrices," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 50, no. 8, pp. 1788–1793, 2004.

[30] Z. Li, L. Chen, L. Zeng, S. Lin, and W. H. Fong, "Efficient encoding of quasi-cyclic low-density parity-check codes," *IEEE Transactions on Communications*, vol. 54, no. 1, pp. 71–81, 2006.

[31] J. Li, K. Liu, S. Lin, and K. Abdel-Ghaffar, "Decoding of quasi-cyclic LDPC codes with section-wise cyclic structure," in *Proceedings of the IEEE Information Theory and Applications Workshop (ITA'14)*, pp. 1–10, Calif, USA, 2014.

[32] F. Cai, X. Zhang, D. Declercq, S. K. Planjery, and B. Vasic, "Finite alphabet iterative decoders for LDPC codes: optimization, architecture and analysis," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 5, pp. 1366–1375, 2014.

[33] H. Liu, Q. Huang, G. Deng, and J. Chen, "Quasi-cyclic representation and vector representation of RS-LDPC Codes," *IEEE Transactions on Communications*, vol. 63, no. 4, pp. 1033–1042, 2015.

[34] Q. Huang, L. Tang, S. He, Z. Xiong, and Z. Wang, "Low-complexity encoding of quasi-cyclic codes based on Galois Fourier transform," *IEEE Transactions on Communications*, vol. 62, no. 6, pp. 1757–1767, 2014.

[35] J. Li, K. Liu, S. Lin, and K. Abdel-Ghaffar, "Algebraic quasi-cyclic ldpc codes: Construction, low error-floor, large girth and a reduced-complexity decoding scheme," *IEEE Transactions on Communications*, vol. 62, no. 8, pp. 2626–2637, 2014.

[36] C. F. Gauss and A. A. Clarke, *Disquisitiones arithmeticae (Second, corrected edition)*, springer, New York, NY, USA, 1966.

[37] J. Li, K. Liu, S. Lin, K. Abdel-Ghaffar, and W. E. Ryan, "An unnoticed strong connection between algebraic-based and protograph-based LDPC codes, Part I: Binary case and interpretation," in *Proceedings of the Information Theory and Applications Workshop (ITA'15)*, pp. 36–45, San Diego, Calif, USA, 2015.

[38] R. M. Tanner, "A recursive approach to low complexity codes," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 27, no. 5, pp. 533–547, 1981.

[39] A. Tasdighi, A. H. Banihashemi, and M.-R. Sadeghi, "Efficient search of girth-optimal QC-LDPC codes," *Institute of Electrical*

*and Electronics Engineers Transactions on Information Theory*, vol. 62, no. 4, pp. 1552–1564, 2016.

[40] C. Sun, H. Xu, D. Feng, and B. Bai, "(3, L) quasi-cyclic LDPC codes: Simplified exhaustive search and designs," in *Proceedings of the 9th International Symposium on Turbo Codes and Iterative Information Processing (ISTC'16)*, pp. 271–275, Brest, France, 2016.

[41] H. Xu, C. Chen, M. Zhu, B. M. Bai, and B. Zhang, "Nonbinary LDPC cycle codes: Efficient search, design, and code optimization," *Science China Information Sciences*, http://engine.scichina.com/doi/10.1007/s11432-017-9271-6.

[42] S. Zhao and X. Ma, "Construction of high-performance array-based non-binary LDPC codes with moderate rates," *IEEE Communications Letters*, vol. 20, no. 1, pp. 13–16, 2016.

[43] H. Xu, D. Feng, R. Luo, and B. Bai, "Construction of quasi-cyclic LDPC codes via masking with successive cycle elimination," *IEEE Communications Letters*, vol. 20, no. 12, pp. 2370–2373, 2016.

[44] H. Xu and B. Bai, "Superposition Construction of Q-Ary LDPC Codes by Jointly Optimizing Girth and Number of Shortest Cycles," *IEEE Communications Letters*, vol. 20, no. 7, pp. 1285–1288, 2016.

[45] Q. Huang, K. Liu, and Z. Wang, "Low-density arrays of circulant matrices: Rank and row-redundancy, and QC-LDPC codes," in *Proceedings of the 2012 IEEE International Symposium on Information Theory, ISIT 2012*, pp. 3073–3077, USA, July 2012.

[46] H. Xu, D. Feng, C. Sun, and B. Bai, "Algebraic-based nonbinary ldpc codes with flexible field orders and code rates," *China Communications*, vol. 14, no. 4, pp. 111–119, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 7063758, 12 pages
https://doi.org/10.1155/2018/7063758

*Research Article*

# Code-Hopping Based Transmission Scheme for Wireless Physical-Layer Security

**Liuguo Yin**[1,2] **and Wentao Hao**[2,3]

[1]*Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China*
[2]*EDA Laboratory, Research Institute of Tsinghua University in Shenzhen, Shenzhen, China*
[3]*School of Aerospace Engineering, Tsinghua University, Beijing 100084, China*

Correspondence should be addressed to Liuguo Yin; yinlg@tsinghua.edu.cn

Received 23 November 2017; Revised 9 February 2018; Accepted 28 February 2018; Published 3 April 2018

Academic Editor: Zesong Fei



Due to the broadcast and time-varying natures of wireless channels, traditional communication systems that provide data encryption at the application layer suffer many challenges such as error diffusion. In this paper, we propose a code-hopping based secrecy transmission scheme that uses dynamic nonsystematic low-density parity-check (LDPC) codes and automatic repeat-request (ARQ) mechanism to jointly encode and encrypt source messages at the physical layer. In this scheme, secret keys at the transmitter and the legitimate receiver are generated dynamically upon the source messages that have been transmitted successfully. During the transmission, each source message is jointly encoded and encrypted by a parity-check matrix, which is dynamically selected from a set of LDPC matrices based on the shared dynamic secret key. As for the eavesdropper (Eve), the uncorrectable decoding errors prevent her from generating the same secret key as the legitimate parties. Thus she cannot select the correct LDPC matrix to recover the source message. We demonstrate that our scheme can be compatible with traditional cryptosystems and enhance the security without sacrificing the error-correction performance. Numerical results show that the bit error rate (BER) of Eve approaches 0.5 as the number of transmitted source messages increases and the security gap of the system is small.

## 1. Introduction

Information security and reliability are two crucial issues in wireless communications. Traditionally, communication systems correct transmission errors at the physical layer based on channel codes and cope with eavesdropping at the application layer based on cryptographic algorithms. In practical scenarios, there will be residual errors in the decoded messages due to the time-varying nature of wireless channels, which may cause severe error diffusion in the decryption. In addition, with the rapid increase of the eavesdropper's computing power, these computational-complexity based encryption algorithms will be easier to break, such as A5/1 in the GSM.

Alternatively, the schemes based on physical-layer security aim to tackle these two crucial issues at the physical layer. Shannon [1] first studied secure communication from an information theoretic perspective in which a preshared secret key between the legitimate parties is used to encrypt the source message. To avoid the key agreement and exploit the inherent randomness of wireless channels, Wyner [2] presented the degraded wiretap channel model in which a transmitter wants to send a secret message to a legitimate receiver through the main channel. This message is also perceived by an eavesdropper through the degraded wiretap channel. The secrecy capacity is defined as the supremum of all the achievable secure and reliable transmission rates. Then, Wyner's original work was generalized to broadcast channels [3] and Gaussian channels [4]. Moreover, the secrecy capacity of fading wiretap channels [5], MIMO wiretap channels [6], and multiuser wiretap channels [7, 8] has been derived in the literature. In these works, the equivocation of Eve is a widely accepted metric for security, which is defined as the conditional entropy of the source message given her noisy observation [9].

Many coding techniques are applied to wiretap channels to make the secrecy transmission rate approach the secrecy capacity, in other words, for the equivocation of Eve to

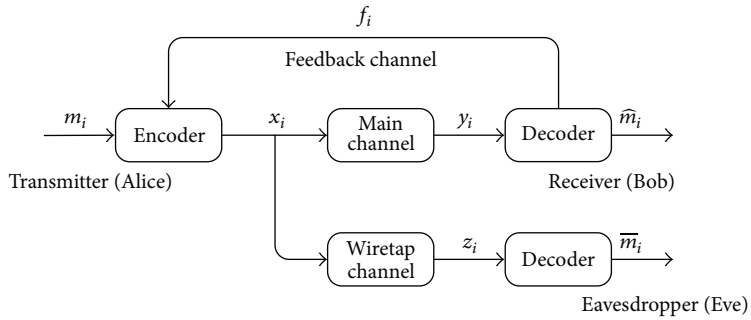



Figure 1: Wiretap channel model with public feedback.

approximate the entropy of the source message. For binary erasure wiretap channels, Thangaraj et al. [10] proposed a coding technique based on the dual of LDPC codes and showed that the secrecy capacity can be achieved by this technique. For symmetric discrete memoryless wiretap channels, Andersson et al. [11] proved that nested polar codes can achieve the whole rate-equivocation region. In addition, this coding technique is further applied to relay-eavesdropper channels [12], block fading channels [13], and multiuser channels [14]. These schemes are really effective when the code length is sufficient, but may be difficult to implement in practical systems.

When we consider the design of practical coding schemes, another valuable metric is the BER [15, 16]. In fact, it is difficult for the eavesdropper to recover any information from the decoded message when she experiences a BER of about 0.5 and the errors are randomly distributed. Security gap is defined as the quantity difference between Bob's and Eve's channels required to achieve a sufficient level of physical-layer security, while ensuring that Bob reliably receives the information [17]. In [17], punctured systematic LDPC codes were exploited to obtain a small security gap. Furthermore, a nonsystematic solution based on scrambled systematic LDPC codes was proposed in [18]. It was proved that the achievable security gap of the scrambled scheme is smaller than that of the punctured method. In [19], scrambling, concatenation, and hybrid automatic repeat-request (HARQ) were combined to reduce the security gap even further. In addition, dynamic LDPC codes are used to enhance the security of the communication system [20]. And protograph LDPC codes [21, 22] can also be used to guarantee the security of the transmission.

In this paper, we propose a scheme based on code-hopping for secrecy transmission over wireless wiretap channels. In the proposed scheme, with ARQ mechanism, the transmitter and the legitimate receiver can select the source messages in real time to distill the secret key. This secret key is then mapped into the parity-check matrix of LDPC codes, which is used to encode the source message. As for the eavesdropper, the uncorrectable decoding errors prevent her from generating the same secret key as the transmitter and the legitimate receiver. Therefore, she cannot obtain the correct parity-check matrix to recover the source message. Theoretical analysis demonstrates that it is difficult for the eavesdropper to generate the same secret key as legitimate parties. Simulation results show that the BER of Eve approaches 0.5 as the number of transmissions increases and the security gap of the system is small.

The remainder of the paper is organized as follows. We introduce our system model and the design of the encoder and decoder in Section 2. In Section 3, the dynamic secret key generation algorithm is proposed and the security of the secret key is well examined. In Section 4, we construct a large number of parity-check matrices of LDPC codes based on the technique we called structured-random protograph expanding. Encoder and Decoder implementation of structured-random LDPC codes are discussed in Section 5. In Section 6, we analyze the reliability and the security performance of our scheme. And some numerical results are given in Section 7. Finally, concluding remarks are provided in Section 8.

## 2. The Proposed Secrecy Transmission Scheme

In this section, we will first introduce the wiretap channel model with public feedback and the concept of security gap. Then, we will propose our secrecy transmission scheme along with the design of encoder and decoder.

*2.1. System Model.* As shown in Figure 1, for $i = 1, 2, \ldots$, message $m_i$ is a sequence of uncoded bits and the length of $m_i$ is $s$. A transmitter named Alice wants to send $m_i$ to a legitimate receiver named Bob through the main channel, but her transmission is also perceived by an eavesdropper named Eve through the wiretap channel. To keep $m_i$ as secret as possible, Alice encodes each length-$s$ message $m_i$ to a length-$n$ codeword $x_i$ by her encoder. The corresponding received codewords by Bob and Eve are denoted by $y_i$ and $z_i$, which are recovered by the decoder as $\widehat{m}_i$ and $\overline{m}_i$, respectively. Additionally, in our model, Bob can use a public feedback channel to inform Alice whether the current codeword is decoded successfully with a feedback signal $f_i$. If there occurs a decoding error at Bob, Alice will retransmit the source message until Bob successfully recovers it or the number of retransmissions reaches the maximum. Taking into account the application in practical scenarios, both channels are assumed to be Gaussian or fading channels:

$$\begin{aligned} y_i &= h_i^B x_i + n_i^B, \\ z_i &= h_i^E x_i + n_i^E, \end{aligned} \tag{1}$$

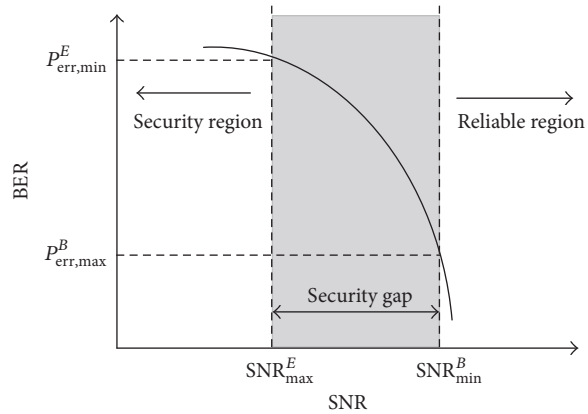



Figure 2: Security gap established by the BER curve.

where $h_i^B$ and $h_i^E$ are the fading coefficients, which are equal to one for Gaussian scenario and follow a certain distribution for fading scenario and $n_i^B$ and $n_i^E$ are zero mean Gaussian noise; $n_i^B \sim \mathcal{N}(0, \sigma_B^2)$ and $n_i^E \sim \mathcal{N}(0, \sigma_E^2)$.

Let $P_e^B$ and $P_e^E$ denote the average BER of Bob and Eve, respectively. As shown in Figure 2, to guarantee the reliability, $P_e^B$ should be lower than a given threshold $P^B_{\text{err,max}}$ ($\approx$0). And to achieve the confidentiality, $P_e^E$ should be larger than a given threshold $P^E_{\text{err,min}}$ ($\approx$0.5). Particularly, if $P_e^E$ is close to 0.5 and the errors are randomly distributed, Eve cannot extract any information from the decoded messages. Based on this observation, the reliability and security of the transmission are guaranteed if conditions (2) and (3) can be satisfied [17], respectively:

$$P_e^B \leq P^B_{\text{err,max}} = P_e\left(\text{SNR}^B_{\text{min}}\right), \tag{2}$$

$$P_e^E \geq P^E_{\text{err,min}} = P_e\left(\text{SNR}^E_{\text{max}}\right), \tag{3}$$

where $\text{SNR}^B_{\text{min}}$ is the lowest signal-to-noise ratio at Bob to guarantee reliability, $\text{SNR}^E_{\text{max}}$ is the highest signal-to-noise ratio at Eve to guarantee security, and $P_e(\cdot)$ denotes the BER as the function of SNR. Then, the security gap is defined as follows [17]:

$$\text{Sg} = \text{SNR}^B_{\text{min}} - \text{SNR}^E_{\text{max}}, \tag{4}$$

where the SNRs are expressed in decibels (dB). Without sacrificing the error-correcting performance of the transmission system, our design targets are making the BER of Eve approach 0.5 and reducing the security gap as much as possible.

*2.2. Design of the Coding Scheme.* To exploit the inherent randomness of wireless channels and the uncorrectable decoding errors of Eve, our scheme is implemented such that the secret keys are distilled from the un-retransmitted source messages, which are then used to generate the parity-check matrices of LDPC codes. During the transmission, the source messages are encoded and decoded by these dynamic parity-check matrices. The block diagrams of the encoder and the decoder are illustrated in Figures 3 and 4, respectively.

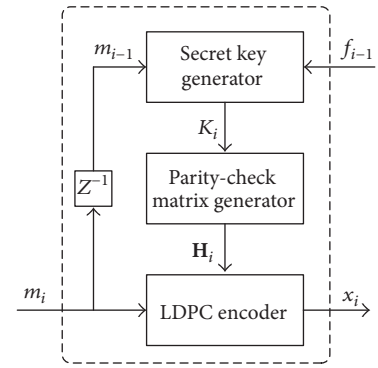



Figure 3: Block diagram of the encoder. Note that the $z^{-1}$ block denotes a delay unit.

In the encoder of Alice, the secret key $K_i$ is updated dynamically according to the received feedback signal $f_{i-1}$ and the source message $m_{i-1}$. If $f_{i-1}$ = ACK, then $K_i$ will be updated according to $m_{i-1}$. If $f_{i-1}$ = NACK, $K_i$ will remain unchanged. The detailed procedure of key update will be discussed in Section 3. Then, the secret key $K_i$ will be used to generate the parity-check matrix of LDPC codes as follows:

$$\mathbf{H}_i = f_H\left(K_i\right), \tag{5}$$

where $f_H(\cdot)$ is the mapping from the secret key to the parity-check matrix. For each source message $m_i$, it will be encoded by the corresponding $\mathbf{H}_i$.

In the decoder of Bob, the integrity of the decoded source message $\widehat{m}_i$ will be checked. If $\widehat{m}_i$ is recovered without errors, the public feedback signal $f_i$ = ACK; otherwise, $f_i$ = NACK. Instead of using the syndrome of the decoded codeword to determine the correctness of $\widehat{m}_i$, we use the cyclic redundancy check (CRC) algorithm to perform integrity check. This is because when the decoded codeword converges to another valid codeword of $\mathbf{H}_i$, the method based on the syndrome cannot detect errors. As for the symmetric key $K_i$ and the

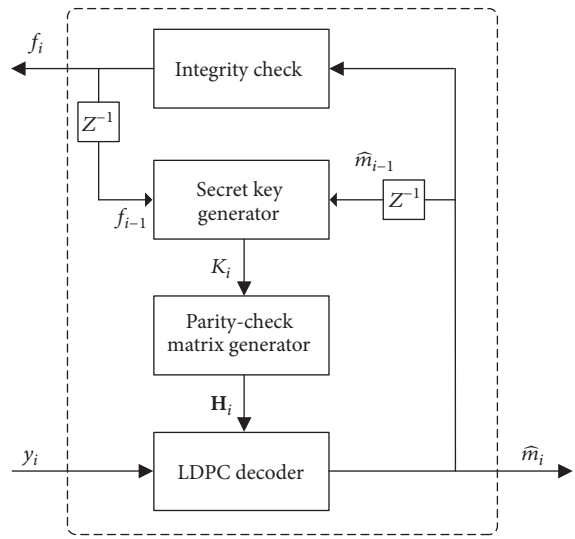



FIGURE 4: Block diagram of the decoder. Note that the $z^{-1}$ block denotes a delay unit.

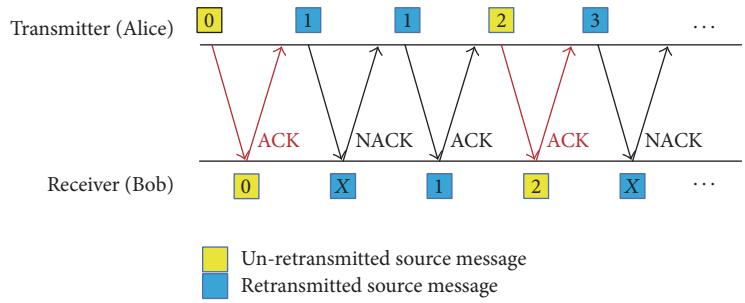



FIGURE 5: The process of automatic source massage selection.

parity-check matrix $\mathbf{H}_i$, they will be generated as in Alice's encoder.

## 3. Dynamic Secret Key Generation Scheme

In this section, we will introduce the dynamic secret key generation algorithm and the mathematical rationales behind it. With this algorithm, Alice and Bob can select the appropriate source messages during the transmission and then distill the secret key based on the universal hashing family.

*3.1. Automatic Source Message Selection.* In this subsection, we will show how Alice and Bob select appropriate source messages in real time during the transmission, which is then hashed into the dynamic secret key. We define $\psi_i^t$ and $\psi_i^r$ as the source message set that is used to generate the secret key $K_i$ at Alice and Bob, respectively. To give Alice and Bob an advantage over Eve, only un-retransmitted source messages will be included in $\psi_i^t$ and $\psi_i^r$. Before the communication begins, $\psi_0^t = \psi_0^r = (\psi_{0,0}, \psi_{0,1}, \ldots, \psi_{0,D-1})$, where $D$ is the number of source messages in the set and $\psi_{0,j}$ is the public agreed initialized binary vector of length-$s$, $j = 0, 1, \ldots, D-1$.

As illustrated in Figure 5, during the transmission, Alice transmits a source message $m_i$ and waits for the corresponding feedback signal $f_i$ before transmitting any new source message. If the received feedback signal $f_i$ = NACK, $\psi_{i+1}^t$ will remain unchanged compared to $\psi_i^t$:

$$\begin{aligned}\psi_{i+1}^t &= (\psi_{i+1,0}, \psi_{i+1,1}, \ldots, \psi_{i+1,D-1}) \\ &= (\psi_{i,0}, \psi_{i,1}, \ldots, \psi_{i,D-1}).\end{aligned} \tag{6}$$

If the received feedback signal $f_i$ = ACK, $\psi_{i+1}^t$ will be updated as follows:

$$\begin{aligned}\psi_{i+1}^t &= (\psi_{i+1,0}, \ldots, \psi_{i+1,D-2}, \psi_{i+1,D-1}) \\ &= (\psi_{i,1}, \ldots, \psi_{i,D-1}, m_i).\end{aligned} \tag{7}$$

As for Bob, if he recovers the source message successfully, he will also update the set $\psi_{i+1}^r$ in the same way and send a feedback signal $f_i$ = ACK. If he fails, he will keep $\psi_{i+1}^r = \psi_i^r$ and send a feedback signal $f_i$ = NACK. This strategy guarantees that $\psi_i^t = \psi_i^r = \psi_i$.

Because there are totally $D$ elements in $\psi_i$ and the length of each element is $s$ bits, the space complexity of storing $\psi_i$ is $O(Ds)$. The update of $\psi_i$ is similar to that of a queue. In the update process, the first element in $\psi_i$ will be removed and discarded. The second element in $\psi_i$ will be moved to

the first location and so on. As for the new element, that is, the source message that has been successfully transmitted, it will be moved to the last location. Considering that the length of each element is $s$ bits, only additional $s$ bits of space are needed to store the element that is being moved. Therefore, the space complexity of updating $\psi_i$ is $O(s)$.

It is very difficult for Eve to reproduce $\psi_i$. She must eavesdrop on not only every source message, but also all of the feedback signals. Whenever the eavesdropper has uncertainty about $\psi_i$, the uncertainty is reflected in the corresponding secret key.

*3.2. Secret Key Distillation.* In this subsection, we will introduce how to distill a secret key from the source message set $\psi_i$. Our target is retaining as much of the eavesdropper's information loss as possible in the secret key. The theory of universal hash family (UHF) provides a powerful solution for us. A UHF is a family of functions such that the random mapping obtained by uniformly choosing a function from this family is almost invertible [23]. In other words, regardless of the actual input distribution, by uniform-randomly choosing a function from a universal hash family, the expected hash output distribution will be close to the uniform. In our considered scenario, $\psi_i$ is hashed into a secret key $K_i$ by using a function $f_{\text{key}}$ that is selected from the universal hash function families $F$. And the conditional distribution of $K_i$ given the eavesdropper's knowledge about $\psi_i$ can be close to the uniform distribution. Because a nearly uniform distribution means nearly maximum entropy, the eavesdropper knows almost nothing about $K_i$. Based on the generalized result from [24], the security of $K_i$ can be evaluated by

$$H\left(K_i \mid F, E_i = e_i\right) \geq H_2\left(K_i \mid F, E_i = e_i\right) \tag{8}$$

$$\geq l - \log_2\left(1 + 2^{l-l_r}\right) \tag{9}$$

$$\geq l - \frac{2^{l-l_r}}{\ln 2}, \tag{10}$$

where $E_i = e_i$ is the eavesdropper's knowledge about $\psi_i$, $l$ is the length of $K_i$ in bits, and $H_2(\cdot)$ is the Renyi entropy of order 2 [24]. When the probability that $E_i = e_i$ is at least $(1 - \delta)$, formula (9) can be generalized as

$$H\left(K_i \mid F, E_i\right) \geq (1 - \delta)\left(l - \log_2\left(1 + 2^{l-l_r}\right)\right). \tag{11}$$

Formulas (9) and (11) show that if the length of the secret key does not exceed $l_r$, $K_i$ is secure because averagely Eve will have less than one-bit information about $K_i$. And $l_r$ can be estimated as follows:

$$l_r \leq H_2\left(\psi_i \mid E_i = e_i\right). \tag{12}$$

It is noteworthy that (9) and (11) are averaged over all uniformly choices of hash functions. It is possible that, for some specific values of $F$, $H(K_i \mid F, E_i)$ is not negligible when $l \leq l_r$. However, it appears with negligible probability [24].

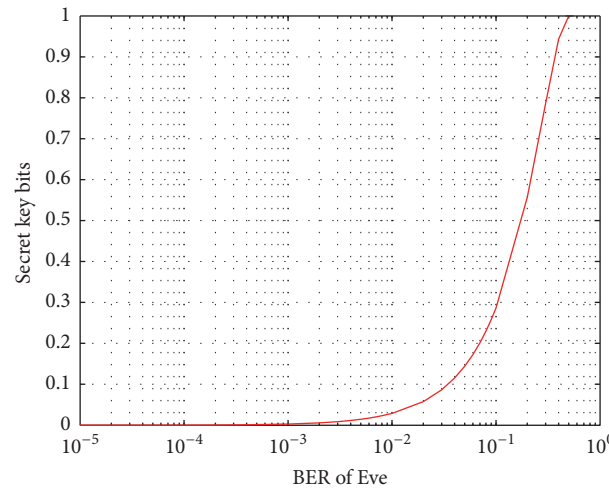



Figure 6: The number of secret key bits we can distill from each source message bit versus Eve's BER.

Because of the randomness of the wireless channel, it is impossible for Eve to recover each source message in $\psi_i$. $H_2(\psi_i \mid E_i = e_i)$ can be calculated as follows [24]:

$$H_2\left(\psi_i \mid E_i = e_i\right) = -d \cdot \log_2\left(\left(1 - \left(P_e^E\right)\right)^2 + \left(P_e^E\right)^2\right), \tag{13}$$

where $d = Ds$ is the length of $\psi_i$ in bits.

Figure 6 illustrates the relationship between Eve's BER and the number of secret key bits we can distill from each source message bit. We can see that, with the increase of Eve's BER, we can distill more secret key bits averagely from each source message bit. It shows that the eavesdropper's information loss is retained in the secret key.

In our considered wiretap channel model, Eve's BER can be calculated by Bob's maximum BER and the security gap:

$$\begin{aligned} P_e^E &= P_e\left(\text{SNR}_E\right) = P_e\left(\text{SNR}_B - \text{Sg}\right) \\ &= P_e\left(P_e^{-1}\left(P_e^B\right) - \text{Sg}\right). \end{aligned} \tag{14}$$

Then, we can calculate $l_r$ as follows:

$$l_r = -d \cdot \log_2\left(\left(1 - P_e^E\right)^2 + \left(P_e^E\right)^2\right) \tag{15}$$

$$\begin{aligned} &= -d \cdot \log_2\left[\left(1 - P_e\left(P_e^{-1}\left(P_e^B\right) - \text{Sg}\right)\right)^2\right. \\ &\quad \left. + \left(P_e\left(P_e^{-1}\left(P_e^B\right) - \text{Sg}\right)\right)^2\right]. \end{aligned} \tag{16}$$

Considering that $d = Ds$, according to (16), we can choose the value of $D$ as follows:

$$D = \frac{-l_r}{\log_2\left[\left(1 - P_e\left(P_e^{-1}\left(P_e^B\right) - \text{Sg}\right)\right)^2 + \left(P_e\left(P_e^{-1}\left(P_e^B\right) - \text{Sg}\right)\right)^2\right] s}. \tag{17}$$

*3.3. Implementation of UHF.* In this subsection, we will show how to implement the universal hash function $f_{\text{key}}(\cdot)$ in

practical scenarios. A Toeplitz matrix is a matrix in which each descending diagonal from left to right is constant and is a kind of UHF that can be implemented with low complexity [25]. In our proposed scheme, we try to generate secret key $K_i$ with length $l$ from the source message set $\psi_i$ with length $d$. The corresponding Toeplitz matrix is as follows:

$$\mathbf{T} = \begin{bmatrix} t_l & t_{l+1} & \cdots & t_{l+d-2} & t_{l+d-1} \\ t_{l-1} & t_l & \cdots & t_{l+d-3} & t_{l+d-2} \\ \vdots & t_{l-1} & \ddots & \vdots & \vdots \\ t_2 & \vdots & \ddots & t_d & t_{d+1} \\ t_1 & t_2 & \cdots & t_{d-1} & t_d \end{bmatrix}, \tag{18}$$

where $t_1, t_2, \ldots, t_l, \ldots, t_{l+d-1}$ is the randomly generated element over GF(2). The secret key can be generated by multiplying $\mathbf{T}$ and $\psi$:

$$K_i = f_{\text{key}}(\psi_i) = \mathbf{T} \times (\psi_i)^T. \tag{19}$$

The computational complexity of (19) is $O(d^2)$. To reduce the computational complexity, we can use the improved algorithm based on fast Fourier transformation (FFT) [26]. Based on the Toeplitz matrix $\mathbf{T}$, we can obtain a new circular matrix $\mathbf{T}_C$ as follows:

$$\begin{aligned} \mathbf{T}_C &= \begin{bmatrix} \mathbf{T} & \mathbf{R}_1 \\ \mathbf{R}_2 & \mathbf{R}_3 \end{bmatrix} \\ &= \text{Circu}\left(t_l, t_{l+1}, \ldots, t_{l+d-1}, t_1, \ldots, t_{l-1}\right), \end{aligned} \tag{20}$$

where $\mathbf{R}_1$, $\mathbf{R}_2$, and $\mathbf{R}_3$ are the submatrices defined in [26], which make the extended matrix $\mathbf{T}_C$ a circular matrix. Circu$(\cdot)$ denotes the circular matrix, which can be represented by its first row.

Then, we generate a new vector $\widehat{\psi}_i = (\psi_i, \mathbf{0})$ by combining $\psi_i$ with a zero vector $\mathbf{0}$, where the length of $\widehat{\psi}_i$ equals the columns of $\mathbf{T}_C$. The secret key can be generated by multiplying $\mathbf{T}_C$ and $\widehat{\psi}_i$, which can be calculated using the FFT-based method:

$$K_i = \mathbf{T}_C \times \widehat{\psi}_i^T = \mathscr{F}^{-1}\left(\mathscr{F}\left(\mathbf{T}_C(1)\right) \circ \mathscr{F}\left(\widehat{\psi}_i\right)\right), \tag{21}$$

where $\mathscr{F}(\cdot)$ is the Fourier transform and $\mathscr{F}^{-1}(\cdot)$ is the inverse, $\mathbf{T}_C(1)$ is the first row of $\mathbf{T}_C$, and $\circ$ denotes the operation that multiplies the corresponding elements in the vector. The computational complexity of (21) is $O(d \log d)$.

## 4. Design of Structure-Random LDPC Codes

In this section, we will show how to construct a large number of parity-check matrices of LDPC codes based on the technique we called structured-random protograph expanding. A protograph is a Tanner graph with a relatively small number of nodes [27], which can be used to construct the parity-check matrix of LDPC codes. Because systematic codes directly expose the secret message bits, all of the $s$ information bits will be punctured and the $n$ parity bits will be transmitted.

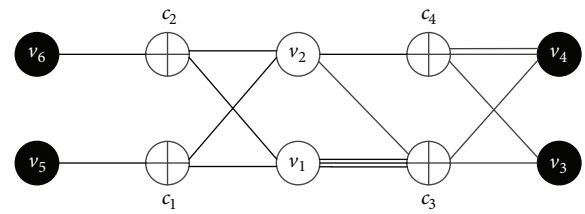



Figure 7: Protograph $P$ for the nonsystematic LDPC code. The rate is 1/2.

We use the code doping method in [28] to design and optimize our protograph to ensure that the iterative decoding of the designed LDPC codes can be triggered successively. Figure 7 shows our optimized protograph $P = (V, C, E)$ for a rate-1/2 nonsystematic LDPC code. We denote $V$ as the set of variable nodes $\{v_1, v_2, \ldots, v_6\}$, $C$ as the set of check nodes $\{c_1, c_2, \ldots, c_4\}$, and $E$ as the set of edges $\{e_1, e_2, \ldots, e_{16}\}$. In the designed protograph, we will puncture the information nodes denoted by $v_1$ and $v_2$ among all the variable nodes to avoid systematic transmission.

To guarantee the convergence of the brief propagation (BP) decoding algorithm, the connection relationship of the check node $c_4$ is specially designed. In our designed protograph, the check node $c_4$ is connected to only one punctured variable node $v_2$. Equivalently, we can use a base parity-check matrix $\mathbf{H}_{B,0}$ with size $4 \times 6$ to represent this protograph.

$$\mathbf{H}_{B,0} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 3 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 2 & 0 & 0 \end{bmatrix}. \tag{22}$$

A "copy-and-permute" operation can be applied to the protograph $P$ to obtain a large derived Tanner graph. We define $T$ as the expanded factor; the "copy-and-permute" operation firstly makes $T$ copies of the protograph $P$ and then permutes the endpoints of each edge among the $T$ variable nodes and $T$ check nodes connected to the set of $T$ edges copied from the same edge from the original protograph $P$.

After this operation, we can obtain a large Tanner graph, where the $T$ copies of the original protograph are connected to each other. Equivalently, we can expand each element of value $w$ in the base matrix $\mathbf{H}_{B,0}$ to a $T \times T$ matrix with $w$ ones in each row or column. As a result, we can obtain a large matrix with size $4T \times 6T$.

Because random permutation is not easy to describe and implement efficiently, in our scheme, we adopt the structured type of permutation, such as cyclic permutation. In other words, we expand each element of value $w$ in the base matrix $\mathbf{H}_{B,0}$ to $T \times T$ circulant permutation matrices $\mathbf{I}_T(t)$. As a result, the expanded parity-check matrix will become a $T$-circulant matrix.

To construct a large number of parity-check matrices of LDPC codes, it is not enough to expand the protograph $P$ with just one single stage. Therefore, we develop a structured-random protograph expanding technique. This technique expands the protograph $P$ with $L > 1$ stages. We denote $T_1, T_2, \ldots, T_L$ as the expanding factors for stages $1, 2, \ldots, L$, respectively. The total expanding factor $T$ can be calculated as $T = T_1 T_2 \cdots T_L$. Finally, the base matrix $\mathbf{H}_{B,0}$ is expanded to the parity-check matrix $\mathbf{H}_{B,L}$.

(i) *Structured expanding*: in the procedure of structured expanding, we expand the protograph $P$ in the first $L-1$ stages to avoid parallel edges, short cycles, and low-weight codewords. As a result, all the nonzero elements in $\mathbf{H}_{B,L-1}$ will be equal to 1.

(ii) *Random expanding*: in the procedure of random expanding, we expand $\mathbf{H}_{B,L-1}$ in the $L$ stage based on the value of the dynamic secret key $K_i$. For each zero element in $\mathbf{H}_{B,L-1}$, we will expand it by a $T_L \times T_L$ zero matrix $\mathbf{0}_{T_L \times T_L}$. For each nonzero element, we will expand it by a $T_L \times T_L$ circulant permutation matrix $\mathbf{I}_{T_L}(t)$. The total number of zero and nonzero elements is $J = |E|T/T_L$.

As for the parameters that are used in the procedure of structured expanding, that is, all the shift values and expanding factors, they are constant and will be shared between Alice and Bob publicly in advance. Now, we rewrite the dynamic secret key $K_i$ as a binary vector:

$$K_i = \left(k_{i,0}, k_{i,1}, \ldots, k_{i,j}, \ldots, k_{i,J-1}\right), \tag{23}$$

where each element $k_{i,j} \in [\![0, T_L - 1]\!]$ is represented by $\log_2 T_L$ bits. Regarding the parameters that are used in the procedure of random expanding, that is, all the random shift values, they are controlled by the dynamic secret key $K_i$, whose length is required to be $l = J \log_2 T_L$ bits.

After expanding the protograph $P$ with $L > 1$ stages, the base matrix $\mathbf{H}_{B,0}$ is expanded to an $n \times (n+s)$ parity-check matrix $\mathbf{H}_i$, where $n = 4T$ and $s = 2T$. As mentioned above, $\mathbf{H}_i$ is a $T_L$-circulant matrix and can be written as $\mathbf{H}_i = [\mathbf{A}(K_i), \mathbf{B}(K_i)]$ such that

$$\mathbf{A}\left(K_i\right) = \left[\mathbf{A}^w_{\alpha\beta}\right]_{2\times 4} = \begin{bmatrix} \mathbf{A}^1_{11} & \mathbf{A}^1_{12} \\ \mathbf{A}^1_{21} & \mathbf{A}^1_{22} \\ \mathbf{A}^3_{31} & \mathbf{A}^1_{32} \\ \mathbf{0} & \mathbf{A}^1_{42} \end{bmatrix},$$
$$\mathbf{B}\left(K_i\right) = \left[\mathbf{B}^w_{\alpha\beta}\right]_{4\times 4} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{B}^1_{13} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{B}^1_{24} \\ \mathbf{B}^1_{31} & \mathbf{B}^1_{32} & \mathbf{0} & \mathbf{0} \\ \mathbf{B}^1_{41} & \mathbf{B}^2_{42} & \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{24}$$

The first $s = 2T$ nodes are punctured as information nodes among all the $n + s = 6T$ variable nodes.

*4.1. An Example.* In this subsection, we construct a large number of nonsystematic (2048, 1024) LDPC codes via $L = 3$ stages. The total expanding factor $T = T_1 T_2 T_3 = 4 \times 4 \times 32 = 512$. With the factor $T_1 = 4$, the first stage aims to separate all the parallel edges. With the factor $T_2 = 4$, the second stage aims to avoid the existence of the cycle of girth 4. With the factor $T_3 = 32$, the third stage aims to randomly expand all the $|E|T/T_3 = 256$ edges. Finally, we get a set of parity-check matrices $\mathscr{H} = \{\mathbf{H}(\mathbf{r}) : \mathbf{r} \in [\![0, 2^{256} - 1]\!]\}$.

During the transmission, we randomly select a parity-check matrix for each source message. The number of iterations is restricted by 63. In Figure 8, we show the average BER of the structured-random nonsystematic (2048, 1024) LDPC codes with different number of retransmissions $r$.

## 5. Encoder and Decoder Implementation of Structured-Random LDPC Codes

*5.1. Encoder Implementation.* To implement the encoder of structured-random LDPC codes, we need to derive the $s \times n$ nonsystematic generator matrix $\mathbf{G}_i$ according to the parity-check matrix $\mathbf{H}_i$. According to (24), $\mathbf{G}_i$ can be derived by

$$\mathbf{G}_i = \left(\mathbf{B}\left(K_i\right)^{-1} \cdot \mathbf{A}\left(K_i\right)\right)^T = \begin{bmatrix} (\mathbf{G}_i)_{11} & (\mathbf{G}_i)_{12} & (\mathbf{G}_i)_{13} & (\mathbf{G}_i)_{14} \\ (\mathbf{G}_i)_{21} & (\mathbf{G}_i)_{22} & (\mathbf{G}_i)_{23} & (\mathbf{G}_i)_{24} \end{bmatrix}$$
$$= \begin{bmatrix} \left(\mathbf{A}^3_{31}\right)^T \mathbf{D}_1 & \left(\mathbf{A}^3_{31}\right)^T \mathbf{D}_3 & \left(\mathbf{A}^1_{11}\right)^T \mathbf{B}_{13} & \left(\mathbf{A}^1_{21}\right)^T \mathbf{B}_{24} \\ \left(\mathbf{A}^1_{32}\right)^T \mathbf{D}_1 \oplus \left(\mathbf{A}^1_{42}\right)^T \mathbf{D}_2 & \left(\mathbf{A}^1_{32}\right)^T \mathbf{D}_3 \oplus \left(\mathbf{A}^1_{42}\right)^T \mathbf{C}^T & \left(\mathbf{A}^1_{12}\right)^T \mathbf{B}_{13} & \left(\mathbf{A}^1_{22}\right)^T \mathbf{B}_{13} \end{bmatrix}, \tag{25}$$

where $\mathbf{D}_1 = (\mathbf{I} \oplus \mathbf{B}^1_{31}(\mathbf{B}^1_{41})^T \mathbf{C}^T (\mathbf{B}^1_{32})^T)\mathbf{B}^1_{31}$, $\mathbf{D}_2 = \mathbf{C}^T(\mathbf{B}^1_{32})^T \mathbf{B}^1_{31}$, $\mathbf{D}_3 = \mathbf{B}^1_{31}(\mathbf{B}^1_{41})^T \mathbf{C}^T$, and $\mathbf{C} = ((\mathbf{B}^2_{42})^T \oplus (\mathbf{B}^1_{32})^T \mathbf{B}^1_{31}(\mathbf{B}^1_{41})^T)^{-1}$.

The multiplication between $m_i$ and $\mathbf{G}_i$ can be calculated in blocks:

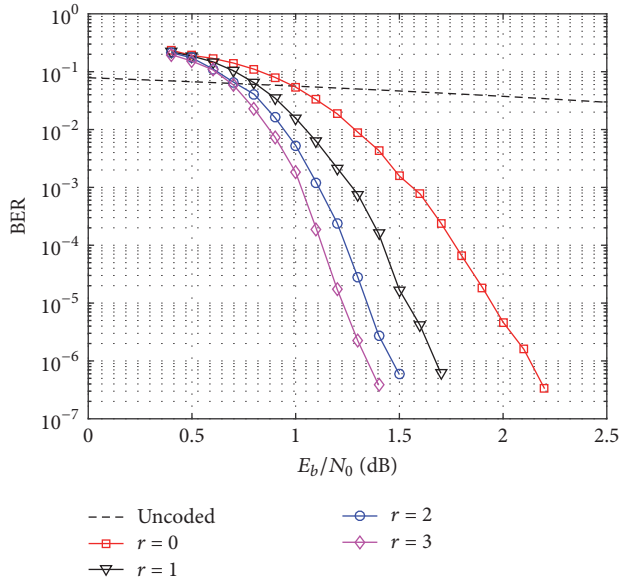



FIGURE 8: The average BER of the nonsystematic (2048, 1024) LDPC codes with different maximum retransmission number $r$.

$$m_i \cdot \mathbf{G}_i = \left[(m_i)_{11} \;\; (m_i)_{12}\right] \times \begin{bmatrix} (\mathbf{G}_i)_{11} & (\mathbf{G}_i)_{12} & (\mathbf{G}_i)_{13} & (\mathbf{G}_i)_{14} \\ (\mathbf{G}_i)_{21} & (\mathbf{G}_i)_{22} & (\mathbf{G}_i)_{23} & (\mathbf{G}_i)_{24} \end{bmatrix} = \left[(x_i)_{11} \;\; (x_i)_{12} \;\; (x_i)_{13} \;\; (x_i)_{14}\right], \tag{26}$$

where $(x_i)_{11} = (m_i)_{11}(\mathbf{G}_i)_{11} + (m_i)_{12}(\mathbf{G}_i)_{21}$, $(x_i)_{12} = (m_i)_{11}(\mathbf{G}_i)_{12} + (m_i)_{12}(\mathbf{G}_i)_{22}$, $(x_i)_{13} = (m_i)_{11}(\mathbf{G}_i)_{13} + (m_i)_{12}(\mathbf{G}_i)_{23}$, and $(x_i)_{14} = (m_i)_{11}(\mathbf{G}_i)_{14} + (m_i)_{12}(\mathbf{G}_i)_{24}$. The multiplication between $(m_i)_{ab}$ and $(\mathbf{G}_i)_{\acute{a}\acute{b}}$ can be further divided as in [29]. For example, to multiply by $(\mathbf{G}_i)_{12}$ can be divided into four steps by successively multiplying by $(\mathbf{A}_{31}^3)^T$, $\mathbf{B}_{31}^1$, $(\mathbf{B}_{41}^1)^T$, and $\mathbf{C}^T$. Because all those submatrices are circulant, all the required multiplications in the encoding process can be finished in $O(n)$ time. The additional computational complexity is from the inversion operation to derive $\mathbf{C}$. In [30], authors have shown the inversion of a binary matrix can be finished in $O(n)$ time by using a parallel hardware architecture. Therefore, the encoding process can be finished in $O(n)$ time.

Consider that the size of $\mathbf{C}$ is about 1/8 of the size of $\mathbf{G}_i$. Thus, the designed encoder for structured-random LDPC codes will increase by 1/8 of the storage compared to the traditional encoder for QC-LDPC codes with a fixed parity-check matrix [31, 32].

*5.2. Decoder Implementation.* As for the decoder of struc\-tured-random LDPC codes, it can be extended from the conventional decoder of quasi-cyclic LDPC codes with a fixed parity-check matrix [33, 34]. This is because the parity-check matrix $\mathbf{H}_i$ of structured-random LDPC codes is also quasi-cyclic as shown in Section 4. The only difference is that the shift values of the circulant permutation matrices in $\mathbf{H}_i$ will be updated according to the dynamic secret key $K_i$. When the shift values are successfully updated, the iterative decoding process is the same. Therefore, the decoder implementation complexity of structured-random LDPC codes will be the same as that of quasi-cyclic LDPC codes with a fixed parity-check matrix.

## 6. Performance Analysis

In this section, we will analyze the security and reliability performances of our proposed scheme. As shown in the previous section, we can construct a large number of nonsystematic LDPC codes that have good error-correction performance. Therefore, we can guarantee that Bob's BER $P_e^B$ will be lower than the given threshold by utilizing these nonsystematic LDPC codes. It guarantees the reliability of the transmission. We will analyze the security of our scheme in two aspects: the complexity when Eve tries to crack the dynamic secret key and Eve's average BER during the whole transmission.

Different from the traditional cryptosystems that have to distribute the secret key before communication begins, our scheme generates the secret key $K_i$ dynamically from the source message set $\psi_i$. During the transmission, an event which is referred to as synchronization error may happen. That is, there exists an index $i_{\mathrm{TH}} \in \mathbb{N}$, such that $z_{i_{\mathrm{TH}}}$ is not correctly decoded by Eve, but $y_{i_{\mathrm{TH}}}$ is successfully recovered by Bob. At this moment, Eve's source message set $\overline{\psi}_{i+1}$ will be different from Alice's and Bob's source message set $\psi_{i+1}$. Therefore, Eve cannot generate the same secret key as Alice and Bob.

As analyzed in Section 3, universal hash function makes the conditional distribution of $K_i$ close to the uniform distribution as follows:

$$P\left(K_i = k_i \mid E_i = e_i\right) \approx \frac{1}{\left|K_i\right|}, \quad \forall k_i \in K_i. \tag{27}$$

From the information theoretic perspective, (27) means that the conditional entropy of $K_i$ is close to its self-information

$$H\left(K_i \mid E_i = e_i\right) \approx H\left(K_i\right) = \log_2 \left|K_i\right|. \tag{28}$$

Therefore, the computational complexity of Eve to crack a dynamic secret key is approximated to $2^{|K_i|} = 2^l$. Even if Eve cracks the secret key by the exhaustive search, the similar synchronization error may happen again and she has to repeat the cracking process.

To evaluate the probability that the synchronization error happens, we denote $P_f(\cdot)$ as the frame error rate (FER) as the function of SNR. Bob's FER and Eve's FER can be expressed as $P_f^B = P_f(\mathrm{SNR}_B)$ and $P_f^E = P_f(\mathrm{SNR}_E)$, respectively. And $i_{\mathrm{TH}}$ is distributed geometrically; $i_{\mathrm{TH}} \sim G(p_0)$, where $p_0 = (1 - P_f^B)P_f^E$. Thus, the probability distribution of $i_{\mathrm{TH}}$ can be calculated as

$$\Pr\left(i_{\mathrm{TH}} = i\right) = \left[1 - \left(1 - P_f^B\right)P_f^E\right]^{i-1}\left(1 - P_f^B\right)P_f^E. \tag{29}$$

As analyzed above, it is difficult for Eve to generate the same secret key as Alice and Bob once the synchronization error happens. In other words, Eve cannot generate the

correct parity-check matrix to decode $z_{i_{\mathrm{TH}}}$. To evaluate Eve's BER during the whole transmission, we can divide the source messages that Eve fails to recover into two categories. The first category contains the source messages that Eve fails to recover before the synchronization error happens. For the messages in the first category, they are recovered by Eve using the correct parity-check matrix. The number of messages in the first category $N_1(i_{\mathrm{TH}})$ obeys the binomial distribution, $N_1(i_{\mathrm{TH}}) \sim B(i_{\mathrm{TH}}-1, p_1)$, where $p_1 = P_f^B P_f^E/(1-(1-P_f^B)P_f^E)$. Thus, the average of $N_1(i_{\mathrm{TH}})$ can be calculated as

$$\overline{N}_1(i_{\mathrm{TH}}) = (i_{\mathrm{TH}})\, p_1 = \frac{(i_{\mathrm{TH}}-1)\, P_f^B P_f^E}{1-\left(1-P_f^B\right)P_f^E}. \tag{30}$$

And the average number of error bits in each error message can be calculated as

$$k_{\mathrm{ER}} = \frac{P_e(\mathrm{SNR}_E)\cdot s}{P_f(\mathrm{SNR}_E)}. \tag{31}$$

For the messages in the second category, half of their bits are wrong, because Eve cannot generate the correct parity-check matrix as Alice and Bob. Finally, Eve's BER can be calculated as

$$\begin{aligned} P_e^E = &\sum_{i=1}^{N} \frac{k_{\mathrm{ER}}\cdot\overline{N}_1(i) + 0.5s\cdot(N-i+1)}{k\cdot N}\cdot \Pr\left(i_{\mathrm{TH}}=i\right) \\ &+ \frac{k_{\mathrm{ER}}\cdot\overline{N}_1(N+1)}{k\cdot N}\cdot\Pr\left(i_{\mathrm{TH}}\geq N+1\right). \end{aligned} \tag{32}$$

Based on (30), $P_e^E$ can be further calculated as

$$\begin{aligned} P_e^E &= \sum_{i=1}^{N}\frac{k_{\mathrm{ER}}\cdot(i-1)\cdot p_1 + 0.5s\cdot(N-i+1)}{s\cdot N} \\ &\quad\cdot\Pr\left(i_{\mathrm{TH}}=i\right) + \frac{k_{\mathrm{ER}}\cdot N\cdot p_1}{s\cdot N}\cdot\Pr\left(i_{\mathrm{TH}}\geq N+1\right) \\ &\geq \sum_{i=1}^{N}\frac{k_{\mathrm{ER}}\cdot(i-1)\cdot p_1 + 0.5s\cdot(N-i+1)}{s\cdot N} \\ &\quad\cdot\Pr\left(i_{\mathrm{TH}}=i\right) \\ &= \frac{0.5s\cdot N\sum_{i=1}^{N}\Pr\left(i_{\mathrm{TH}}=i\right)}{s\cdot N} \\ &\quad+\frac{\left(k_{\mathrm{ER}}\cdot p_1 - 0.5s\right)\sum_{i=1}^{N}(i-1)\cdot\Pr\left(i_{\mathrm{TH}}=i\right)}{s\cdot N} \\ &\geq 0.5\cdot\left(1-\Pr\left(i_{\mathrm{TH}}\geq N+1\right)\right) \\ &\quad-\frac{0.5\cdot\sum_{i=1}^{\infty} i\cdot\Pr\left(i_{\mathrm{TH}}=i\right)}{N} \\ &= 0.5\cdot\left(1-\Pr\left(i_{\mathrm{TH}}\geq N+1\right)\right) - 0.5\cdot\frac{\bar{i}_{\mathrm{TH}}}{N} \\ &= 0.5\cdot\left(1-(1-p_0)^N - \frac{\bar{i}_{\mathrm{TH}}}{N}\right), \end{aligned} \tag{33}$$

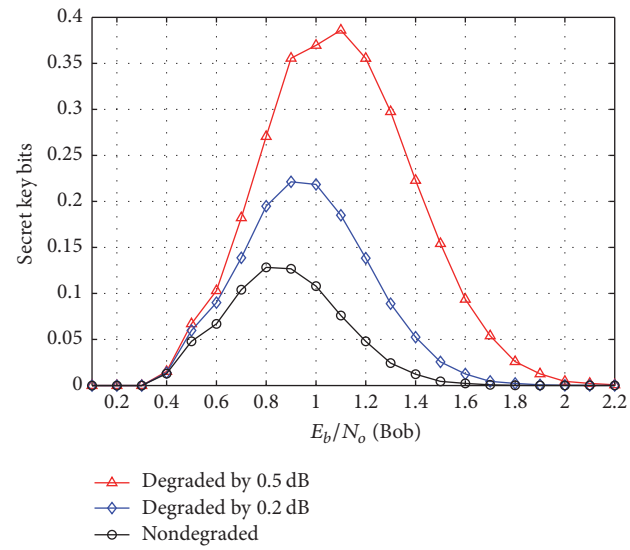



Figure 9: The number of secret key bits we can distill averagely from each transmitted source message bit using (2048, 1024) nonsystematic LDPC codes.

where $\bar{i}_{\mathrm{TH}}$ is defined as follows:

$$\bar{i}_{\mathrm{TH}} = \sum_{i=1}^{\infty} i\cdot\Pr\left(i_{\mathrm{TH}}=i\right) = \frac{1}{p_0} = \frac{1}{\left(1-p_f^B\right)p_f^E}. \tag{34}$$

Finally, $P_e^E$ can be lower bounded as follows:

$$P_e^E \geq 0.5\cdot\left(1-(1-p_0)^N - \frac{1}{N\cdot p_0}\right) \xrightarrow{N\to\infty} 0.5. \tag{35}$$

From the above analysis, we can know that Eve's BER $P_e^E$ will approach 0.5 when the number of the transmitted messages goes to infinity. In addition, when the security gap of the system increases, $(1-P_f^B)$ and $P_f^E$ will increase, and thus $p_0$ will increase. Therefore, we can make Eve's BER $P_e^E$ approach 0.5 with faster speed by increasing the security gap of the system.

## 7. Simulation Results

In this section, we will evaluate the performance of our proposed scheme by Monte-Carlo simulations.

Figure 9 illustrates the number of secret key bits we can distill averagely from each transmitted source message bit. In the region with very low or very high $E_b/N_0$, we can see that the number of secret key bits decreases. The reasons are as follows: in the region with very low $E_b/N_0$, the retransmission happens frequently and the proportion of un-retransmitted source messages is small; in the region with very high $E_b/N_0$, the BER of Eve is very low and therefore the number of secret key bits we can distill averagely from each source message bit is small. In addition, we can see that the more the channel of Eve is degraded compared to that of Bob, the more secret key bits we can distill.

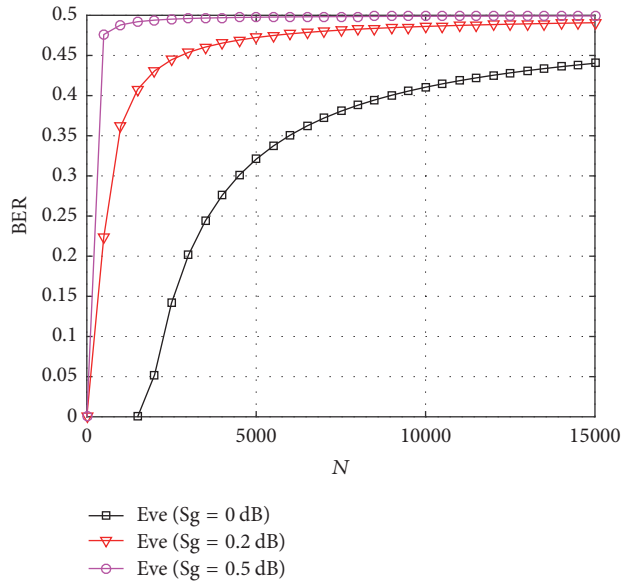



Figure 10: The BER of eavesdropper versus the number of transmitted source messages when $N = 1, 2, \ldots, 15000$.

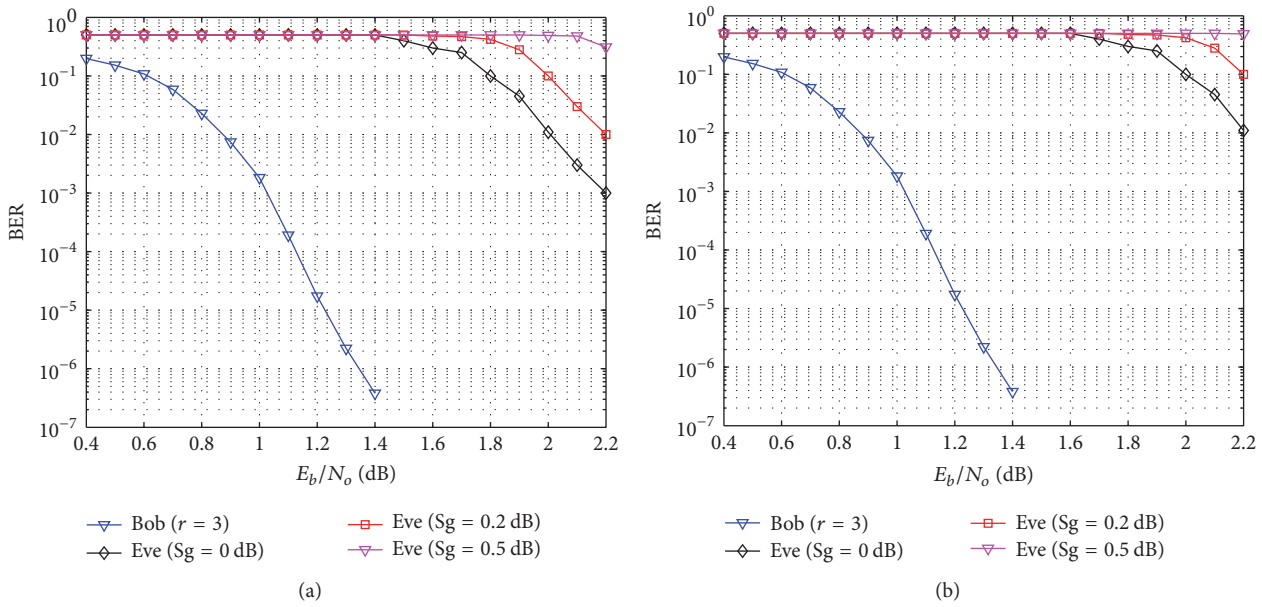



Figure 11: BER of our framework for Gaussian wiretap channel using (2048, 1024) nonsystematic LDPC codes for different security gaps when $N = 1000$ and $N = 10000$ source message are transmitted with maximum retransmission number $r = 3$.

In Figure 10, the BER of Eve versus the number of transmitted source messages $N$ for different security gaps $S_g$ is plotted. We can see from Figure 8 that Bob's BER will be lower than $10^{-6}$ in four conditions: $r = 0$ and $E_b/N_0 = 2.2$ dB, $r = 1$ and $E_b/N_0 = 1.7$ dB, $r = 2$ and $E_b/N_0 = 1.5$ dB, or $r = 3$ and $E_b/N_0 = 1.4$ dB. Therefore, to guarantee the reliability of the transmission ($P^B_{e,\max} < 10^{-6}$), the quality of the main channel can be fixed to $E_b/N_0 = 1.7$ dB and the maximum transmission number can be fixed to $r = 1$. For different security gaps, we can see that the BER of Eve will always approach 0.5 as the number of transmitted source messages increases. This is owing to the fact that the secret keys generated by Eve are the same as the keys generated by Alice and Bob before the first synchronization error happens. Therefore, she can recover the corresponding source messages successfully. After the first synchronization error happens, Eve can no longer decode the following source messages anymore, because the uncorrected decoding errors prevent her from generating the correct secret key. Thus, as the number of transmitted source messages increases, Eve's average BER will approach 0.5.

In Figure 11(a), the BER curves of Bob and Eve are plotted when $N = 1000$ for different security gaps $S_g$. The maximum retransmission number is fixed to $r = 3$. If Bob's BER

threshold is set to $P^B_{e,\max} < 10^{-6}$ and Eve's BER threshold is set to $P^E_{e,\min} = 0.49$, the security gap $S_g = 0$ dB can be achieved. In Figure 11(b), the BER curves of Bob and Eve are plotted when $N = 10000$. We can see that security gap $S_g$ can be further reduced to lower than 0 dB. It means that security of the source message can be guaranteed even when the wiretap channel is better than the main channel. We can see that the security gap performance of our scheme is really small and can be improved by increasing $N$.

## 8. Conclusions

In this paper, we have proposed a secrecy transmission scheme based on code-hopping to encrypt and encode the source messages at the physical layer for wireless communications. First, we present a dynamic secret key generation algorithm based on ARQ mechanism. With this algorithm, Alice and Bob can distill the secret keys from the unretransmitted source messages based on the universal hash families. Second, we present a structured-random LDPC codes design algorithm. Based on this algorithm, we generate a large amount of parity-check matrices of LDPC codes. During the transmission, Alice and Bob dynamically select the parity-check matrices of LDPC codes to encode and recover the source messages based on the dynamic secret keys. Theoretical analysis demonstrates that it is difficult for Eve to generate the same secret key as Alice and Bob. Simulation results show that the BER of Eve will approach 0.5 as the number of transmitted source messages increases and the security gap of our system is small.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC, 91538203), the New Strategic Industries Development Projects of Shenzhen City (JCYJ20150403155812833), and the Joint Research Foundation of the General Armaments Department and the Ministry of Education of China (6141A02033322).

## References

[1] C. E. Shannon, "Communication theory of secrecy systems," *Bell Labs Technical Journal*, vol. 28, pp. 656–715, 1949.

[2] A. D. Wyner, "The wire-tap channel," *Bell Labs Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[3] I. Csiszar and J. Korner, "Broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 339–348, 1978.

[4] S. K. Leung-Yan-Cheong, "On a Special Class of Wiretap Channels," *IEEE Transactions on Information Theory*, vol. 23, no. 5, pp. 625–627, 1977.

[5] P. K. Gopala, L. Lai, and H. El Gamal, "On the secrecy capacity of fading channels," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4687–4698, 2008.

[6] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4961–4972, 2011.

[7] E. Ekrem and S. Ulukus, "The secrecy capacity region of the Gaussian MIMO multi-receiver wiretap channel," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2083–2114, 2011.

[8] A. S. Mansour, R. F. Schaefer, and H. Boche, "The individual secrecy capacity of degraded multi-receiver wiretap broadcast channels," *IEEE International Conference on Communications*, pp. 4181–4186, 2015.

[9] A. Yener and S. Ulukus, "Wireless physical-layer security: lessons learned from information theory," *Proceedings of the IEEE*, vol. 103, no. 10, pp. 1814–1825, 2015.

[10] A. Thangaraj, S. Dihidar, A. R. Calderbank, S. W. McLaughlin, and J.-M. Merolla, "Applications of LDPC codes to the wiretap channel," *IEEE Transactions on Information Theory*, vol. 53, no. 8, pp. 2933–2945, 2007.

[11] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund, "Nested polar codes for wiretap and relay channels," *IEEE Communications Letters*, vol. 14, no. 8, pp. 752–754, 2010.

[12] B. Duo, P. Wang, Y. Li, and B. Vucetic, "Secure transmission for relay-eavesdropper channels using polar coding," *IEEE International Conference on Communications*, pp. 2197–2202, 2014.

[13] H. Si, O. O. Koyluoglu, and S. Vishwanath, "Hierarchical polar coding for achieving secrecy over state-dependent wiretap channels without any instantaneous CSI," *IEEE Transactions on Communications*, vol. 64, no. 9, pp. 3609–3623, 2016.

[14] Y.-P. Wei and S. Ulukus, "Polar Coding for the General Wiretap Channel with Extensions to Multiuser Scenarios," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 2, pp. 278–291, 2016.

[15] Z. Chen, L. Yin, and J. Lu, "Hamming distortion based secrecy systems: To foil the eavesdropper with finite shared key," *IEEE Communications Letters*, vol. 19, no. 5, pp. 711–714, 2015.

[16] P. Wang, L. Yin, and J. Lu, "An efficient helicopter-satellite communication scheme based on check-hybrid ldpc coding," *Tsinghua Science and Technology*, 2018.

[17] D. Klinc, J. Ha, S. W. McLaughlin, J. Barros, and B.-J. Kwak, "LDPC codes for the Gaussian wiretap channel," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 532–540, 2011.

[18] M. Baldi, M. Bianchi, and F. Chiaraluce, "Non-systematic codes for physical layer security," *IEEE Information Theory Workshop*, pp. 1–5, 2010.

[19] M. Baldi, M. Bianchi, and F. Chiaraluce, "Coding with scrambling, concatenation, and HARQ for the AWGN wire-tap channel: A security gap analysis," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 883–894, 2012.

[20] Z. Chen, L. Yin, Y. Pei, and J. Lu, "CodeHop: physical layer error correction and encryption with LDPC-based code hopping," *Science China Information Sciences*, vol. 59, no. 10, Article ID 102309, pp. 1–15, 2016.

[21] Y. Fang, G. Bi, Y. L. Guan, and F. C. M. Lau, "A survey on protograph LDPC codes and their applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 1989–2016, 2015.

[22] Y. Fang, S. C. Liew, and T. Wang, "Design of distributed protograph LDPC Codes for multi-relay coded-cooperative networks," *IEEE Transactions on Wireless Communications*, pp. 7235–7251, 2017.

[23] H. Tyagi and A. Vardy, "Universal hashing for information-theoretic security," *Proceedings of the IEEE*, vol. 103, no. 10, pp. 1781–1795, 2015.

[24] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, "Generalized privacy amplification," *IEEE Transactions on Information Theory*, vol. 41, no. 6, part 2, pp. 1915–1923, 1995.

[25] M. Hayashi, "Exponential decreasing rate of leaked information in universal random privacy amplification," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3989–4001, 2011.

[26] M. Hayashi and T. Tsurumaru, "More efficient privacy amplification with less random seeds via dual universal hash function," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 2213–2232, 2016.

[27] J. Thorpe, "Low-density parity-check (ldpc) codes constructed from protographs," *IPN progress report*, vol. 42, no. 154, pp. 42–154, 2003.

[28] S. Ten Brink and G. Kramer, "Design of repeat-accumulate codes for iterative detection and decoding," *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2764–2772, 2003.

[29] T. J. Richardson and R. L. Urbanke, "Efficient encoding of low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 638–656, 2001.

[30] A. Bogdanov, M. C. Mertens, C. Paar, J. Pelzl, and A. Rupp, "Smith-a parallel hardware architecture for fast gaussian elimination over gf (2)," in *Workshop on Special-Purpose Hardware for Attacking Cryptographic Systems*, 2006.

[31] Z. Li, L. Chen, L. Zeng, S. Lin, and W. H. Fong, "Efficient encoding of quasi-cyclic low-density parity-check codes," *IEEE Transactions on Communications*, vol. 54, no. 1, pp. 71–81, 2006.

[32] Q. Huang, L. Tang, S. He, Z. Xiong, and Z. Wang, "Low-complexity encoding of quasi-cyclic codes based on Galois Fourier transform," *IEEE Transactions on Communications*, vol. 62, no. 6, pp. 1757–1767, 2014.

[33] Y.-L. Ueng, B.-J. Yang, C.-J. Yang, H.-C. Lee, and J.-D. Yang, "An efficient multi-standard LDPC decoder design using hardware-friendly shuffled decoding," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 3, pp. 743–756, 2013.

[34] Q. Huang, L. Song, and Z. Wang, "Set message-passing decoding algorithms for regular non-binary LDPC codes," *IEEE Transactions on Communications*, vol. 65, no. 12, pp. 5110–5122, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 6101853, 9 pages
https://doi.org/10.1155/2018/6101853

*Research Article*

# Research and Implementation of Rateless Spinal Codes Based Massive MIMO System

**Liangliang Wang**[1,2] **Xiang Chen**[1,2] **and Hongzhou Tan**[1]

[1]*School of Electronic and Information Technology, Sun Yat-sen University, Guangzhou 510006, China*
[2]*Key Lab of EDA, Research Institute of Tsinghua University in Shenzhen (RITS), Shenzhen 518075, China*

Correspondence should be addressed to Xiang Chen; chenxiang@mail.sysu.edu.cn

Received 23 November 2017; Revised 1 February 2018; Accepted 28 February 2018; Published 2 April 2018

Academic Editor: Zesong Fei



The potential performance gains promised by massive multi-input and multioutput (MIMO) rely heavily on the access to accurate channel state information (CSI), which is difficult to obtain in practice when channel coherence time is short and the number of mobile users is huge. To make the system with imperfect CSI perform well, we propose a rateless codes-aided massive MIMO scheme, with the aim of approaching the maximum achievable rate (MAR) as well as improving the achieved rate over that based on the fixed-rate codes. More explicitly, a recently proposed family of rateless codes, called spinal codes, are applied to massive MIMO systems, where the spinal codes bring the benefit of approximately achieving the MAR with sufficiently large encoding block size. In addition, the multilevel puncturing and dynamic block-size allocation (MPDBA) scheme is proposed, where the block sizes are determined by user MAR to curb the average retransmission delay for successfully decoding the messages, which further enhances the system retransmission efficiency. Multilevel puncturing, which is MAR dependent, narrows the gap between the system MAR and the related achieved rate. Theoretical analysis is provided to demonstrate that spinal codes with the MPDBA can guarantee the system retransmission efficiency as well as achieved rate, which are also verified by numerical simulations. Finally, a simplified but comparable MIMO testbed with 2 transmit antennas and 2 single-antenna users, based on NI Universal Software Radio Peripheral (USRP) and LabVIEW communication toolkits, is built up to demonstrate the effectiveness of our proposal in realistic wireless channels, which is easy to be extended to massive MIMO scenarios in future.

## 1. Introduction

Massive multi-input and multioutput (MIMO), achieving high spectral efficiency and low power consumption, has been widely regarded as a promising technique for 5G wireless communication systems. However, its benefits rely heavily on the accuracy of the channel state information (CSI) to perform the multiuser procoding. Unfortunately, the collection of accurate CSI is costly because of the short channel coherence time and the huge number of mobile users. In fact, pilot contamination is an essential factor to result in imperfect CSI, and pilot contamination appears to have a more profound effect than classical MIMO [1, 2]. Therefore, a critical question is how to improve the throughput performance of massive MIMO systems under imperfect CSI. Traditional fixed-rate codes, such as LDPC [3] and Turbo codes [4], may suffer from the significant throughput loss resulting from the rate mismatching under inaccurate CSI. Therefore, developing a resilient transmission scheme for massive MIMO in the presence of imperfect CSI is of great importance.

In fact, for multiuser cellular systems, where the base station is not equipped with a large number of antennas, it has been proved that rateless codes perform well under inaccurate CSI in [5, 6]. Therefore, rateless codes based transmission scheme for massive MIMO is considered in this paper. When CSI is not available at transmitter, rateless codes with adaptive code rates perform well; some related works about rateless codes, such as spinal codes, strider codes, and raptor codes, have been presented in [7–9]. In [7], the authors proved that spinal codes outperform LDPC codes as well as strider and raptor codes in fading channels with inaccurate CSI. Therefore, spinal codes are integrated into resilient transmission scheme, where spinal encoders encode users'

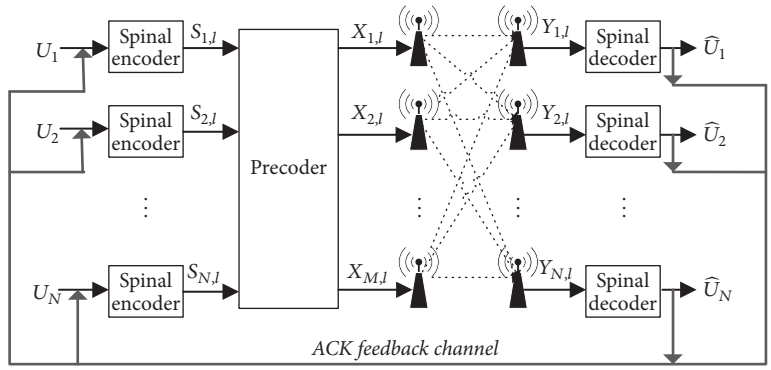



Figure 1: Spinal codes based massive MIMO systems.

original messages into multiple infinite streams of symbols, which are transmitted continuously through numbers of (re)transmissions, called passes, until the senders receive the acknowledge (ACK) indicating the successful decoding, from the receivers. Thus, the system will benefit from the multi-retransmission diversity and achieves a robust performance. Once the encoding block sizes for spinal codes are allowed to be sufficiently large, users will gradually approach their MAR with maximum-likelihood (ML) decoding algorithm [10].

For classical MIMO, nonlinear precoding techniques, such as dirty-paper-coding (DPC) [11], vector perturbation (VP) [12], and lattice-aided [13], have better performance. However, with antennas increasing at the base station in massive MIMO, linear precoders, such as zero forcing (ZF) [14], have limited throughput loss compared with nonlinear precoders [15]. ZF has low complexity and is more practical for massive MIMO. Therefore, ZF is considered in our work to mitigate multiuser interferences.

However, in order to guarantee the system retransmission efficiency (measured in bits per symbol per second), reducing the retransmission delay, which is equivalent to reducing the pass number, should be considered before spinal codes are integrated into massive MIMO. Reference [16] studied this problem from the link-layer; however, user-schedule scheme involved in the link-layer will bring some extra complexity. To that end, an easily implemented scheme in the physic-layer is proposed in this paper. From [10], it is known that the pass number is proportional to the block size while being inversely proportional to the maximum achievable rate (MAR). Therefore, once we initialize the encoders for all users with a large enough static block size for MAR-approaching purpose, the users with lower MARs, locating in the cell-edge or trapping in a deep fading channel, will significantly enlarge the average pass number for decoding, which further degrades the system retransmission efficiency performance. Because of this, we developed a multilevel puncturing and dynamic block-size allocation (MPDBA) scheme, where the MARs are obtained as a priori knowledge to determine the block sizes dynamically, which can reduce the pass number as well as retransmission delay. Different puncturing method is implemented for different MARs, which is proved to guarantee the system retransmission efficiency and achieved rate. The numerical simulation results show that spinal codes with MPDBA can make massive MIMO with imperfect CSI work reliably with efficient retransmission.

Moreover, in order to make the spinal codes based massive MIMO be practical from theory, we also consider building up a system with 2 transmit antennas and 2 single-antennas users, where NI USRP and LabVIEW communication toolkits are involved as the hardware and software platforms, respectively. In this implementation demo system, the retransmission efficiency by our proposal is verified in the fading environments. This demo can also be fast extended to massive MIMO scenarios by massive hardware scale expansion.

The rest of this paper is organized as follows: Section 2 presents the system model. The efficient transmission scheme for the system with spinal codes is proposed in Section 3. Section 4 shows some simulation results to illustrate the benefits of the proposed scheme, Section 5 presents some details about the USRP based spinal codes MIMO demo system. Finally, conclusions are drawn in Section 6.

## 2. System Model

Throughout this paper, uppercase boldface letters are used to denote matrices. $(\bullet)^H$ and $(\bullet)^T$ represent the Hermitian transpose and transpose, respectively. $E[\bullet]$ denotes the expectation operator. $\mathrm{Tr}(\bullet)$ stands for trace, $[\bullet]$ is the round to integer operator, and $\lceil \bullet \rceil$ is the round up to integer operator.

A $N \times M$ downlink massive MIMO system is shown in Figure 1, where the single-cell system has a base station with $M$ antennas and $N$ single-antenna users, $M \gg N$. For user $n$, $n = 1, 2, \ldots, N$, $U_n$ is used to denote an unexpected message. Spinal encoder [7], illustrated in Figure 2, divides $U_n$, denoted by message $I$, $I = 1, 2, \ldots$, into $B$ blocks with size $K_n(I)$ bits; hash function uses each block to generate a sequence of spine values; then the modulated symbols are yielded by a mapper and pseudorandom number generator (RNG) function, which uses spine values as its seeds. After encoding, ZF maps the modulated symbols to the precoded symbols, which are transmitted over fading channels. If decoding happened successfully, the receiver will send the ACK to the corresponding sender through the feedback channel; otherwise, the sender will generate more redundant symbols to transmit in the following passes until receiving the ACK.

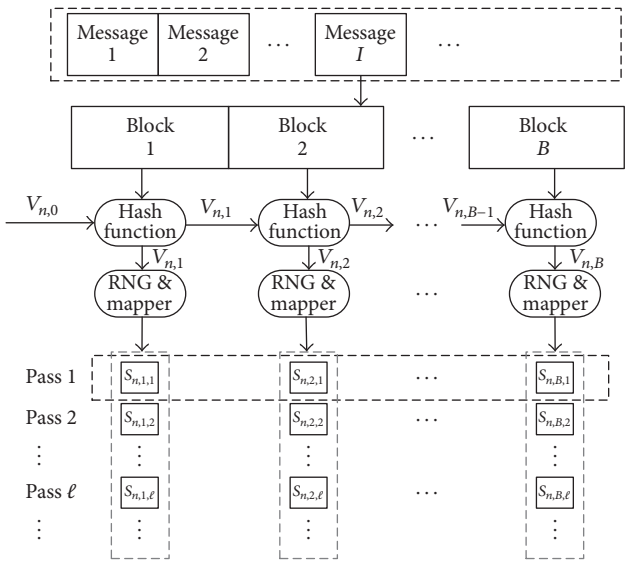



Figure 2: Structure of spinal encoder.

Given $L_n(I) \geq 1$ is the pass number for user $n$ to successfully decode the message $I$. At $l$th pass, for $l = 1, 2, \ldots, L_n(I)$, all users' modulated symbols are mapped to a $N \times B$ matrix of symbols $\mathbf{S}_l = (\mathbf{S}_{1l}, \mathbf{S}_{2l}, \ldots, \mathbf{S}_{Bl})$. $\mathbf{S}_{il} = (S_{1il}, S_{2il}, \ldots, S_{Nil})^T$, for $i = 1, 2, \ldots, B$, follows Gaussian distribution and $\mathrm{Tr}(\mathbf{S}_{il}\mathbf{S}_{il}^H) = P_l$, $P_l$ is total transmit power. Assuming CSI is prior knowledge obtained by channel estimation methods at transmitter, and downlink MIMO channel remains ergodic and stationary during $B$ symbols periods, $\mathbf{S}_l$ is then precoded to transmit signal $\mathbf{X}_l \in \mathbb{C}^{M\times B}$ by

$$\mathbf{X}_l = \beta_l \mathbf{W}_l \mathbf{S}_l, \tag{1}$$

where $\mathbf{W}_l = \widehat{\mathbf{H}}_l^H(\widehat{\mathbf{H}}_l\widehat{\mathbf{H}}_l^H)^{-1}$ denotes the ZF precoding matrix for $l$th pass, $\widehat{\mathbf{H}}_l \in \mathbb{C}^{N\times M}$ denotes the estimated channel matrix, and $\beta_l = \sqrt{N/\mathrm{Tr}(\mathbf{W}_l\mathbf{W}_l^H)}$ is power control factor used to normalize $\widehat{\mathbf{W}}_l$ such that $\mathrm{Tr}(\mathbf{X}_{il}\mathbf{X}_{il}^H) = P_l$.

$\mathbf{X}_l$ is then transmitted through a MIMO fading channel, and the received signal matrix $\mathbf{Y}_l \in \mathbb{C}^{N\times B}$ is given by

$$\mathbf{Y}_l = \beta_l^{-1}\mathbf{H}_l\mathbf{X}_l + \beta_l^{-1}\mathbf{N}_l, \tag{2}$$

where $\mathbf{H}_l = \mathbf{D}_l^{1/2}\mathbf{G}_l$ denotes the MIMO channel matrix, $\mathbf{G}_l \in \mathbb{C}^{N\times M}$ and $\mathbf{D}_l \in \mathbb{R}^{N\times N}$ represent the complex small-scale and large-scale fading coefficients matrix, respectively. Assuming that $\mathbf{G}_l$ has zero mean and unit variance independent and identically distributed (i.i.d.) complex Gaussian entries, large-scale fading coefficients, denoted as $\mathbf{D}_l = \mathrm{diag}(d_{1l}, d_{2l}, \ldots, d_{Nl})$, are the same for different antennas but user-dependent. $\mathbf{N}_l \in \mathbb{C}^{N\times B}$ denotes the noise matrix with i.i.d. complex Gaussian random variables with zero mean and unit variance. Then the average transmit SNR is $P_l/N$.

If perfect CSI is available at transmitter, then $\widehat{\mathbf{H}}_l = \mathbf{H}_l$. However, imperfect CSI always arises in any practical estimation schemes. According to the channel estimation model described in [14], $\widehat{\mathbf{H}}_l$ can be given by

$$\widehat{\mathbf{H}}_l = \mathbf{H}_l + \delta_e\Delta\mathbf{H}_l, \tag{3}$$

where $\Delta\mathbf{H}_l \in \mathbb{C}^{N\times M}$ denotes the channel estimation error matrix, with zero mean and unit variance i.i.d. complex Gaussian random variables uncorrelated with that of $\mathbf{H}_l$, and $\delta_e$ is a nonzero parameter to measure the quality of channel estimation and is appropriately chosen depending on the channel dynamics. Since the focus of this paper is to study the performance of spinal codes, $\delta_e$ is assumed to be known at transmitter.

Based on (3), the ZF precoding matrix can be expressed as

$$\begin{aligned}\mathbf{W}_l = {} & \beta_l\left(\mathbf{H}_l + \delta_e\Delta\mathbf{H}_l\right)^H \\ & \cdot\left[\left(\mathbf{H}_l + \delta_e\Delta\mathbf{H}_l\right)\left(\mathbf{H}_l + \delta_e\Delta\mathbf{H}_l\right)^H\right]^{-1},\end{aligned} \tag{4}$$

substituting (4) into (2), the $i$th symbol vector at $l$th pass is given by

$$\mathbf{Y}_{il} = \mathbf{S}_{il} + \beta_l^{-1}\mathbf{N}_{il} - \delta_e\Delta\mathbf{H}_l\widehat{\mathbf{H}}_l^H\left(\widehat{\mathbf{H}}_l\widehat{\mathbf{H}}_l^H\right)^{-1}\mathbf{S}_{il}, \tag{5}$$

where $\delta_e\Delta\mathbf{H}_l\widehat{\mathbf{H}}_l^H(\widehat{\mathbf{H}}_l\widehat{\mathbf{H}}_l^H)^{-1}\mathbf{S}_{il}$ is the additional interference item brought by the channel estimation error.

Once $l = L_n(I)$, spinal decoder for user $n$ collects enough mutual information to retrieve the message $I$ successfully [7]. Then the user current data rate (measured in bits per symbol) is achieved by

$$R_n(I) = \frac{K_n(I)}{L_n(I)}, \tag{6}$$

based on (6), the average achieved rate of spinal codes based massive MIMO is given by $R_{\text{achieved}} = \lim_{I\to\infty}(1/I)\sum_{I=1}^{\infty}\sum_{n=1}^{N}R_n(I)$.

Based on (5), assuming that $P_{nl}$ is the average signal power for the desired symbols $S_{nil} \in \mathbf{S}_{il}$, we have $\sum_{n=1}^{N}P_{nl} = \mathrm{Tr}(\mathbf{S}_{il}\mathbf{S}_{il}^H) = P_l$. Denoting

$$\widehat{\mathbf{N}}_{il} = \beta_l^{-1}\mathbf{N}_{il} - \delta_e\Delta\mathbf{H}_l\widehat{\mathbf{H}}_l^H\left(\widehat{\mathbf{H}}_l\widehat{\mathbf{H}}_l^H\right)^{-1}\mathbf{S}_{il} \tag{7}$$

as the received noise, the covariance for $\widehat{\mathbf{N}}_{il}$ can be proved to be

$$E\left[\widehat{\mathbf{N}}_{il}\widehat{\mathbf{N}}_{il}^H\right] = \beta_l^{-2}\mathbf{I}_N + \boldsymbol{\lambda}_l^{-1}P_l\delta_e^2, \tag{8}$$

where $\boldsymbol{\lambda}_l = \mathrm{diag}(\lambda_{1l}, \ldots, \lambda_{Nl})$ is the singular value of $\widehat{\mathbf{H}}_l\widehat{\mathbf{H}}_l^H$ and is approximated by

$$\boldsymbol{\lambda}_l \approx M\left(\mathbf{D}_l + \delta_e^2\mathbf{I}_N\right). \tag{9}$$

*Proof.* Assuming that $M \gg N$ in massive MIMO system, let the singular value decomposition (SVD) of matrix $\widehat{\mathbf{H}}_l\widehat{\mathbf{H}}_l^H$ be $\widehat{\mathbf{H}}_l\widehat{\mathbf{H}}_l^H = \mathbf{U}_l\boldsymbol{\lambda}_l\mathbf{V}_l^H$, where $\mathbf{U}_l$ and $\mathbf{V}_l$ are unitary matrix and $\boldsymbol{\lambda}_l$ is a diagonal matrix.

Given $\widehat{\mathbf{N}}_{il} = \beta_l^{-1}\mathbf{N}_{il} - \delta_e\Delta\mathbf{H}_l\widehat{\mathbf{H}}_l^H(\widehat{\mathbf{H}}_l\widehat{\mathbf{H}}_l^H)^{-1}\mathbf{S}_{il}$, the covariance of $\widehat{\mathbf{N}}_{il}$ can be computed as

$$\begin{aligned} E\left[\widehat{\mathbf{N}}_{il}\widehat{\mathbf{N}}_{il}^H\right] &= \beta_l^{-2}E\left[\mathbf{N}_{il}\mathbf{N}_{il}^H\right] \\ &+ \delta_e^2 E\left[\Delta\mathbf{H}_l\widehat{\mathbf{H}}_l^H\left(\widehat{\mathbf{H}}_l\widehat{\mathbf{H}}_l^H\right)^{-1}\mathbf{S}_{il}\mathbf{S}_{il}^H\left(\widehat{\mathbf{H}}_l\widehat{\mathbf{H}}_l^H\right)^{-1}\right. \\ &\left.\cdot\,\widehat{\mathbf{H}}_l\Delta\mathbf{H}_l^H\right] = \beta_l^{-2}\mathbf{I}_N + \boldsymbol{\lambda}_l^{-1}P_l\delta_e^2, \end{aligned} \tag{10}$$

where we use the fact that $E[\Delta\mathbf{H}_l\widehat{\mathbf{H}}_l^H(\widehat{\mathbf{H}}_l\widehat{\mathbf{H}}_l^H)^{-1}\mathbf{S}_{il}\mathbf{N}_{il}^H] = 0$ and $E[\mathbf{N}_{il}\mathbf{N}_{il}^H] = \mathbf{I}_N$.

In massive MIMO, when $M \gg N$, $M \to \infty$, the rows of $\mathbf{G}_l$ are asymptotically orthogonal; that is, $\mathbf{G}_l\mathbf{G}_l^H \approx M\mathbf{I}_N$. Hence we have $\widehat{\mathbf{H}}_l\widehat{\mathbf{H}}_l^H = (\mathbf{D}_l^{1/2}\mathbf{G}_l + \delta_e\Delta\mathbf{H}_l)(\mathbf{D}_l^{1/2}\mathbf{G}_l + \delta_e\Delta\mathbf{H}_l)^H \approx M(\mathbf{D}_l + \delta_e^2\mathbf{I}_N)$ and $\boldsymbol{\lambda}_l \approx M(\mathbf{D}_l + \delta_e^2\mathbf{I}_N)$. □

Based on (8) and (9), the average MAR for user $n$ is given by

$$R_n^{\text{upper}} = E\left[\log_2\left(1 + \frac{M\beta_l^2P_{nl}}{M + \beta_l^2P_l\delta_e^2\left(d_{nl} + \delta_e^2\right)^{-1}}\right)\right]. \tag{11}$$

Thus, the upper bound of ergodic achievable rate for massive MIMO with ZF under imperfect CSI is expressed as $R_{\text{upper}} = \sum_{n=1}^{N} R_n^{\text{upper}}$.

**Lemma 1.** *If $K_n(I)$ is sufficiently large and $K_n(I) \geq R_n^{upper}$, then there exists $(L_1(I), L_2(I), \ldots, L_N(I))$ for users to achieve $R_n^{upper} - R_n(I) = O((R_n^{upper})^2/K_n(I))$.*

*Proof.* From [10], $L_n(I)$ can be expressed as

$$L_n(I) = \left\lceil \frac{K_n(I)}{R_n^{\text{upper}}} \right\rceil + O(1). \tag{12}$$

Substituting (12) into (6), the corresponding data rate is rewritten as

$$R_n(I) = \frac{K_n(I)}{\lceil K_n(I)/R_n^{\text{upper}} \rceil + O(1)}. \tag{13}$$

Based on the fact that $[K_n(I)/R_n^{\text{upper}}] \leq K_n(I)/R_n^{\text{upper}}$, we have

$$\begin{aligned} R_n(I) &= \frac{K_n(I)}{\lceil K_n(I)/R_n^{\text{upper}} \rceil + O(1)} \\ &= \frac{K_n(I)}{\lceil K_n(I)/R_n^{\text{upper}} \rceil + 1 + O(1)} \\ &\geq \frac{K_n(I)}{K_n(I)/R_n^{\text{upper}} + 1 + O(1)} \\ &= \frac{R_n^{\text{upper}}}{1 + R_n^{\text{upper}}(1 + O(1))/K_n(I)}, \end{aligned} \tag{14}$$

when $K_n(I) \to \infty$, $R_n^{\text{upper}}(1 + O(1))/K_n(I) \to 0$, $R_n^{\text{upper}}/(1 + R_n^{\text{upper}}(1+O(1))/K_n(I)) \approx R_n^{\text{upper}}(1 - R_n^{\text{upper}}(1+O(1))/K_n(I))$. Therefore, we can obtain $R_n^{\text{upper}}(1 - R_n^{\text{upper}}(1+O(1))/K_n(I)) \leq R_n(I) \leq R_n^{\text{upper}}$, which can be further expressed as

$$0 \leq R_n^{\text{upper}} - R_n(I) \leq \frac{\left(R_n^{\text{upper}}\right)^2(1 + O(1))}{K_n(I)}. \tag{15}$$

From (15), there exists $c_1$ to satisfy $R_n^{\text{upper}} - R_n(I) \leq c_1((R_n^{\text{upper}})^2/K_n)$, and we can obtain $R_n^{\text{upper}} - R_n(I) = O((R_n^{\text{upper}})^2/K_n)$, where $O(\cdot)$ denotes the upper bound of $R_n^{\text{upper}} - R_n(I)$. □

This lemma indicates that when block size $K_n(I) = K_{\max}$, where $K_{\max}$ denotes the sufficiently large block size in this paper, the gap between $R_n(I)$ and $R_n^{\text{upper}}$ can be arbitrarily small such that each user can approach $R_n^{\text{upper}}$ with $L_n(I)$, and spinal codes based massive MIMO approaches $R_{\text{upper}}$ as well.

When the small large-scale fading coefficients $d_{nl}$ deteriorate $R_n^{\text{upper}}$ in (11), based on (12), $K_n(I) = K_{\max}$ will enlarge $L_n(I)$ significantly. If each pass occupies the constant time $T$ (measured in seconds), then the larger $L_n(I)$ makes the average retransmission delay spent on decoding the message $I$ be costly, and the average system retransmission efficiency, defined as

$$r = \lim_{I\to\infty}\frac{1}{I}\sum_{I=1}^{\infty}\sum_{n=1}^{N}\frac{R_n(I)}{TL_n(I)}, \tag{16}$$

will be limited as well. We use $r_n(I) = R_n(I)/TL_n(I)$ to denote current retransmission efficiency for each user.

The greedy idea is an effective way to achieve a preferable system retransmission efficiency across different SNRs. That is, we try to make sure that $r_n(I)$ for each message $I$ is better, which will further achieve a satisfied value for $r$. When $R_n(I)$ is ideal, to decrease $L_n(I)$ is an effective way to enhance $r_n(I)$. Therefore, based on (12), $K_n(I)$ can be dynamically determined by $R_n^{\text{upper}}$ to reduce $L_n(I)$.

## 3. Multilevel Puncturing and Dynamic Block-Size Allocation Scheme

In this section, a multilevel puncturing and dynamic block-size allocation scheme is proposed. For each user, the dynamic block-size allocation problem is formulated as

$$K_n^*(I) = \underset{R_n^{\text{upper}} \leq K_n(I) < K_{\max}}{\arg\min} L_n(I), \tag{17}$$

where $K_n^*(I)$ is the desired block size for message $I$.

However, we cannot obtain the solution for (17) by solving (12) directly, which includes an unexpected constant item $O(1)$. Instead, we use the cumulative distribution function (CDF) of pass numbers $L_n(I)$ to determine $K_n^*(I)$. Figure 3 presents the CDF of pass numbers corresponding to the different block sizes and SNRs, from which we obtain that, under different SNRs, smaller block sizes guarantee the system to decode successfully with less pass number and high probability. Therefore, we choose

$$K_n^*(I) = \lceil R_n^{\text{upper}} \rceil \tag{18}$$

for each spinal encoder for encoding the message $I$.

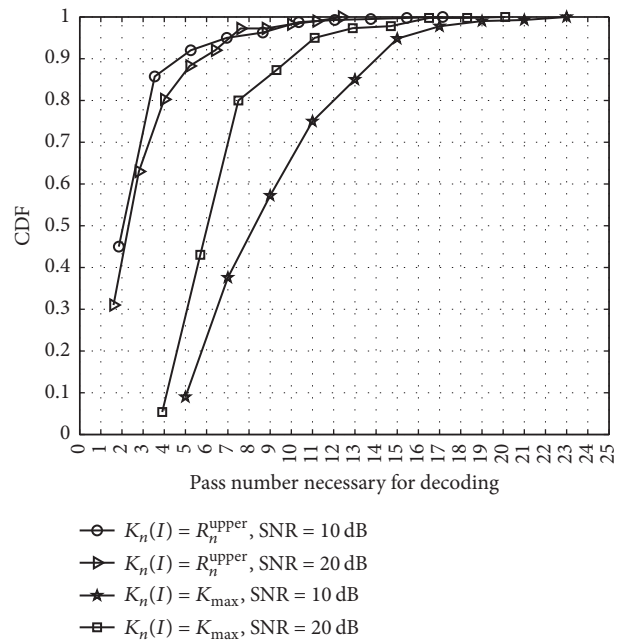



Figure 3: CDF curves of pass number needed for successfully decoding under different SNRs, $K_{\max} > R_n^{\text{upper}}$, $\delta_e = 0.1$.

When $K_n(I) = K_n^*(I)$, according to the proof of Lemma 1, the gap between the MAR and the achieved rate is given by

$$R_n^{\text{upper}} - R_n(I) = O\left(R_n^{\text{upper}}\right), \tag{19}$$

which indicates that the dynamic block-size scheme will cause an unavoidable achieved-rate performance loss at higher $R_n^{\text{upper}}$. To reduce the loss as much as possible, a multilevel puncturing method is implemented for spinal codes.

Assume message $I$ with $m$ bits is encoded by spinal encoder with desired encoding block size $K_n(I) = K_n^*(I)$ and yields $B = m/K_n(I)$ coded symbols. If these $B$ symbols transmitted in the first pass, that is, $l = 1$, cannot be used to retrieve message successfully, then, from the second transmit pass, that is, $l > 1$, each $B$ symbol is transmitted within $L_{\text{sub},n}(I)$ subpasses with $m/K_n(I)L_{\text{sub},n}(I)$ symbols transmitted at each subpass.

When $l = L_{\text{sub},n}(I)$, $m$ bits can be retrieved successfully by $m/K_n(I) + (L_n(I) - 1)(m/K_n(I)L_{\text{sub},n}(I))$ symbols, and the user $n$ current rate in (6) can be rewritten as

$$R_n(I) = \frac{K_n(I) L_{\text{sub},n}(I)}{L_{\text{sub},n}(I) + L_n(I) - 1}, \tag{20}$$

where $L_{\text{sub},n}(I) \in \{1, 2, \ldots, B\}$ and is determined according to the practical cases.

**Theorem 2.** *Consider that the symbols of MPDBA based spinal codes with parameter $K_n^*(I)$, $P_l$ are transmitted over a MIMO channel with imperfect CSI, then the system can achieve $R_n^{upper} - R_n(I) = \theta(R_n^{upper}/L_{sub,n}(I))$ with $r_n(I) \geq \theta((R_n^{upper})^2/L_{sub,n}(I))$.*

*Proof.* From [10], $L_n(I)$ under MPDBA scheme is given by

$$L_n(I) = \left\lceil \frac{K_n^*(I) L_{\text{sub},n}(I)}{R_n^{\text{upper}}} \right\rceil - L_{\text{sub},n}(I) + 1 + O(1). \tag{21}$$

Substituting (21) into (20), the current data rate under MPDBA scheme is achieved as

$$R_n(I) = \frac{K_n^*(I) L_{\text{sub},n}(I)}{\left\lceil K_n^*(I) L_{\text{sub},n}(I)/R_n^{\text{upper}} \right\rceil + O(1)}, \tag{22}$$

when $K_n^*(I) = \lceil R_n^{\text{upper}} \rceil$, (22) can be further written by

$$R_n(I) = \frac{\left\lceil R_n^{\text{upper}} \right\rceil L_{\text{sub},n}(I) + L_{\text{sub},n}(I)}{\left[\left\lceil R_n^{\text{upper}} \right\rceil L_{\text{sub},n}(I)/R_n^{\text{upper}} + L_{\text{sub},n}(I)/R_n^{\text{upper}}\right] + 1 + O(1)}. \tag{23}$$

Based on the fact that

$$\begin{aligned} &\frac{\left\lceil R_n^{\text{upper}} \right\rceil L_{\text{sub},n}(I)}{R_n^{\text{upper}}} + \frac{L_{\text{sub},n}(I)}{R_n^{\text{upper}}} - 1 \\ &< \left\lceil \frac{\left\lceil R_n^{\text{upper}} \right\rceil L_{\text{sub},n}(I)}{R_n^{\text{upper}}} + \frac{L_{\text{sub},n}(I)}{R_n^{\text{upper}}} \right\rceil \\ &\leq \frac{\left\lceil R_n^{\text{upper}} \right\rceil L_{\text{sub},n}(I)}{R_n^{\text{upper}}} + \frac{L_{\text{sub},n}(I)}{R_n^{\text{upper}}}, \end{aligned} \tag{24}$$

we can obtain

$$\begin{aligned} &\frac{\left\lceil R_n^{\text{upper}} \right\rceil L_{\text{sub},n}(I) + L_{\text{sub},n}(I)}{\left\lceil R_n^{\text{upper}} \right\rceil L_{\text{sub},n}(I)/R_n^{\text{upper}} + L_{\text{sub},n}(I)/R_n^{\text{upper}} + 1 + O(1)} \\ &\quad \leq R_n(I) \\ &\quad < \frac{\left\lceil R_n^{\text{upper}} \right\rceil L_{\text{sub},n}(I) + L_{\text{sub},n}(I)}{\left\lceil R_n^{\text{upper}} \right\rceil L_{\text{sub},n}(I)/R_n^{\text{upper}} + L_{\text{sub},n}(I)/R_n^{\text{upper}} + O(1)}. \end{aligned} \tag{25}$$

From (25), we have

$$\begin{aligned} \frac{R_n^{\text{upper}} O(1)}{L_{\text{sub},n}(I)(2 + O(1))} &< R_n^{\text{upper}} - R_n(I) \\ &< \frac{R_n^{\text{upper}}(1 + O(1))}{L_{\text{sub},n}(I)}. \end{aligned} \tag{26}$$

From (26), there exist $c_2$ and $c_3$ to satisfy $c_2(R_n^{\text{upper}}/L_{\text{sub},n}(I)) \leq R_n^{\text{upper}} - R_n(I) \leq c_3(R_n^{\text{upper}}/L_{\text{sub},n}(I))$. Thus we have

$$R_n^{\text{upper}} - R_n(I) = \theta\left(\frac{R_n^{\text{upper}}}{L_{\text{sub},n}(I)}\right), \tag{27}$$

where $\theta(\cdot)$ is used to denote the lower and upper bound of $R_n^{\text{upper}} - R_n(I)$.

From (21), the pass number satisfies $1 + O(1) < L_n(I) \leq L_{\text{sub},n}(I)/R_n^{\text{upper}} + 2 + O(1)$; therefore we have

$$L_n(I) = O\left(\frac{L_{\text{sub},n}(I)}{R_n^{\text{upper}}}\right). \tag{28}$$

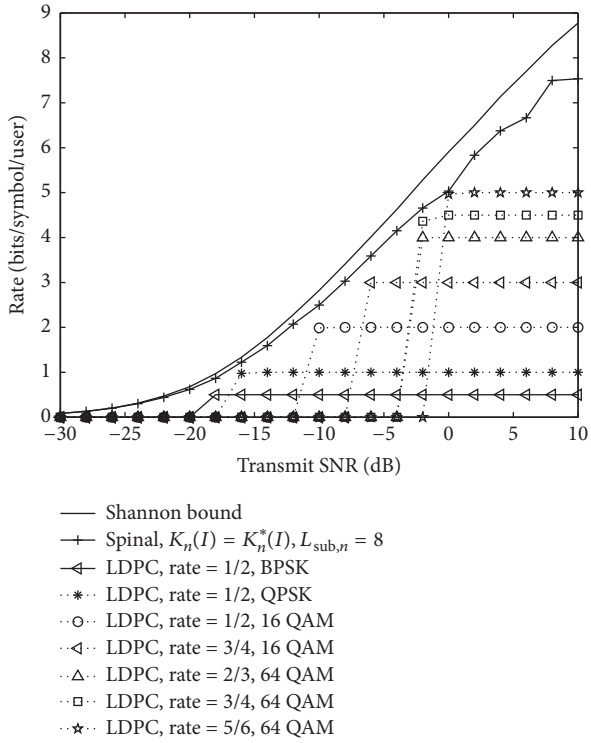



Figure 4: Average rates achieved by spinal codes and LDPC codes based massive MIMO, $\delta_e = 0.1$.

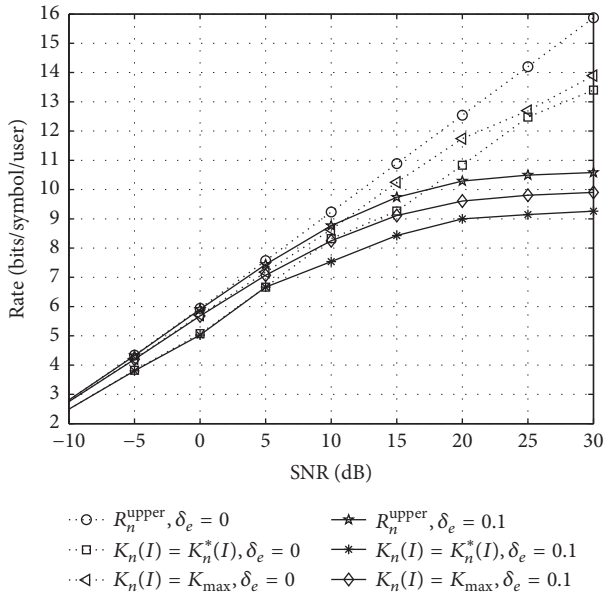



Figure 5: Average rates achieved by spinal codes based massive MIMO, $K_{\max} > K_n^*, L_{\text{sub},n}(I) = 8$.

Based on (27) and (28), there exist $c_4$ and $c_5$ such that the minimum achieved data rate and maximum pass number can be given by $R_n^{\text{upper}} - c_4(R_n^{\text{upper}}/L_{\text{sub},n}(I))$ and $c_5(L_{\text{sub},n}(I)/R_n^{\text{upper}})$, respectively. Therefore, the lower bound of $r_n(I)$ is given by

$$r_n^{\text{low}}(I) = \frac{1}{c_5}\frac{\left(R_n^{\text{upper}}\right)^2}{L_{\text{sub},n}(I)} - \frac{c_4}{c_5}\frac{\left(R_n^{\text{upper}}\right)^2}{\left(L_{\text{sub},n}(I)\right)^2}, \tag{29}$$

which can be written as $r_n^{\text{low}}(I) = \theta((R_n^{\text{upper}})^5/L_{\text{sub},n}(I))$. □

This theorem indicates that for the transmission scheme in massive MIMO with MPDBA based spinal codes, we can choose a proper $L_{\text{sub},n}(I)$ (i.e., lower $L_{\text{sub},n}(I)$ for lower $R_n^{\text{upper}}$ and higher $L_{\text{sub},n}(I)$ for higher $R_n^{\text{upper}}$) for different $R_n^{\text{upper}}$ such that the gap between the achieved rate and the MAR is limited as well as better retransmission efficiency is guaranteed under different $L_{\text{sub},n}(I)$.

## 4. Simulation Results and Analysis

In this section, some numerical simulation results are illustrated to verify the performance of MPDBA based spinal codes scheme for a $4 \times 64$ MIMO system, $K_{\max} = 20$.

Figure 4 compares the average achieved-rates of massive MIMO based on spinal codes and LDPC codes, respectively. LDPC codes are from 802.16e, decoded with a powerful decoder. We can obtain that spinal codes outperform LDPC codes across all SNRs. Moreover, the simple decoder structure of spinal codes also avoids the demapping complexity caused by different constellations mapping operations of LDPC codes.

In Figure 5, we compare the average achieved-rates of spinal codes based massive MIMO with $K_n(I) = K_{\max}$ and $K_n(I) = K_n^*(I)$, respectively. Massive MIMO can approach its MAR gradually by enlarging the block size, which verifies the conclusion in Lemma 1. We also obtain that the achieved-rate performance of spinal codes with $K_n(I) = K_n^*(I)$ follows conclusions in Theorem 2; that is, larger $R_n^{\text{upper}}$ will cause bigger achieved-rate performance loss, especially for lower $L_{\text{sub},n}(I)$.

Figure 6 shows that we can enlarge $L_{\text{sub},n}(I)$ to minimize the gap between the MARs and the rates achieved by spinal codes with MPDBA. For lower $R_n^{\text{upper}}$, smaller $L_{\text{sub},n}(I)$ can achieve a satisfying achieved-rates result. For higher $R_n^{\text{upper}}$, a larger $L_{\text{sub},n}(I)$ is necessary for spinal codes to achieve the MAR.

Figure 7 shows that the application of MPDBA for spinal encoders will enhance the system retransmission efficiency. Different proper $L_{\text{sub},n}(I)$ for the MPDBA based spinal codes can achieve preferable retransmission efficiency, which also indicates that once the system achieves better retransmission efficiency, there will be more proper $L_{\text{sub},n}(I)$ to choose from to obtain better achieved-rates performance illustrated in Figure 6. The numerical results in Figures 6 and 7 verify Theorem 2; that is, spinal codes with MPDBA can guarantee the system reliability as well as retransmission efficiency.

## 5. Simplified Spinal Codes Based MIMO Demo System with NI USRP 2920

This section will present some details about making spinal codes based massive MIMO practical from theory. Some works about implementation of massive MIMO prototyping

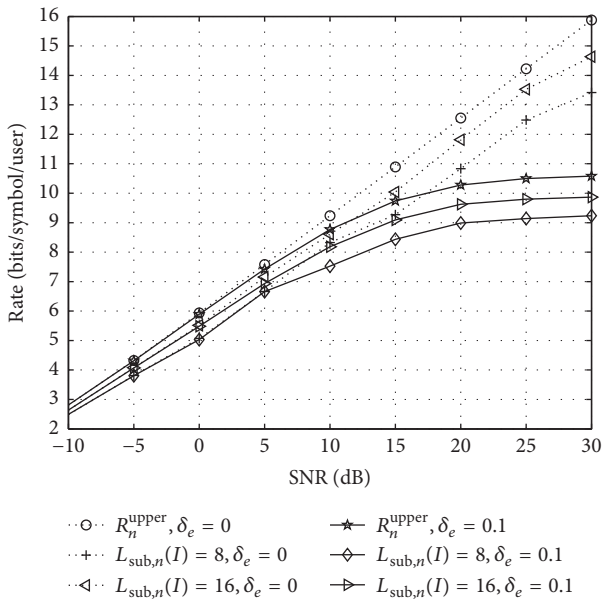



FIGURE 6: The effects of $L_{\text{sub},n}(I)$ on average rates achieved by spinal codes based massive MIMO.

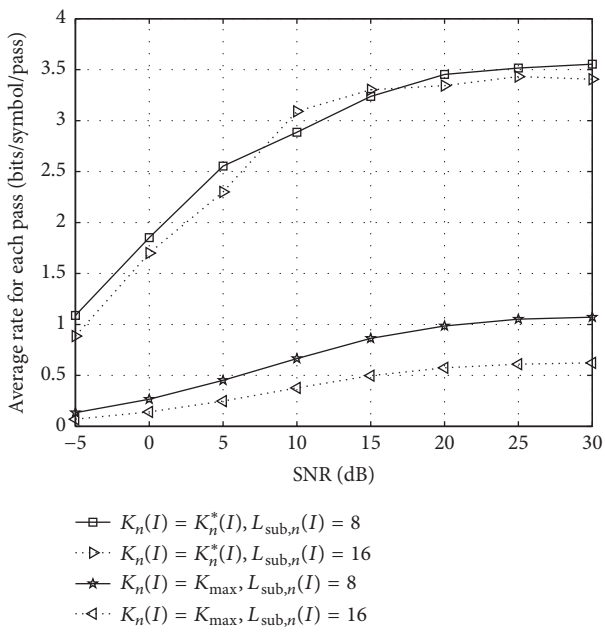



FIGURE 7: Average system retransmission efficiency, $K_{\max} > K_n^*(I)$, $\delta_e = 0.1$.

system have been presented in [17], where large-scale antennas are used to verify the significant gains in data rates. A rapid prototyping system architecture is proposed in [18], where the proposed system has high scalability in terms of the number of antennas. In order to verify the performance of spinal codes based MIMO with imperfect CSI, to equip large-scale antennas is unnecessary and also too expensive to achieve. Therefore, we first only consider a system with 2 transmit antennas and 2 single-antenna users as a simplified and comparable case of the massive MIMO. We use NI Universal Software Radio Peripheral (USRP, here we use USRP 2920 as the basic module) and LabVIEW communication toolkits to build up the testbed, which is used to verify the benefits presented in our theoretical analysis. Owing to the decouple design of software and hardware, the simplified MIMO can be easily deployed and expanded to a massive MIMO system with little extra operations.

*5.1. Simplified MIMO System Architecture.* The MIMO system architecture is shown in Figure 8. In our demo system, specific MIMO cables are used to synchronize the clock sources between 2 transmitters and 2 receivers, respectively [19]. Then, they are connected to one host computer through an Ethernet switch. By using the unique host computer, the CSI can be obtained perfectly from the uplink to perform precoding, being also suitable for adding some more CSI errors.

We consider a Time Division Duplexing (TDD) system, where two users separately send orthogonal pilots to base station in the first time slot, and then the received pilots at the base station are used to estimate the uplink channel matrix. Base station then calculates precoding matrix with estimated CSI (in general imperfect CSI as presented above in this paper) and precodes symbols after parallel spinal encoders as Figure 1 shows. Precoded symbols are then transmitted through the downlink channel to users in the second time slot. Each user tries to retrieve original messages from received symbols and then sends ACK to based station if messages were decoded successfully. Figure 9 illustrates the user interference in the unique host computer. Three parts are mainly included, including parameters configuration (located on the left), retrieved-message display (on the middle), and numerical results display (on the right).

The decouple design of software and hardware can easily customize USRP as transmitter or receiver, which enables USRP to work correctly. USRP supports the input of synchronized signal from external clock sources; more than 2 functioned-well USRPs can be synchronized to build up a system with large transmit antennas and receivers. Therefore, both the function designs and devices support the scalability from a $2 * 2$ MIMO system to a massive MIMO system.

*5.2. Radio Frequency Calibration.* In our implemented TDD system as Figure 8 shows, the uplink and downlink channels are assumed to be reciprocal. However, real channels are not reciprocal due to the differences in transmitter and receiver hardware. An internal calibration method [20] is considered in our work, where we first find a reference radio; then if we know the calibration coefficient between any two radios and a reference radio, we can derive the direct calibration coefficient between them. With these calibration coefficients, the real channels will be derived.

*5.3. Demo System's Signal Frame Structure Design.* Figure 10 gives a simple frame structure design based on TDD mode for the $2 \times 2$ demo system. At the uplink scheduling slot, we use preamble to obtain the start point of data and synchronize the system frequency; the orthogonal pilots are transmitted through uplink sound channel (UL-SCH) to obtain the CSI.

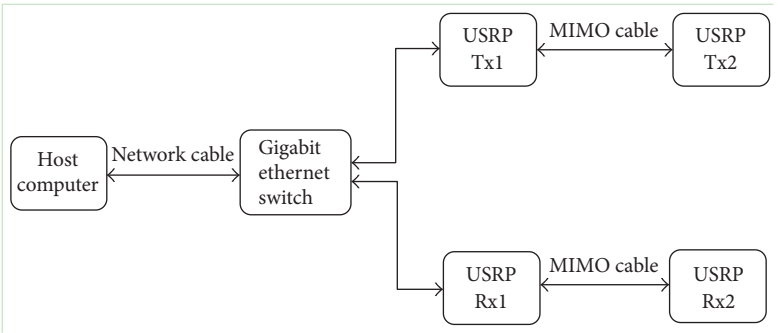



Figure 8: Simplified 2 × 2 MIMO demo system layout.

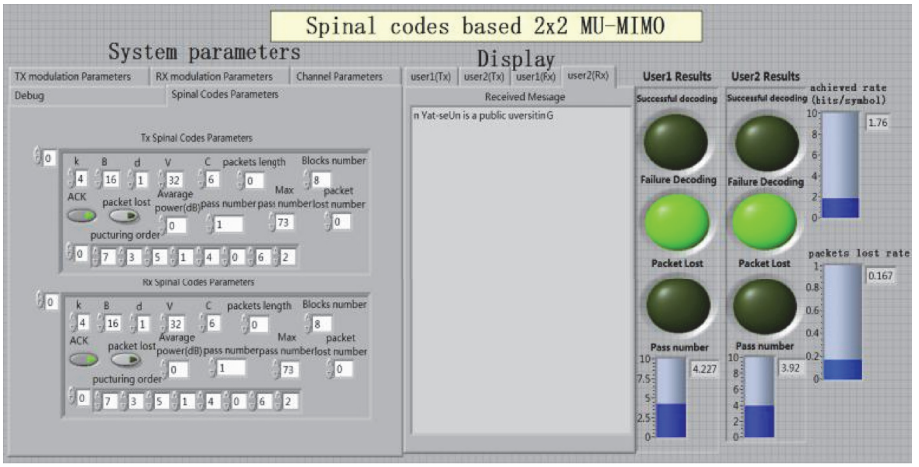



Figure 9: Simplified 2 × 2 MIMO demo system's user interface in the host computer.

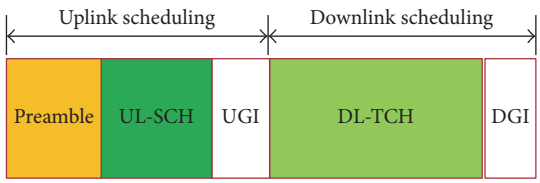



Figure 10: Data frame structure.

At downlink scheduling slot, users' messages are precoded and transmitted through the downlink traffic channel (DL-TCH). UGI and DGI are guard intervals in the uplink and downlink channel, respectively.

*5.4. Signal Detection Implementation Test Results.* Due to complicated field test condition limitations, we firstly test the above demo system's performance in fading channels by the channel simulator in LabVIEW software environments. The uplink and downlink signals are truly transmitted by radio. However, the distance limitation between transmit and receive antennas is so near, and we have to pass the signals through fading channels in LabVIEW software environments. The detailed parameters of the fading channels are presented in Table 1. Since the fading channel remains stationary, we use the real achieved rate to replace the MAR to obtain the dynamic block size.

Table 1: Detailed parameters of the system.

| Parameter | Value |
|---|---|
| Sample rate (MHz) | 1 |
| Number of multipaths | 3 |
| Path delay (us) | 0, 0.01, 0.02 |
| Path power (dB) | 0, −0.9, −1.7 |

Table 2: Retransmission times (pass number) of 2 × 2 MIMO.

| Block size | SNR (dB) | | |
|---|---|---|---|
| | 0 | 10 | 20 |
| Fixed $K_n(I) = 8$ | 22 | 13 | 10 |
| Dynamic $K_n(I) = K_n^*(I)$ | 4 | 3 | 3 |

The performance of retransmission times is compared by our proposed MPDBA and by fixed block-size scheme in Table 2. From this table, it can be seen that, by the MPDBA, the retransmission times can be reduced obviously in different SNR scenarios compared with the fixed block-size scheme. It also helps to verify the retransmission efficiency

improvements of the proposed spinal codes based MIMO scheme in fading channels.

## 6. Conclusions

The application of spinal codes in massive MIMO with imperfect CSI has great benefits. Before spinal codes are integrated into the system, an efficient transmission scheme for massive MIMO with spinal codes is proposed, where the maximum achievable rate (MAR) is used to determine the block size dynamically to reduce the retransmission delay spent on decoding the messages, and multilevel puncturing scheme is considered for different MARs to limit the gap between the MAR and the achieved rate. Some theoretical analyses are provided to prove that the multilevel puncturing and dynamic block-size allocation (MPDBA) based spinal codes can guarantee the system achieved rate as well as retransmission efficiency. Numerical simulation results illustrate that spinal codes with MPDBA can improve the achieved rate efficiently over that based on the fixed-rate codes as well as enhancing the system retransmission efficiency with limited achieved-rate performance degradation. Therefore, the proposed method can guarantee spinal codes based massive MIMO system to be reliable and practical. Finally, we implement a simplified MIMO testbed over NI's USRP platform to verify the proposal's performance improvements, which can help to understand our theoretical analysis and can be easily extended to massive MIMO scenarios.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work has been supported by Science, Technology and Innovation Commission of Shenzhen Municipality (JCYJ20160429170032960), National Natural Science Foundation of China (no. 61501527), State's Key Project of Research and Development Plan (no. 2016YFE0122900-3), the Fundamental Research Funds for the Central Universities, Guangdong Science and Technology Project (no. 2016B010126003), and 2016 Major Project of Collaborative Innovation in Guangzhou (no. 201604046008).

## References

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[2] J. Hoydis, S. Ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: how many antennas do we need?" *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, 2013.

[3] R. G. Gallager, *Low-density parity-check codes*, Springer International Publishing, 1960.

[4] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit errorcorrecting coding and decoding: Turbo-codes. 1," in *Proceedings of the IEEE International Conference on Communications*, pp. 1064–1070, 1993.

[5] Y. Sun, C. E. Koksal, K.-H. Kim, and N. B. Shroff, "Scheduling of Multicast and Unicast Services under Limited Feedback by using Rateless Codes," in *Proceedings of the IEEE INFOCOM*, pp. 1671–1679, Toronto, Canada, 2014.

[6] Y. Sun, C. E. Koksal, S.-J. Lee, and N. B. Shroff, "Network control without CSI using rateless codes for downlink cellular systems," in *Proceedings of the IEEE INFOCOM*, pp. 1016–1024, 2013.

[7] J. Perry, P. A. Iannucci, K. Fleming, H. Balakrishnan, and D. Shah, "Spinal codes," *ACM SIGCOMM Computer Communication Review*, vol. 42, no. 42, pp. 49–60, 2012.

[8] A. Gudipati and S. Katti, "Automatic rate adaptation and collision handling," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 4, pp. 158–169, 2011.

[9] O. Etesami and A. Shokrollahi, "Raptor codes on binary memoryless symmetric channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2033–2051, 2006.

[10] H. Balakrishnan, P. Iannucci, J. Perry, and D. Shah, "Derandomizing shannon: The design and analysis of a capacity-achieving rateless code," arXiv preprint arXiv:1206.0418, 2012.

[11] M. H. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[12] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part ii: perturbation," *IEEE Transactions on Communications*, vol. 53, no. 3, pp. 537–544, 2005.

[13] C. Windpassinger, R. F. H. Fischer, and J. B. Huber, "Lattice-reduction-aided broadcast precoding," *IEEE Transactions on Communications*, vol. 52, no. 12, pp. 2057–2060, 2004.

[14] C. Wang, E. K. S. Au, R. D. Murch, W. H. Mow, R. S. Cheng, and V. Lau, "On the performance of the MIMO zero-forcing receiver in the presence of channel estimation error," *IEEE Transactions on Wireless Communications*, vol. 6, no. 3, pp. 805–810, 2007.

[15] F. Rusek, D. Persson, B. K. Lau et al., "Scaling up MIMO: opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[16] P. Iannucci, J. Perry, H. Balakrishnan, and D. Shah, "No symbol left behind: A link-layer protocol for rateless codes," in *Proceedings of the International Conference on Mobile Computing and Networking*, pp. 17–27, 2012.

[17] X. Yang, W. J. Lu, N. Wang, K. Nieman, S. Jin, and H. Zhu, *Design and implementation of a tdd-based 128-antenna massive mimo prototyping system*, 2016.

[18] X. Yang, Z. Huang, B. Han et al., "RaPro: A Novel 5G Rapid Prototyping System Architecture," *IEEE Wireless Communications Letters*, vol. 6, no. 3, pp. 362–365, 2017.

[19] T. Zhao, *Design and implementation of MIMO system based on USRP*, Nanjing University of Posts and Telecommunications, June 2016.

[20] S. Clayton, H. Yu, A. Narendra et al., "Argos: practical many-antenna base stations," in *proceedings of the International Conference on Mobile Computing and Networking*, vol. 11, pp. 53–64, 2012.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 7658093, 13 pages
https://doi.org/10.1155/2018/7658093

*Research Article*

# Design and Analysis of Adaptive Message Coding on LDPC Decoder with Faulty Storage

**Guangjun Ge[1,2] and Liuguo Yin[2,3]**

[1]*School of Aerospace, Tsinghua University, Beijing, China*
[2]*EDA Laboratory, Research Institute of Tsinghua University in Shenzhen, Shenzhen, China*
[3]*School of Information Science and Technology, Tsinghua University, Beijing, China*

Correspondence should be addressed to Liuguo Yin; yinlg@tsinghua.edu.cn

Received 23 November 2017; Revised 9 February 2018; Accepted 19 February 2018; Published 22 March 2018

Academic Editor: Qin Huang



Unreliable message storage severely degrades the performance of LDPC decoders. This paper discusses the impacts of message errors on LDPC decoders and schemes improving the robustness. Firstly, we develop a discrete density evolution analysis for faulty LDPC decoders, which indicates that protecting the sign bits of messages is effective enough for finite-precision LDPC decoders. Secondly, we analyze the effects of quantization precision loss for static sign bit protection and propose an embedded dynamic coding scheme by adaptively employing the least significant bits (LSBs) to protect the sign bits. Thirdly, we give a construction of Hamming product code for the adaptive coding and present low complexity decoding algorithms. Theoretic analysis indicates that the proposed scheme outperforms traditional triple modular redundancy (TMR) scheme in decoding both threshold and residual errors, while Monte Carlo simulations show that the performance loss is less than 0.2 dB when the storage error probability varies from $\mathbf{10^{-3}}$ to $\mathbf{10^{-4}}$.

## 1. Introduction

Low-Density Parity-Check (LDPC) codes are widely used in space communications due to their capacity-approaching capabilities [1]. The outstanding performance of LDPC is based on the soft-decoding algorithms [2] which consume a large number of memories. However, the radiation environment will give rise to fault problems for memories when LDPC decoders are used in the spacecraft [3]. Such unreliable storage will severely degrade the performance of LDPC codes. Thus, it is important to consider the robustness of LDPC decoders utilizing unreliable memories.

There are studies on the effects of unreliable hardware on LDPC decoders. Varshney considered the thresholds and residual errors of LDPC codes with the faulty Gallager A decoding in the earlier stage [4]. Extended studies on faulty Gallager B decoders were then developed in [5–7]. Besides these bit flipping decoding algorithms, the belief propagation (BP) decoding of LDPC on noisy hardware was studied in [8, 9], where infinite-precision message with additive Gaussian noise was considered. Finite-precision message for the min-sum decoding of LDPC was studied in [10–12]. It showed that quantizing messages with more bits was not always beneficial for LDPC decoders with hardware errors.

In general, the existing works treated each finite-precision message as an integer, while this paper discusses the various impacts of different bits of the finite-precision message. We develop a discrete density evolution analysis for LDPC decoders with faulty messages. It indicates that the sign bits of the messages play the most important role in the decoding performance of LDPC codes, which means setting protection on sign bits is efficient enough. To protect the sign bit inside each quantized message, the traditional method is the static triple modular redundancy (TMR) scheme as applied in [13]. However, since two quantization bits are occupied for protecting the sign bit, the TMR scheme is not always beneficial for various storage error levels due to the loss of quantization precision. By analyzing the convergence process of LDPC decoding as well as referring to the results in [12, 14], it shows that when the magnitude of message is small, the precision bits, that is, the least significant bits (LSBs), are nonnegligible for decoding performance, while when the

message has a large magnitude value, the sign bit becomes even more critical for the residual errors.

Based on the aforementioned observations, we propose an adaptive embedded coding scheme for the unreliable messages to achieve a robust LDPC decoder. First, we put the messages into packages by taking advantage of the parallel message architecture of the quasi-cyclic (QC) LDPC decoders. The structure of message package permits more efficient block coding schemes for the sign bits other than simple TMR method. Then, two LSBs are adaptively employed for sign bits protection based on the magnitude level of message package. Moreover, we introduce a construction of Hamming product code for the adaptive coding, which has a multistage coding structure and outstanding error-correcting capability. We also discuss low complexity iterative decoding algorithms for the Hamming product code. Both theoretical analysis and Monte Carlo simulations demonstrate that the proposed adaptive message coding scheme outperforms the TMR scheme in decoding both thresholds and residual errors for various storage error levels.

The paper is organized as follows. In Section 2, the system models are introduced. Section 3 presents the discrete density evolution analysis on unreliable LDPC decoders. The adaptive message coding scheme and construction of Hamming product code are proposed in Section 4. We give the decoding algorithms for adaptive Hamming product codes in Section 5. Monte Carlo simulations are provided in Section 6. Section 7 concludes the entire paper.

## 2. System Models

*2.1. LDPC Decoder.* The hardware architecture of the QC-LDPC decoder is shown in Figure 1, which consists of interleaver ($\Pi_{H(\mathrm{LDPC})}$), variable node units (VNU), check node units (CNU), and data buffers (RAM). Since the matrix of QC-LDPC is divided into subblocks, the decoders are always implemented with the partially parallel architecture [15–17], which means the messages of each subblock are calculated by the same VNU or CNU node in the pipeline operations. The constraint of the LDPC code is executed by the interleaver, which is used to deliver the messages between the VNU and CNU based on the parity-check matrix of LDPC in various decoding algorithms [18, 19]. To execute the BP decoding of LDPC, the decoder firstly obtains the log-likelihood (LLR) $m_o$ from the channel. Then, VNU and CNU perform iterative computations, where the internal messages $m_{v2c}$ and $m_{c2v}$ are produced. Specifically, in VNU,

$$m_{v2c}(i) = m_o + \sum_{c' \in \mathcal{N}(v) \backslash c} m_{c'2v}(i-1), \tag{1}$$

while, in CNU,

$$m_{c2v}(i) = 2\tanh^{-1}\left[\prod_{v' \in \mathcal{N}(c) \backslash v} \tanh\left(\frac{m_{v'2c}(i)}{2}\right)\right], \tag{2}$$

where $\mathcal{N}(v)$ and $\mathcal{N}(c)$ are defined as the sets of nodes connected to node $v$ and node $c$, respectively. These messages are stored in memories during the decoding process. To implement LDPC decoders on integrated circuits, all of the messages will be quantized into bits. Existing studies [20] have shown that 4–6 bits' quantization on messages can provide ideal compromise between complexity and performance for LDPC decoders. Among the quantized bits, one is used for the sign, while the rest are used for the magnitude value.

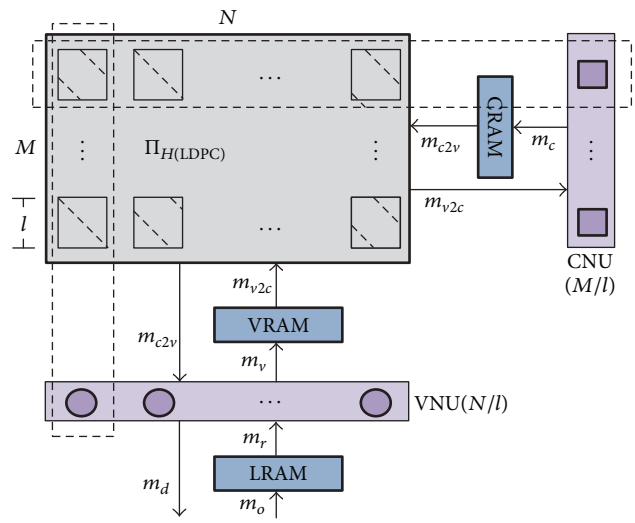



Figure 1: The partially parallel architecture of LDPC decoders.

*2.2. Error Model of Memory.* For the existing studies on LDPC decoders with faulty hardware, there are several widely accepted error models. As shown in Figure 2, where the model shown as Figure 2(a) is adopted in [4, 5, 7], the model shown as Figure 2(b) is adopted in [8, 9]. These models are both assumed to connect the error-free operation results with error channels, such as the binary symmetric channel (BSC) or the additive white Gaussian noise (AWGN) channel.

However, these two error models still have limitations for practical LDPC decoders. For example, the BSC error model is mostly utilized in bit flipping decoding algorithms, such as Gallager A decoding and Gallager B decoding, which make more sense in theoretical analysis. The AWGN error model is adopted in infinite-precision soft-decoding algorithms, where the messages are in continuous domain and assumed to be added with Gaussian noise by the faulty hardware.

In this paper, we consider the practical LDPC decoders, where finite-precision decoding algorithm is utilized. Following the studies in [11, 12], we assume a quantized BSC model for the storage errors, as shown in Figure 3. In the quantized BSC error model, the decoding messages are quantized into bits, each of which is assumed to pass a BSC error channel. The BSC errors for different bits are assumed to be independent, and the error ratios are assumed to be the same. The cross-over parameter $\alpha$ of the BSC channel is the flipping probability of the RAM cell, which is relevant to the radiation level and the service duration. As shown in Figure 1, there are 3 memories for the message storage: the LLR message storage, the V2C message storage, and the C2V message storage. In this paper, we assume the same bit flipping probability $\alpha_0$ for all message memories.

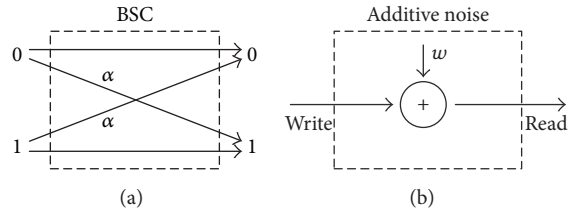



Figure 2: Existing faulty storage models.

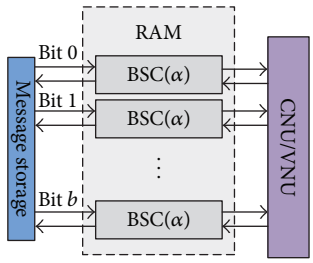



Figure 3: The quantized BSC model for the unreliable memories.

## 3. Analysis on LDPC Decoder with Unreliable Messages

*3.1. Discrete Density Evolution.* In this section, we define a discrete density evolution method for the analysis of finite-precision BP decoding of LDPC codes, which will give the performance of decoding thresholds and residual error ratios for LDPC decoders with different message protection schemes.

It has been proved by Varshney [4] that the symmetric conditions of density evolution are still suitable for the faulty LDPC decoders with symmetric hardware errors. Therefore, we can utilize the discrete density evolution by assuming all zero sequences are transmitted to analyze the finite-precision LDPC decoders with the BSC storage models. In this paper, only the regular $(d_v, d_c)$ LDPC codes are considered for the sake of simplicity.

In the density evolution analysis, we define $\mathbf{P}_m^k = \{p_1, p_2, \ldots, p_{2^Q-1}\}$ as the probability mass function (PMF) of the corresponding message $m$ at iteration $k$, where $Q$ is the number of quantization bits and $p_i$ is the probability of the $i$th quantization symbol. For example, $\mathbf{P}_o$ is the PMF vector of the message $m_o$ shown in Figure 1, which is the LLR from channel, while $\mathbf{P}_{v2c}^k$ is the PMF of message $m_{v2c}$ at the $k$th decoding iteration.

Since the codewords are assumed to be all zero sequences, to initialize the discrete density evolution, $\mathbf{P}_o$ is calculated as

$$\mathbf{P}_o = \left\{ p_i = \frac{1}{\sqrt{2\pi}\cdot 2/\sigma} \int_{q_{i-1}}^{q_i} \exp\left( -\frac{\left(t + 2/\sigma^2\right)^2}{8/\sigma^2} \right) dt \right\}, \quad (3)$$

where $q_i$ is the value of the $i$th quantization symbol, $q_0 = -\infty$, and $q_{2^Q-1} = +\infty$. Meanwhile, $\mathbf{P}_{c2v}^0$ is initialized as

$$\mathbf{P}_{c2v}^0 = \{0, \ldots, 0, p_{2^{(Q-1)}} = 1, 0, \ldots, 0\}. \quad (4)$$

After the initialization, the density evolution executes its iterations. Firstly, in the VNU nodes,

$$\mathbf{P}_v^k = \mathbf{P}_r^k \otimes \left(\mathbf{P}_{c2v}^{k-1}\right)^{\otimes(d_v-1)}. \quad (5)$$

It is worth noting that, after the convolution operations, we shall combine the extra elements of $\mathbf{P}_v^k$ so as to ensure a length of $2^Q - 1$.

Secondly, in the CNU nodes, the magnitude values of messages are mapped into log-domain by function $\lambda(x) = -\log(\tanh(x/2))$. The corresponding PMF of the magnitude values is mapped by $\Lambda(\mathbf{P}_{v2c}^k)$. Further define

$$\mathbf{\Gamma}_{v2c}^k = \left[\operatorname{sign}\left(\mathbf{P}_{v2c}^k\right), \Lambda\left(\mathbf{P}_{v2c}^k\right)\right]^{\otimes(d_c-1)}, \quad (6)$$

and thus the output PMF of CNU is updated by

$$\mathbf{P}_c^k = \operatorname{sign}\left(\mathbf{\Gamma}_{v2c}^k\right) \cdot \Lambda\left(\operatorname{abs}\left(\mathbf{\Gamma}_{v2c}^k\right)\right). \quad (7)$$

Similarly, the extra elements of $\mathbf{\Gamma}_{v2c}^k$ shall be combined after the convolution operations.

Finally, after the maximum iterations, the decoding decision is made in the VNU nodes, where the PMF is calculated as

$$\mathbf{P}_d^k = \mathbf{P}_r^k \otimes \left(\mathbf{P}_{c2v}^{k-1}\right)^{\otimes d_v}, \quad (8)$$

and the probability of residual error $p_e$ can be obtained by

$$p_e = \frac{1}{2} \cdot \mathbf{P}_d^k\left(2^{Q-1}\right) + \sum_{i=2^{(Q-1)}+1}^{2^Q-1} \mathbf{P}_d^k(i). \quad (9)$$

Above is the conventional discrete density evolution method for finite-precision LDPC decoders. However, this paper considers the issue of message storage errors, which means each message will suffer transformation of PMF outside the nodes. In the following, we will model the PMF transformation of unreliable message in density evolution.

Define $\mathbf{E} = \{e_1, e_2, \ldots, e_Q\}$ as the quantization bit error vector, where $e_i$ is the error probability of the $i$th quantization bit ($e_1$ for the sign bit, $e_Q$ for the LSB). For example, we can make $e_1 = e_2 = \cdots = e_Q = \alpha_0$ for the VRAM and CRAM

error models described in Section 2.2, where all quantized bits experience the same error probability. Further, define PMF transfer matrix $\mathbf{\Pi}(\mathbf{E})$, which is calculated as follows:

$$\mathbf{\Pi}(\mathbf{E}) = \left\{ p_{ij} = \prod_{k=1}^{Q} \beta_k(i,j) \right\}, \quad \left(i, j = 1, 2, \ldots, 2^Q - 1\right), \tag{10}$$

where $p_{ij}$ is the transfer probability from the $i$th quantization symbol $s_i$ to the $j$th one $s_j$. And $\beta_k(i,j) = 1 - e_k$ if $s_i$ and $s_j$ have the same bit in the $k$th quantization position; otherwise $\beta_k(i,j) = e_k$. Since $\beta_k(i,j) = \beta_k(j,i)$, we know that $\mathbf{\Pi}(\mathbf{E})$ is a symmetric matrix. As a result, the PMF transformations between RAM's input and output can be described as

$$\begin{aligned} \mathbf{P}_r &= \mathbf{P}_o \cdot \mathbf{\Pi}(\mathbf{E}(\alpha_0)), \\ \mathbf{P}_{v2c}^k &= \mathbf{P}_v^k \cdot \mathbf{\Pi}(\mathbf{E}(\alpha_0)), \\ \mathbf{P}_{c2v}^k &= \mathbf{P}_c^k \cdot \mathbf{\Pi}(\mathbf{E}(\alpha_0)). \end{aligned} \tag{11}$$

We could set various error vectors $\mathbf{E}$ for corresponding protection schemes for unreliable messages in discrete density evolution, which will give asymptotic performance of different protection schemes.

*3.2. Analysis on Various Bit Errors for Finite-Precision Messages.* Based on the discrete density evolution method defined in Section 3.1, a threshold analysis is provided to demonstrate the various effects of finite-precision message bits. It is shown that the sign bits *have the most influence* on the decoding thresholds of LDPC codes.

We execute the discrete density evolution on a $(4, 32)$ regular LDPC code with the 6 bits' quantized decoder in this paper. To analyze the various effects of quantized bits, we set $\mathbf{E} = \{0, \alpha_0, \alpha_0, \alpha_0, \alpha_0, \alpha_0\}$ for the one highest bit protected memory error model, which means the sign bit is assumed to be error-free. Similarly, several $k$ highest bits protected models can be defined, where $\mathbf{E} = \{0, \ldots, 0, \alpha_0, \ldots, \alpha_0\}$. The decoding thresholds are obtained by the discrete density evolution, as shown in Figure 4. We can see that if the sign bit is protected, the threshold will not be severely affected, while additional protection on the extra bits can provide little gain.

*3.3. Triple Modular Redundancy Scheme for Sign Bit Protection.* For LDPC decoders, the cost is overwhelming to protect every message bit. However, as mentioned before, it is not necessary since the sign bits are demonstrated to be the most important. Thus, following the idea of unequal error protection [21], we can simply set protection on the sign bits to promise a low complexity. In this section, we will firstly introduce traditional TMR protection scheme for the sign bits and then discuss its advantages and disadvantages.

As in [13], TMR has been applied in protecting the messages for LDPC decoder on unreliable hardware. However, TMR will charge two extra bits for protecting the sign bit while the messages are typically quantized into only 4 to 6 bits [20]. As a result, if we maintain the quantity of message quantization bits under the constraint of complexity, introducing TMR will bring a loss of quantization precision, which is not always beneficial for various storage error ratios. In the following, using the discrete density evolution method described in Section 3.1, we will analyze the performance of TMR scheme for the sign bit protection. We set $Q = 6$ for the quantity of quantization bits, which is adopted for most practical LDPC decoders. Moreover, to model the storage error of the TMR-protected messages, the error vector is set to be $\mathbf{E}_t = \{3\alpha_0^2 - 2\alpha_0^3, \alpha_0, \alpha_0, \alpha_0, \alpha_0\}$, which is quantized with 4 bits actually. While the unprotected LDPC decoder is set to be $\mathbf{E}_u = \{\alpha_0, \alpha_0, \alpha_0, \alpha_0, \alpha_0, \alpha_0\}$, the results of discrete density evolution under different storage error ratio $\alpha_0$ are shown in Figure 5.

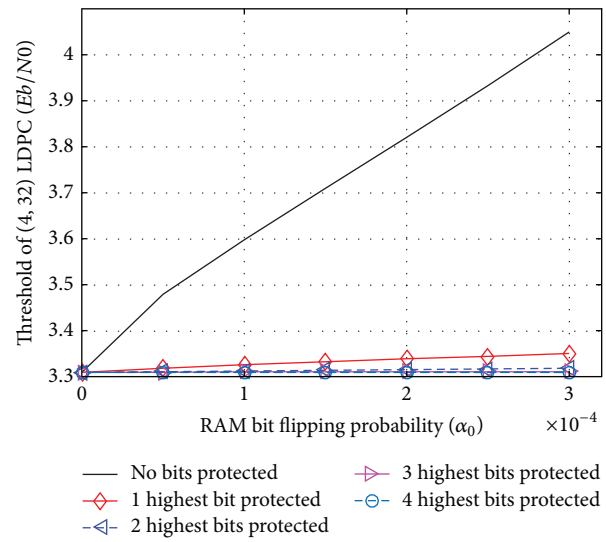



Figure 4: The thresholds of protecting various message bits.

From the analysis, it can be observed that when the storage error ratio is high (e.g., $\alpha_0 = 10^{-3}$), the LDPC decoder without protection cannot work anymore, while the TMR-protected one can work with a dramatic degradation of decoding threshold. However, when the storage error ratio is low enough, for $\alpha_0 = 10^{-4}$ and $\alpha_0 = 10^{-5}$, due to the loss of quantization precision, the TMR-protected LDPC decoders will show disadvantages in decoding threshold compared with the unprotected ones. However, the TMR protection scheme still has its advantages: we can observe that TMR-protected decoders have lower decoding residual errors for all levels of storage error ratios.

*3.4. Existing Adaptive Messages Coding Scheme.* We noticed that a similar adaptive coding scheme for approximate computing with faulty storage has been proposed in [22]. In [22], an adaptive message coding scheme on faulty min-sum LDPC decoders is mentioned. In detail, when the messages were written into RAMs, if the MSB was 1, the last two LSBs were neglected, while the corresponding memory addresses were used for a $(3, 1)$ repetition coding on the sign bit. Otherwise, the messages would be stored in the RAMs directly. When the LDPC decoder reads a message from the RAMs, the MSB

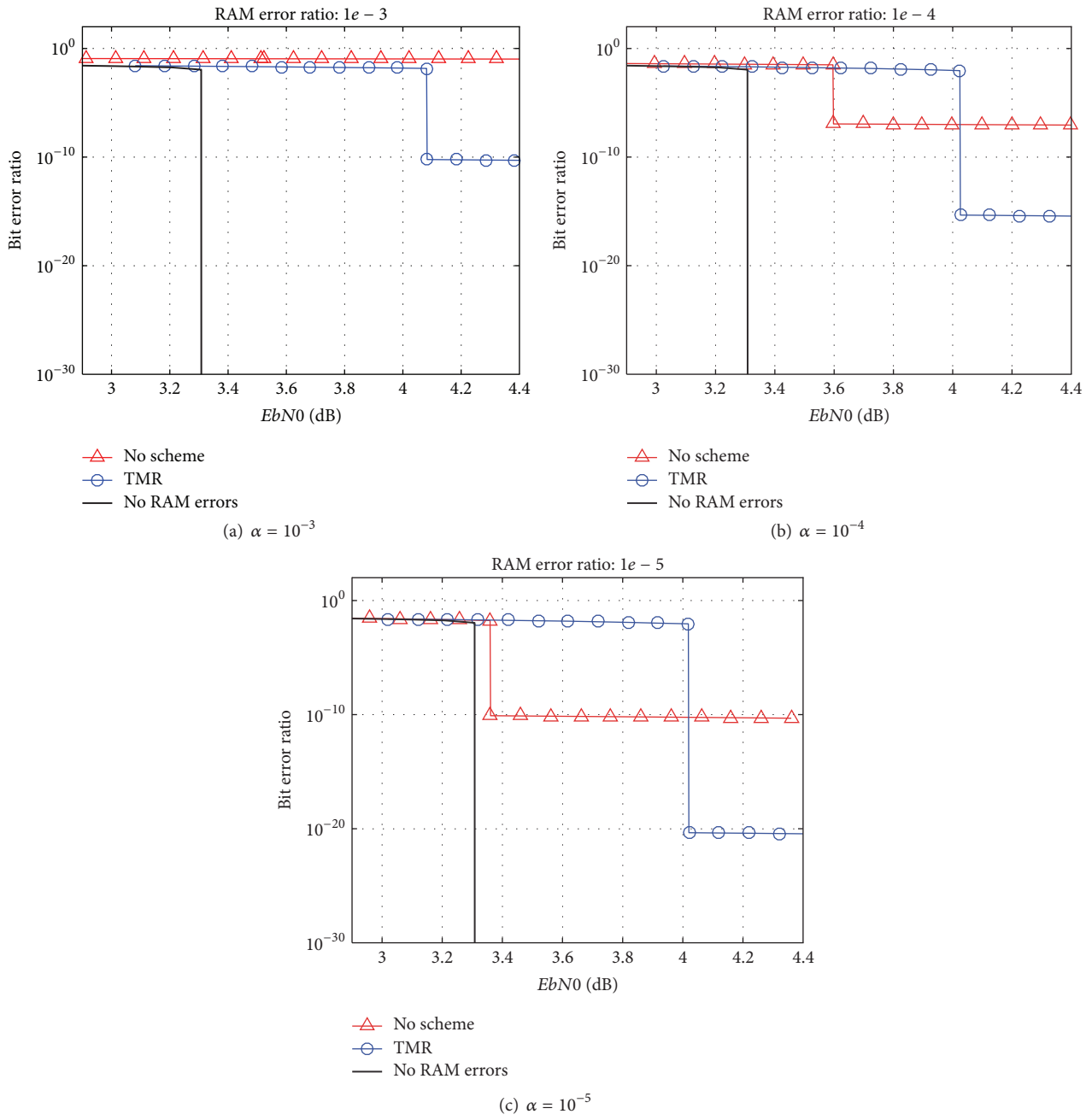



Figure 5: Performance analysis under different storage error ratios.

was checked. If the MSB was read as 1, a decoding on the (3, 1) code was executed to obtain the sign bit, while the last two LSBs were selected from {0, 1} randomly. Otherwise, the messages were assigned with the read values.

The aforementioned scheme makes full use of the LSBs in the messages. It has efficiently protected the unreliable messages without using any storage redundancy. However, there are some disadvantages for this protection scheme. Firstly, the adaptive coding is executed inside the single message, which is typically quantized with no more than 7 bits for the reason of complexity [20]. Consequently, there are not enough bits for the efficient coding schemes. For example, when the number of quantization bits is from 4 to 6, only simple repetition codes can be utilized. And this scheme is even inapplicable when the messages are quantized into less than 4 bits. Secondly, whether the adaptive coding is executed or not is totally based on the MSB, which is also subject to the storage errors. In such a case, the decoding of the adaptive code may be incorrectly executed, which further degrades the performance of the sign bit protection.

We demonstrate the exact error-correcting performance of the sign bits for this coding scheme as follows. In the first case where the MSB is 1, the encoding will be executed. If the MSB is read correctly, the (3, 1) code will be properly decoded

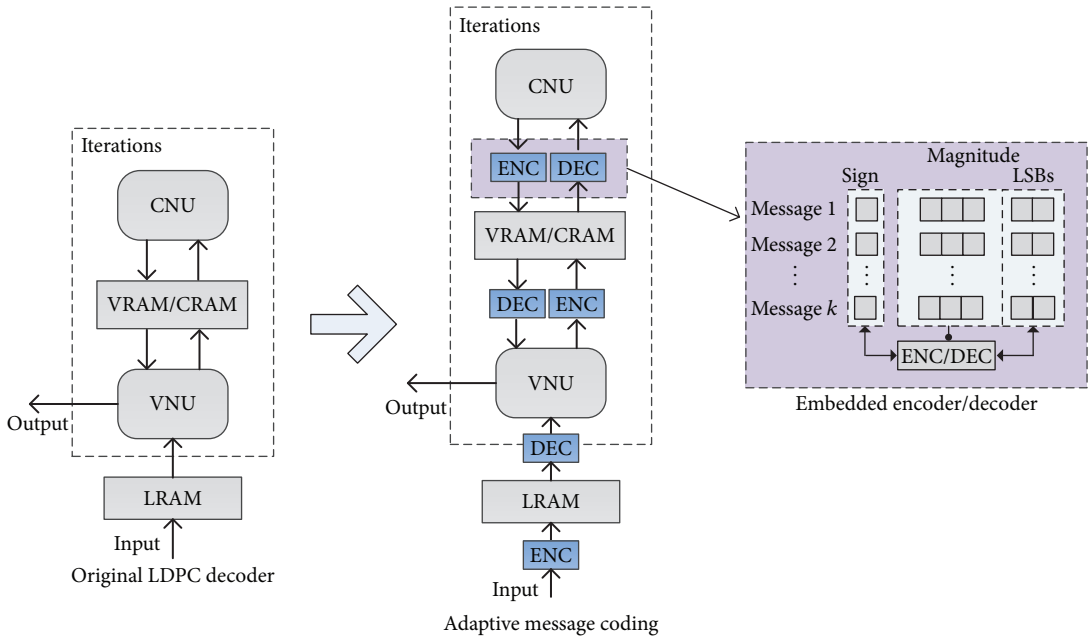



Figure 6: Structure of adaptive package coding.

with an output error ratio of $3\alpha^2 - 2\alpha^3$. If the MSB is read in error, the decoding will be neglected, which results in an error rate of $\alpha$ for the sign bit. That means that the expectation of the error rate for the sign bit is

$$\begin{aligned} p_e\,(\text{MSB} = 1) &= (1-\alpha) \times \left(3\alpha^2 - 2\alpha^3\right) + \alpha \times \alpha \\ &= 4\alpha^2 - 5\alpha^3 + 2\alpha^4. \end{aligned} \tag{12}$$

In the second case where the MSB is 0, similarly the error rate should be calculated with two cases, which is derived by

$$\begin{aligned} p_e\,(\text{MSB} = 0) &= (1-\alpha) \times \alpha + \alpha \times \left[\frac{1}{4}(1-\alpha) + \frac{3}{4}\alpha\right] \\ &= \frac{5}{4}\alpha - \frac{1}{2}\alpha^2. \end{aligned} \tag{13}$$

Unfortunately, since the storage error probability $\alpha$ is small enough, when the MSB is 1, this coding scheme cannot achieve the error-correcting capability of the (3, 1) repetition code, while when the MSB is 0, the error probability of the sign bit is even higher than the one without protection.

## 4. Adaptive Message Coding Scheme

In this section, we firstly present the architecture of the proposed adaptive coding scheme. Then, a specific construction of Hamming product code for the adaptive strategy is provided. Next, we analyze the performance of the proposed scheme theoretically.

*4.1. Protecting Sign Bits Utilizing LSBs Adaptively.* As analyzed in Section 3.4, protecting the sign bits of unreliable messages by occupying extra bits is not always the best scheme. The degradation of decoding threshold is mainly caused by the loss of quantization precision. However, we notice that quantization precision affects decoding performance specifically when the magnitude of message is small; that is, when the message has a large magnitude, the LSBs are less important. On the other hand, with the convergence of LDPC decoding process, the sign bits of messages are shown to be have significant effects on the decoding residual errors when most messages have large magnitudes. Based on these observations, while referencing the idea of adaptation as in [23], we introduce an adaptive scheme for protecting the sign bits of unreliable messages. The basic idea is that when the message magnitude is small, the LSBs are used for maintaining quantization precision, while when the magnitude is large enough, the storage space for LSBs is used to protect sign bits to ensure a lower residual error.

What is more, existing studies execute protection on each single message, where only simple coding scheme (such as TMR) can be utilized. However, we notice that LDPC decoders are usually implemented with a partially parallel architecture, as described in Section 2.1. In other words, a group of messages are produced simultaneously. It inspires us to put the sign bits into packages so that we can introduce efficient block coding schemes instead of the traditional TMR.

As shown in Figure 6, the structure of our proposed adaptive coding scheme is described as follows: first, put $k$ concurrently produced messages into a package. Then, define $T_1$ and $T_2$ as the adaptive thresholds, where $0 < T_1 < T_2 < L_{\max}$ ($L_{\max}$ is the maximum absolute value of quantization). Next, when the messages are written into RAMs, for each message package, calculate the average magnitude value $t$ of the $k$ messages. Based on the value of $t$, the adaptive coding is divided into 3 stages as below.

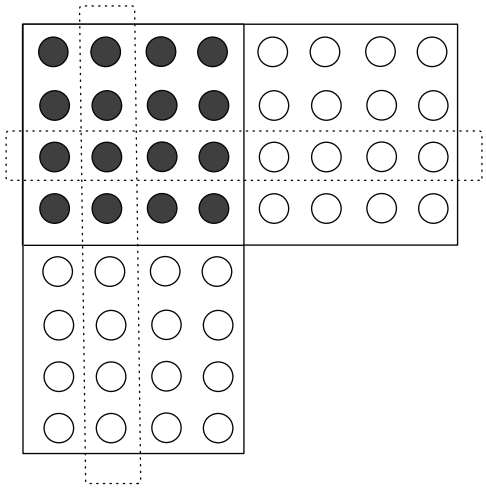


Figure 7: The (48, 16) Hamming product code.

(i) If $0 \le t < T_1$, all LSBs of the message package are reserved for quantizing messages.

(ii) If $T_1 \le t < T_2$, the storage space for the last $k$ LSBs are occupied for coding on the $k$ sign bits with a code rate of 1/2.

(iii) If $T_2 \le t \le L_{\max}$, the storage space for the last $2k$ LSBs are occupied for coding on the $k$ sign bits with a code rate of 1/3.

Reversibly, when the messages are read from RAMs, adaptive decoding is executed based on the value of $t$. If the storage of LSBs has been occupied for the sign bits, the LSBs of messages are randomly assigned.

*4.2. Construction of Adaptive Hamming Product Code.* In this section, we will give a specific code construction for the adaptive coding scheme.

To adaptively protect the sign bits of message packages, the ideal block code should have the features of multistage coding structure, as well as low coding complexity and appropriate block length. We introduce Hamming product codes as the adaptive package codes based on the following advantages. First, the product codes are constructed by several subcodes, whose coding process can be easily designed into multistage. Second, Hamming codes have the simplest decoders and encoders among all of the block codes, which only consist of several basic logic gates. Moreover, as the data is usually operated in bytes, where each byte contains 8 bits, in order to make the package codes suitable for the data operations, we choose the modified Hamming product code, which is $(n, k) = (48, 16)$. It is worth noting that some other short algebraic codes could be adopted to constitute the product code, such as Gray codes in [24], at the cost of complexity.

As shown in Figure 7, the dark points are the sign bits in one message package and the white points are the LSBs. The row and column subcodes are both (8, 4) Hamming codes. For such multistage (48, 16) Hamming product codes, package size $k = 16$, the first coding stage is that both row and column subcodes are inactive when $0 \le t < T_1$, while the second coding stage is executed by only activating the row subcodes when $T_1 \le t < T_2$. And the third coding stage is executed by activating the whole subcodes when $T_2 \le t \le L_{\max}$.

*4.3. Theoretical Analysis.* In this section, we will also utilize the discrete density evolution method to analyze our proposed adaptive package coding scheme. As mentioned before, we should deduce the error vector $\mathbf{E}$ for the proposed scheme.

As defined in Section 4.1, since the stage of adaptive coding is based on the average magnitude values of message packages, we should firstly calculate the PMF of the summation of magnitude values in one message package. First, the PMF of message magnitudes $|\mathbf{P}|$ is derived as

$$|\mathbf{P}| = \left\{\mathbf{P}\left(2^{(Q-1)}\right), \mathbf{P}\left(2^{(Q-1)}-1\right) + \mathbf{P}\left(2^{(Q-1)}+1\right), \ldots\right\}. \tag{14}$$

Then, the PMF of the summation of $k$ magnitudes can be obtained by

$$\mathbf{P}_{\text{sum}} = \left(|\mathbf{P}|\right)^{\otimes k}. \tag{15}$$

Based on the PMF of magnitude summation, we can obtain the probabilities of the 3 stages of adaptive coding, respectively, by

$$\begin{aligned} p_{t0} &= \sum_{\text{mag}(i)=0}^{k\cdot T_1} \mathbf{P}_{\text{sum}}(i), \\ p_{t1} &= \sum_{\text{mag}(i)=k\cdot T_1}^{k\cdot T_2} \mathbf{P}_{\text{sum}}(i), \\ p_{t2} &= \sum_{\text{mag}(i)=k\cdot T_2}^{k\cdot L_{\max}} \mathbf{P}_{\text{sum}}(i), \end{aligned} \tag{16}$$

where $\text{mag}(i)$ is the corresponding magnitude value of the $i$th element of $\mathbf{P}_{\text{sum}}$. Next, we need to calculate the error ratio for each stage of the adaptive Hamming product codes. As for $(n, k, d)$ block codes with a raw error ratio of $\alpha_0$, the error upper bound is derived by

$$e_{\text{blk}}(n, k, d, \alpha_0) = \sum_{i=\lfloor (d+1)/2 \rfloor}^{n} \binom{n}{i} (\alpha_0)^i (1-\alpha_0)^{n-i}. \tag{17}$$

Based on (17), the error vector for the first coding stage is $\mathbf{E}_{t0} = (\alpha_0, \alpha_0, \alpha_0, \alpha_0, \alpha_0, \alpha_0)$, while for the second stage it is $\mathbf{E}_{t1} = (e_{\text{blk}}(8, 4, 4, \alpha_0), \alpha_0, \alpha_0, \alpha_0, \alpha_0, 0.5)$, and for the third stage it is $\mathbf{E}_{t2} = (e_{\text{blk}}(48, 16, 7, \alpha_0), \alpha_0, \alpha_0, \alpha_0, 0.5, 0.5)$. As a result, the eventual error vector is obtained by

$$\mathbf{E} = p_{t0} \cdot \mathbf{E}_{t0} + p_{t1} \cdot \mathbf{E}_{t1} + p_{t2} \cdot \mathbf{E}_{t2}. \tag{18}$$

We set $T_1 = 0.4L_{\max}$ and $T_2 = 0.8L_{\max}$ and analyze the performance of our proposed scheme with $\alpha_0 = 10^{-3}$, $\alpha_0 = 10^{-4}$, and $\alpha_0 = 10^{-5}$. Firstly, we simulate with the parameter

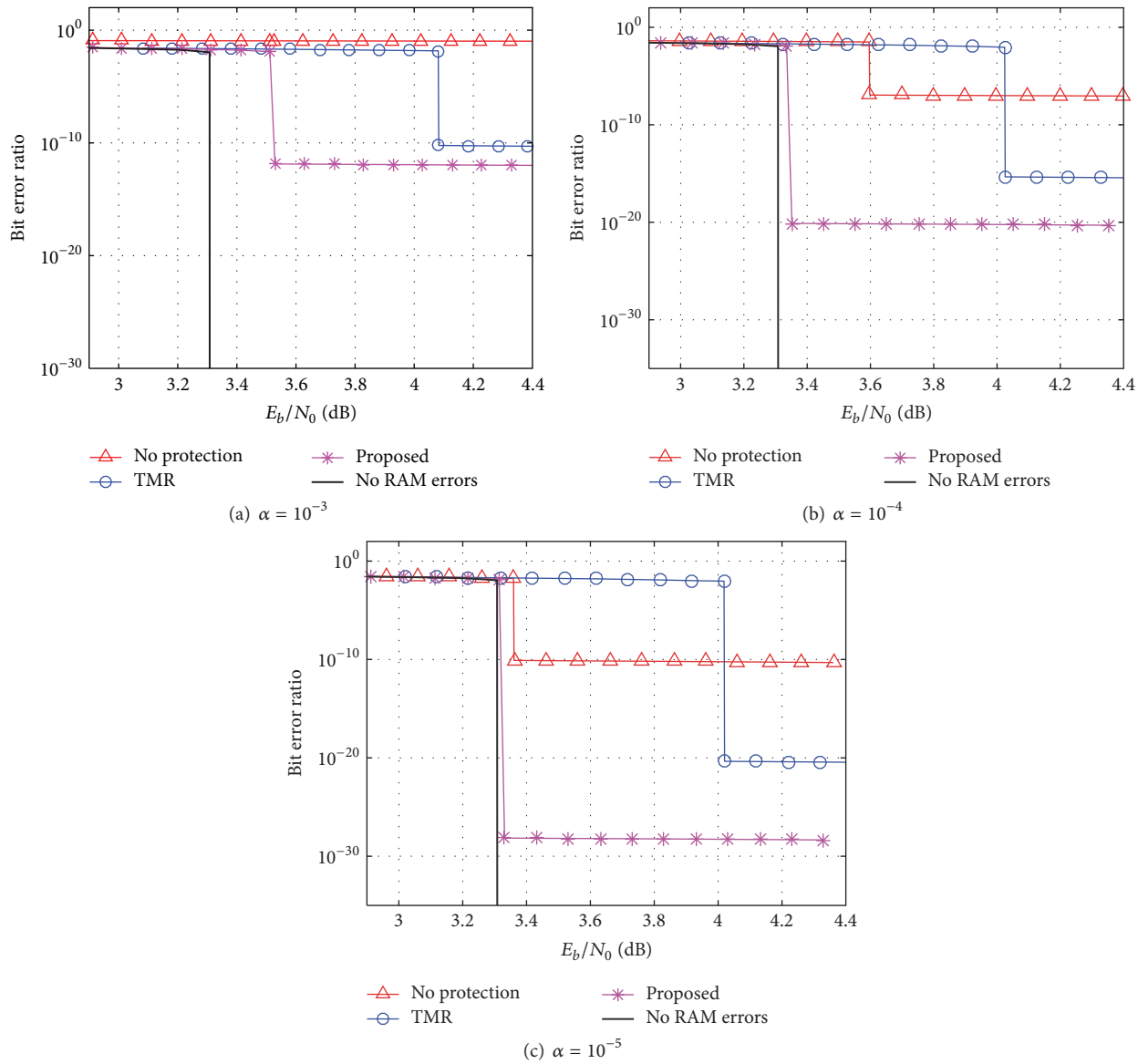



FIGURE 8: Performance analysis on adaptive package coding scheme with $(d_v, d_c) = (4, 32)$.

$(d_v, d_c) = (4, 32)$, whose results are shown in Figure 8. Then, as a comparison, we set $(d_v, d_c) = (4, 8)$ to verify the effects of different row weights (or the coding rates) on the performance, the results of which are shown in Figure 9. We can see that, with the proposed adaptive package coding scheme, the performances of both decoding threshold and residual error are significantly improved. What is more, our proposed scheme is effective in different coding rates.

## 5. Decoding of Hamming Product Codes

The Hamming product code we have introduced has an outstanding minimum distance characteristic. However, its error-correcting capability can only be achieved under the maximum likelihood (ML) decoding, which has high complexity and is not practical for LDPC message protection. In this section, we will discuss specific decoding algorithms for the Hamming product code, which achieves good performance with low complexity.

*5.1. Iterative Decoding of the Hamming Product Code.* For the subcode (8, 4) Hamming code, the Hamming distance is 4; that is, it can correct any one-bit error in one block. But when there are two error bits, the decoder can only declare a block error without locating the error bits. Based on the Hamming codes' characters of both error correcting and detecting, we define two states for the output bits of the Hamming product decoder: the fixed bits and the erasure bits. The decoding algorithm of the Hamming product code is described as follows:

(i) Iterative step: the row subcodes and the column subcodes execute their decoding algorithms iteratively. During the decoding, if the Hamming decoder cannot locate the error bits, keep the block unchanged;

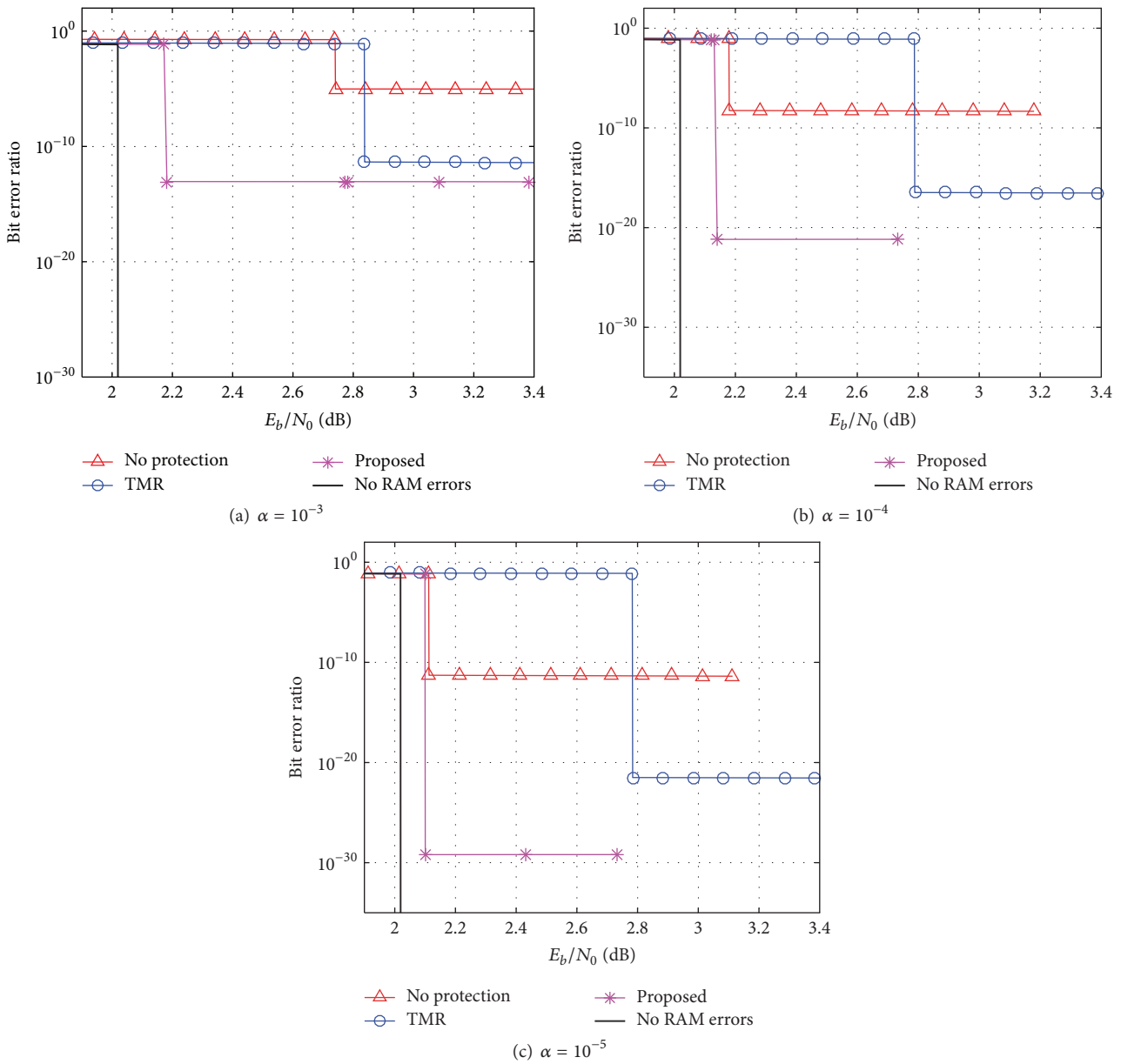



FIGURE 9: Performance analysis on adaptive package coding scheme with $(d_v,\ d_c) = (4, 8)$.

otherwise, update the block. After several iterations (we set it to 2 iterations here), stop the iterative decoding.

(ii) Decision step: firstly, the error detecting is executed by the Hamming decoders. Define $\mathscr{R} \subset \{1, 2, 3, 4\}$ as the set of indices where the row subcodes detect block errors. Similarly, define index set $\mathscr{C}$ for the column subcodes. Then, declare the bits located at $(i, j)$ in the $4 \times 4$ information bit matrix as erasure bits, where $i \in \mathscr{R}$ and $j \in \mathscr{C}$. The rest of the bits are declared as fixed bits.

We utilize a low-order approximation method to evaluate the performance of the Hamming product code with the proposed decoding algorithm. Since two states are defined for the output bits, we use two parameters to describe the decoding performance: the bit erasure ratio $r_e$ and the bit error ratio $r_p$. Both parameters can be approximately derived by the most likely error patterns. The probability of the $i$th-order error pattern can be derived as

$$P(i) = \alpha_0^i \left(1 - \alpha_0\right)^{48-i}, \tag{19}$$

where $\alpha_0$ is the bit flipping probability of the RAM. As listed in Table 1, the error patterns with orders lower than 4 are analyzed. Apparently, if there are less than two-bit errors in a block of the Hamming product code, no erasure or error bit exists. Therefore, only the error patterns with 3rd order and 4th order are adopted to deduce the approximate

Table 1: Low-order error pattern analysis.

| Error pattern order | Bit erasures total | Bit errors total |
|---|---|---|
| 1 | 0 | 0 |
| 2 | 0 | 0 |
| 3 | 256 | 16 |
| 4 | 13008 | 1680 |

performance. Using the 3rd order only, the error performance can be derived by

$$\begin{aligned} \tilde{r}_e(3) &= \frac{256}{16} \times P(3) = 16\alpha^3(1-\alpha)^{45}, \\ \tilde{r}_p(3) &= \frac{16}{16} \times P(3) = \alpha^3(1-\alpha)^{45}. \end{aligned} \tag{20}$$

While using both the 3rd order and 4th order, the performance is derived as follows:

$$\begin{aligned} \tilde{r}_e(3,4) &= \frac{256}{16} \times P(3) + \frac{13008}{16} \times P(4) \\ &= 16\alpha^3(1-\alpha)^{45} + 813\alpha^4(1-\alpha)^{44}, \\ \tilde{r}_p(3,4) &= \frac{16}{16} \times P(3) + \frac{1680}{16} \times P(4) \\ &= \alpha^3(1-\alpha)^{45} + 105\alpha^4(1-\alpha)^{44}. \end{aligned} \tag{21}$$

Meanwhile, Monte Carlo simulations for the iterative decoding of Hamming product code are provided. The curves of the performance are shown in Figure 10. We can see that $\tilde{r}_e(3,4)$ and $\tilde{r}_p(3,4)$ are very close to the results of Monte Carlo simulations. Moreover, the Hamming product code outperforms the traditional TMR scheme dramatically.

*5.2. Enhanced Decoding of the Hamming Product Code.* In fact, the performance of Hamming product code can be further improved at the expense of complexity for the iterative decoding. In this section, an enhanced decoding scheme is proposed to obtain better performance by introducing more decision logics.

As analyzed in Section 5.1, the 3rd-order error pattern has the most influence on the decoding performance of the Hamming product code. Specifically, the most typical error patterns for decoding erasure and error are depicted in Figure 11. Without loss of generality, we can assume that the row subcodes are decoded firstly in the iterative step. If there are three check bit errors (denoted by the dark points in Figure 11(a)) in one column subcode, one of the column subcode's information bits (denoted by the point marked with an asterisk) will be incorrectly decoded. In such a case, when it comes to the decision step, this incorrect information bit, together with the other three incorrect check bits, will constitute a valid codeword for the column subcode. Thus, it will result in a decoding error. Similarly, as shown in Figure 11(b), if there are three errors (denoted by the dark points) located, respectively, in the row subcode and the column subcode, they will simultaneously disable

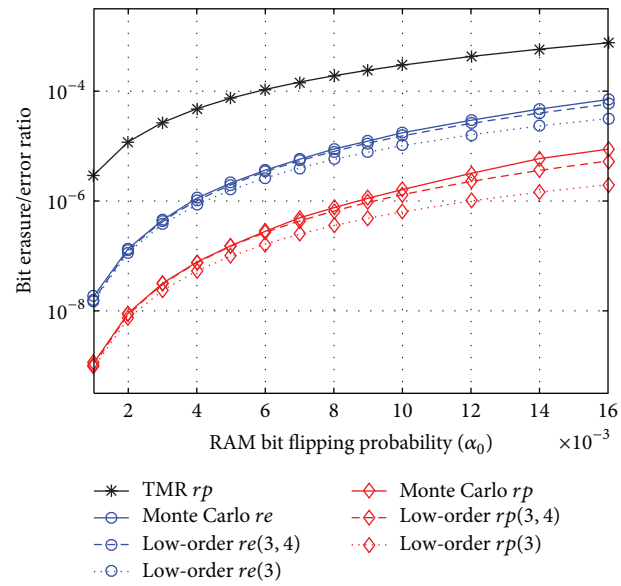



Figure 10: The low-order approximation of $r_e$ and $r_p$.

the decoding of the row and column subcodes. In such a case, the error in the intersection position will be declared as an erasure bit according to the decision logic of the aforementioned decoding algorithm.

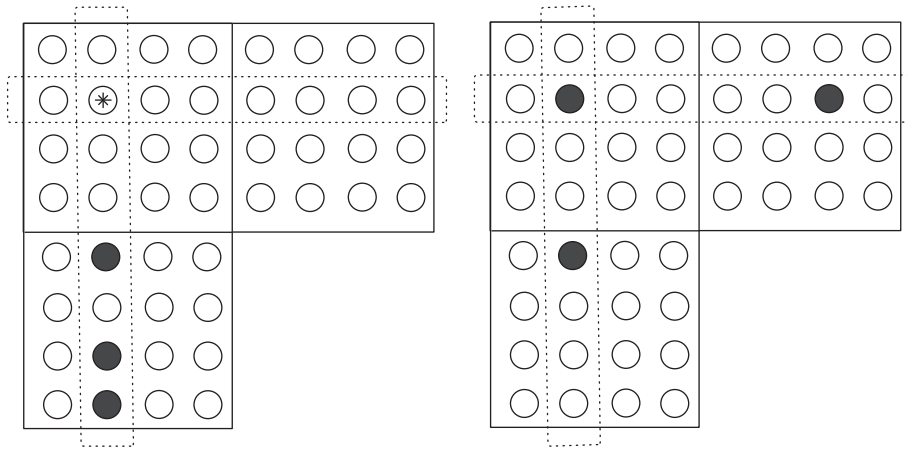
As a matter of fact, the decoding erasure bits and error bits caused by the 3rd-order error patterns all occurred in the similar ways mentioned above. Based on this analysis, an enhanced decoding scheme is proposed by adding the following two decision logics in the decision step:

(i) 3rd-order error bit decision: after the error detecting, if the index sets $\mathscr{R} = \{i\}$ and $\mathscr{C} = \{j\}$ are both single element sets, the bit located at $(i, j)$ is flipped and declared as a fixed bit.

(ii) 3rd-order erasure bit decision: if one bit is decoded into different values by the row subcode and the column subcode, it is declared as an erasure bit.

The performance of the enhanced decoding scheme is shown in Figure 12. There is an improvement for both $r_e$ and $r_p$. It should be noted that the additional decision logics are only used to cope with the 3rd-order error patterns. In fact, more decision logics for the higher-order error patterns can be introduced to obtain better performance.

*5.3. Decoding Complexity of the Hamming Product Code.* Another key issue is the complexity of decoding Hamming product code compared with traditional TMR scheme. As TMR only consumes a majority decision logic module to decode the duplicate check, it is generally believed that introducing advanced long block codes will definitely increase the hardware complexity. However, in this section, based on the Field Programmable Gate Array (FPGA) implementation, we will see that the hardware complexity of Hamming product code can be even lower than TMR scheme in some cases. Moreover, there will be a flexible tradeoff between hardware



(a) Error pattern (b) Erasure pattern

Figure 11: The 3rd-order error patterns.

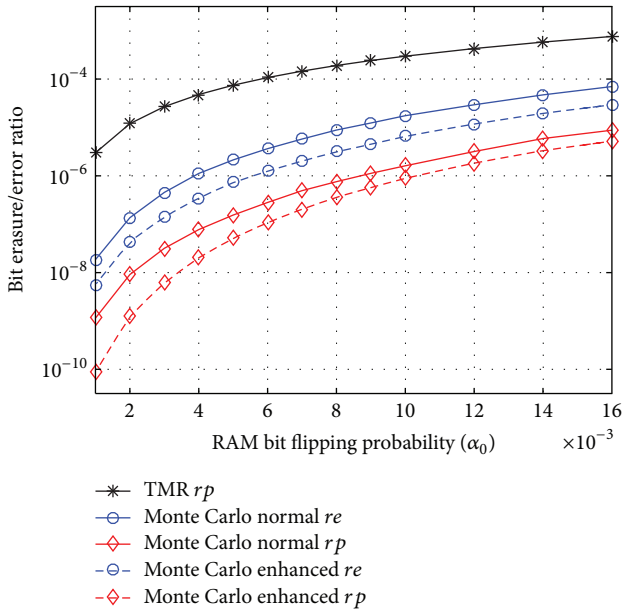



Figure 12: The performance of the enhanced decoding scheme.

consumption and decoding delay for the Hamming product code.

In applications, LDPC encoder and decoder are mostly implemented based on FPGA, which is reconfigurable and widely adopted in communication systems. A major difference between FPGA and the Application Specific Integrated Circuit (ASIC) is the structure of combinational logic circuit. For FPGA, the combinational logic is not composed of actual logic gates. Instead, it is based on a structure called Lookup Table (LUT), which is actually a small block of RAM. The input of combinational logic is connected to the RAM's address, and the logical output is presynthesized and stored into the RAM. Thus arbitrary logical operation can be implemented by looking up into the storage for each input logic combination. For conventional FPGA, 4-input and 6-input LUTs are mostly equipped. As a result, the TMR decision is actually processed by a 4-input LUT on FPGA. Next, we will compare the consumption of LUTs for the TMR and proposed schemes. In our proposed adaptive message coding scheme, each 16 messages are grouped into one package. Consequently, the corresponding consumption for TMR scheme is 16 4-input LUTs totally. Comparatively, the consumption of the proposed scheme is shown in Figure 13. We can see that, for the $(8, 4)$ Hamming encoder, only four 4-input LUTs are required, while, for the decoder, four 4-input LUTs are utilized to generate the correctors, and then each decoded information bit outputs through a 6-input LUT by logically processing the correctors and original value. To sum up, the total consumption is eight 4-input and four 6-input LUTs for $(8, 4)$ Hamming encoder and decoder. Based on these analyses, the hardware complexity of proposed Hamming product code is no more than TMR scheme, while roughly speaking it even consumes less resources on FPGA. Actually, in our iterative decoding algorithm for the Hamming product codes, the cost for improving error-correcting performance of unreliable message is decoding delay instead of hardware complexity. As the subcodes of Hamming produce code are decoded iteratively, the decoding of each message package will occupy a certain number of clocks. Thus if the LDPC decoder is poor in timing margin, the iterative decoding of Hamming product code will severely degrade the decoding throughput. Fortunately, the subcodes of Hamming product code can be decoded in parallel, which means we can compress the decoding clocks by parallel processing with multiple Hamming decoders. In this case, there can be a flexible tradeoff between hardware complexity and decoding delay. The specific consumption of space-time resource for various arrangements is shown in Table 2.

## 6. Simulations

In this section, Monte Carlo simulations are executed on the finite length codewords of LDPC. We utilize the $(8176, 7154)$ LDPC code defined by CCSDS in [25], which is publicly

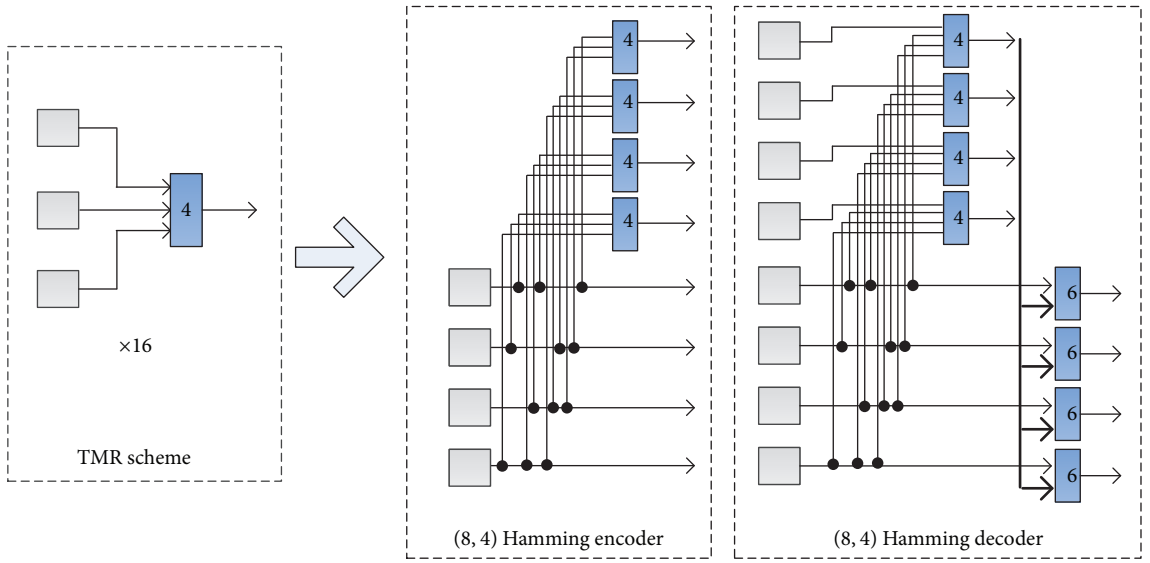



Figure 13: The implementation structure of various schemes based on LUT.

Table 2: Tradeoff between complexity and decoding delay.

| Serial number | Consumption of LUTs | Decoding clocks |
|---|---|---|
| 1 | 4 inputs × 4 + 6 inputs × 4 | 8 |
| 2 | 4 inputs × 8 + 6 inputs × 8 | 4 |
| 3 | 4 inputs × 16 + 6 inputs × 16 | 2 |

available and has outstanding performance. In the simulations, the messages are quantized into 6 bits, while the maximum number of iterations is set to 15. The communication channel is assumed to be the additive white Gaussian noise (AWGN) channel. To demonstrate the effectiveness of our proposed scheme under various storage error levels, the flipping probability of BSC model is set from $\alpha_0 = 10^{-3}$ to $\alpha_0 = 10^{-4}$. We compare the adaptive message coding scheme (labeled as "proposed") with both the traditional TMR scheme (labeled as "TMR") and the one without protection (labeled as "no sch"). The results are shown in Figure 14. We can see that when $\alpha_0 = 10^{-3}$, the proposed scheme has a gain of 0.2 dB compared to TMR scheme, while the unprotected one even cannot work. When $\alpha_0 = 10^{-4}$, The proposed scheme still outperforms the other schemes.

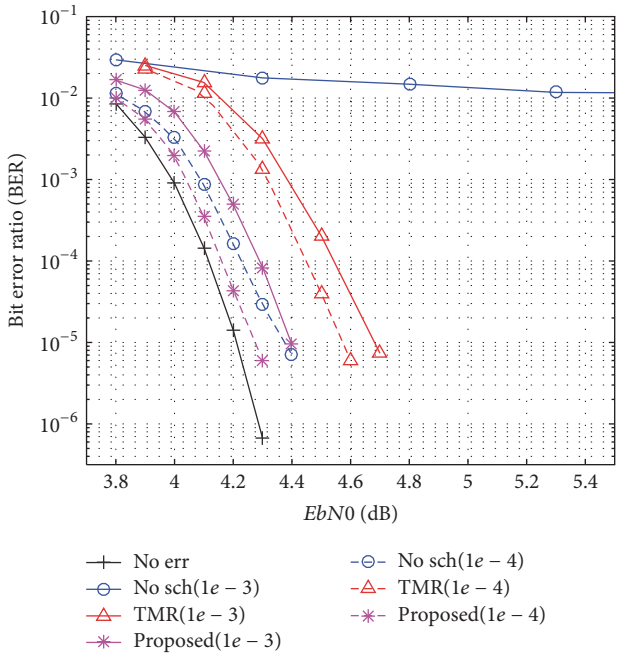



Figure 14: Simulations on the (8176, 7154) LDPC code with message storage errors.

## 7. Conclusion

This paper considered the challenge of implementing LDPC decoders on unreliable memories. We explored the effects of various message bits on finite-precision LDPC decoders and introduced an effective adaptive coding scheme based on the magnitude level of messages. We put the messages into packages and proposed a Hamming product code to adaptively correct the sign bits, as well as discussing two low complexity decoding algorithms. The discrete density evolution analysis showed that the proposed scheme outperforms traditional TMR scheme in decoding both threshold and residual errors under various storage error levels. Moreover, Monte Carlo simulations showed that the proposed scheme could at least obtain a gain of 0.3 dB to the static TMR scheme when the storage error probability was from $10^{-3}$ to $10^{-4}$.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC 91538203), the new strategic industries development projects of Shenzhen City (JCYJ20150403155812833), and the Beijing Innovation Center for Future Chips, Tsinghua University.

## References

[1] K. S. Andrews, D. Divsalar, S. Dolinar, J. Hamkins, C. R. Jones, and F. Pollara, "The development of turbo and LDPC codes for deep-space applications," *Proceedings of the IEEE*, vol. 95, no. 11, pp. 2142–2156, 2007.

[2] Q. Huang, Q. Xiao, L. Quan, Z. Wang, and S. Wang, "Trimming soft-input soft-output viterbi algorithms," *IEEE Transactions on Communications*, vol. 64, no. 7, pp. 2952–2960, 2016.

[3] D. Stone, A. Lindenmoyer, G. French et al., "NASA's approach to commercial cargo and crew transportation," *Acta Astronautica*, vol. 63, no. 1-4, pp. 192–197, 2008.

[4] L. R. Varshney, "Performance of LDPC codes under faulty iterative decoding," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4427–4444, 2011.

[5] F. Leduc-Primeau and W. J. Gross, "Faulty Gallager-B decoding with optimal message repetition," in *Proceedings of the 2012 50th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2012*, pp. 549–556, USA, October 2012.

[6] C. H. Huang, Y. Li, and L. Dolecek, "Gallager B LDPC decoder with transient and permanent errors," *IEEE Transactions on Communications*, vol. 62, no. 1, pp. 15–28, 2014.

[7] S. M. S. Tabatabaei Yazdi, H. Cho, and L. Dolecek, "Gallager B decoder on noisy hardware," *IEEE Transactions on Communications*, vol. 61, no. 5, pp. 1660–1673, 2013.

[8] C. H. Huang, Y. Li, and L. Dolecek, "Noisy belief propagation decoder," in *Proceedings of the 48th Asilomar Conference on Signals, Systems and Computers, ACSSC 2015*, pp. 2111–2115, USA, November 2014.

[9] C. H. Huang, Y. Li, and L. Dolecek, "Belief propagation algorithms on noisy hardware," *IEEE Transactions on Communications*, vol. 63, no. 1, pp. 11–24, 2015.

[10] E. Dupraz, D. Declercq, B. Vasic, and V. Savin, "Analysis and Design of Finite Alphabet Iterative Decoders Robust to Faulty Hardware," *IEEE Transactions on Communications*, vol. 63, no. 8, pp. 2797–2809, 2015.

[11] C. K. Ngassa, V. Savin, and D. Declercq, "Min-Sum-based decoders running on noisy hardware," in *Proceedings of the 2013 IEEE Global Communications Conference, GLOBECOM 2013*, pp. 1879–1884, USA, December 2013.

[12] A. Balatsoukas-Stimming and A. Burg, "Density evolution for min-sum decoding of LDPC codes under unreliable message storage," *IEEE Communications Letters*, vol. 18, no. 5, pp. 849–852, 2014.

[13] M. May, M. Alles, and N. Wehn, "A case study in reliability-aware design: A resilient LDPC code decoder," in *Proceedings of the Design, Automation and Test in Europe, DATE'08*, pp. 456–461, 2008.

[14] C. H. Huang, Y. Li, and L. Dolecek, "Adaptive error correction coding scheme for computations in the noisy min-sum decoder," in *Proceedings of the IEEE International Symposium on Information Theory, ISIT'15*, pp. 1906–1910, June 2015.

[15] Q. Li, X. Qu, L. Yin, and J. Lu, "Generalized Low-Density Parity-Check coding scheme with Partial-Band Jamming," *Tsinghua Science and Technology*, vol. 19, no. 2, pp. 203–210, 2014.

[16] Z. Chen, L. Yin, Y. Pei, and J. Lu, "CodeHop: physical layer error correction and encryption with LDPC-based code hopping," *Science China Information Sciences*, vol. 59, no. 10, Article ID 102309, 2016.

[17] P. Wang, L. Yin, and J. Lu, "An efficient helicopter-satellite communication scheme based on check-hybrid ldpc coding," *Tsinghua Science and Technology*, pp. 10–26599, 2018.

[18] Q. Huang, L. Song, and Z. Wang, "Set Message-Passing Decoding Algorithms for Regular Non-Binary LDPC Codes," *IEEE Transactions on Communications*, 2017.

[19] Q. Huang, M. Zhang, Z. Wang, and L. Wang, "Bit-reliability based low-complexity decoding algorithms for non-binary LDPC codes," *IEEE Transactions on Communications*, vol. 62, no. 12, pp. 4230–4240, 2014.

[20] J. Lee and J. Thorpe, "Memory-efficient decoding of LDPC codes," in *Proceedings of the Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.*, pp. 459–463, Adelaide, Australia, September 2005.

[21] J. Huang, Z. Fei, C. Cao, M. Xiao, and D. Jia, "On-Line Fountain Codes with Unequal Error Protection," *IEEE Communications Letters*, vol. 21, no. 6, pp. 1225–1228, 2017.

[22] C. H. Huang, Y. Li, and L. Dolecek, "ACOCO: Adaptive Coding for Approximate Computing on Faulty Memories," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 4615–4628, 2015.

[23] X. Ji, J. Xu, Y. L. Che, Z. Fei, and R. Zhang, "Adaptive Mode Switching for Cognitive Wireless Powered Communication Systems," *IEEE Wireless Communications Letters*, vol. 6, no. 3, pp. 386–389, 2017.

[24] L. Wang, Z. Wang, Q. Huang, and M. Zhang, "Balanced Gray Codes with Flexible Lengths," *IEEE Communications Letters*, vol. 20, no. 5, pp. 894–897, 2016.

[25] CCSDS, "Low density parity check codes for use in near-earth and deep space applications," CCSDS, 2011.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 8923478, 7 pages
https://doi.org/10.1155/2018/8923478

*Research Article*

# A CCM-Based OFDM System with Low PAPR for Sparse Source

**Qinbiao Yang,[1] Zulin Wang,[1,2] and Qin Huang [1]**

[1]*Electronic and Information Engineering, Beihang University, Beijing 100191, China*
[2]*Collaborative Innovation Center of Geospatial Technology, Wuhan 43079, China*

Correspondence should be addressed to Qin Huang; qinhuang@buaa.edu.cn

Received 24 November 2017; Revised 6 February 2018; Accepted 19 February 2018; Published 21 March 2018

Academic Editor: Michael McGuire



Orthogonal frequency division multiplexing (OFDM) usually suffers high peak-to-average power ratio (PAPR). As shown in this paper, PAPR becomes even severe for sparse source due to many identical nonzero frequency OFDM symbols. Thus, this paper introduces compressive coded modulation (CCM) in order to restrain PAPR by reducing identical nonzero frequency symbols for sparse source. As a result, the proposed CCM-based OFDM system, together with iterative clipping and filtering, can efficiently restrain the high PAPR for sparse source. Simulation results show that it outperforms about 4 dB over the traditional OFDM system when source sparsity is 0.1.

## 1. Introduction

Due to its high-speed data rate, orthogonal frequency division multiplexing (OFDM) [1, 2] has been widely applied to the 4G mobile communication system, Wi-Fi, and some military communication systems [3–7]. However, its signals may have high peak-to-average power ratio (PAPR). Because of the nonlinearity of power amplifier (PA), it will lead to signal distortion and restrict implementation of the OFDM system [8, 9].

To solve this problem, many PAPR reduction techniques have been proposed in recent few years. These techniques can be classified into three categories [10]: multiple signaling and probabilistic techniques [11–13], coding techniques [14], and signal distortion techniques [15–19]. In general, the first two categories, that is, the multiple signaling and probabilistic techniques and coding techniques, do not increase the bit error rate (BER), but they involve high computational complexity and data rate loss [10]. Accordingly, signal distortion techniques reduce the PAPR by distorting the transmitted OFDM signal before it passes through the PA. Although signal distortion techniques slightly increase BER, they have lower computational complexity and do not result in data rate loss. Thus, in order to achieve high data rate in OFDM systems, we adopt signal distortion techniques for PAPR reduction.

In signal distortion techniques, clipping and filtering is the most popular way to reduce PAPR because of its low complexity and moderate signal distortion. It clips the OFDM signal to a predefined threshold and uses a filter to restrain the out-of-band radiation. However, the filtering operation may cause peak regrowth. Hence, iterative clipping and filtering (ICF) [15] has been proposed to suppress the peak regrowth. Some variants of ICF have been developed to increase convergence rate, for example, simplified ICF (SICF) [16], optimized ICF (OICF) [19], and simplified optimized ICF (SOICF) [20]. Only after several iterations is the SOICF able to achieve sufficient PAPR reduction with less BER degradation and low complexity.

In a traditional OFDM system, it usually takes the M-array Phase Shift Keying (MPSK) or the M-array Quadrature Amplitude Modulation (MQAM) bit-to-symbol mapping operations. If source is sparse, these bit-to-symbol mapping operations will generate more identical nonzero frequency OFDM symbols which will result in high PAPR. So PAPR becomes even severe in the case of sparse source. As a result, ICF and its variants may fail to obtain sufficient PAPR reduction.

This paper focuses on the PAPR reduction of OFDM for sparse source. The key idea to reduce PAPR for sparse source is to convert source bits into frequency OFDM symbols by

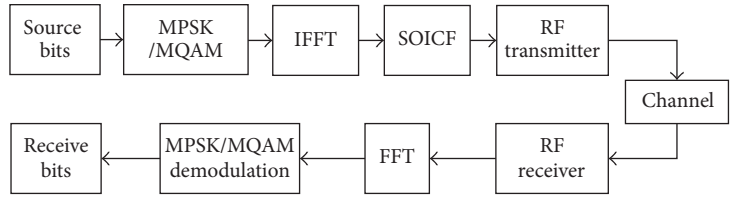



Figure 1: Diagram of the traditional OFDM system.

compressive coded modulation (CCM) [21] instead of traditional MPSK or MQAM. Usually, the CCM is used for seamless rate adaptation by adopting a special random projection (RP) to generate symbols from source bits. However, we verify that the RP symbols of the CCM for sparse source concentrate in zero, and there are much less nonzero symbols than those of MPSK and MQAM in the RP symbols. As a result, we propose using CCM together with SOICF to efficiently reduce the PAPR of OFDM for sparse source. Simulation results show that the proposed CCM-based OFDM system, together with the SOICF method, almost has the same performance in terms of PAPR reduction as a traditional OFDM system regardless of source sparsity. Furthermore, the proposed OFDM system outperforms about 4 dB over the traditional OFDM system when source sparsity is 0.1.

The rest of this paper is organized as follows. Section 2 briefly introduces necessary background. Section 3 presents the related theoretical analysis and the proposed novel OFDM system. Numerical and simulation results are provided in Section 4. The conclusion is given in Section 5.

## 2. Background

In this section, we first review the traditional OFDM system and its PAPR problem. Then, the CCM is briefly described.

*2.1. Traditional OFDM System.* As shown in Figure 1, a traditional OFDM system usually takes the MPSK or MQAM to convert the source bits into the frequency OFDM symbols. Then, IFFT process part generates the discrete-time OFDM signal. Furthermore, we adopt the SOICF method to restrain PAPR. Finally, the optimized signal is transmitted by the RF transmitter, which includes oversampling, upconversion, and PA. In the receiver, the source bits can be recovered by the FFT and the corresponding demodulation.

Let us define $\mathbf{b} = (b_1, b_2, \ldots, b_M)^{\mathrm{T}} \in \{0, 1\}$ as the source vector, where T indicates the transpose of the vector. Through the bit-to-symbol mapping operations with $\mathbf{b}$, the frequency OFDM symbols with $N$ subcarriers can be written as $\mathbf{X} = [X(0), X(1), \ldots, X(N-1)]$, and the discrete-time signal can be obtained by

$$\begin{aligned} x(n) &= \frac{1}{\sqrt{N}} \sum_{k=1}^{N} X(k) \cdot e^{j(2\pi nk/NL)} \\ &= \sqrt{N} \cdot \mathrm{IFFT}\left(X(k)\right)_{NL}, \quad n = 0, 1, \ldots, NL-1, \end{aligned} \tag{1}$$

where $L$ is the oversampling factor and the oversampling operation is implemented by zero padding in the end of $\mathbf{X}$. For signal $x(n)$, the PAPR is defined as the ratio of the maximum power to the average power and can be formulated as

$$\mathrm{PAPR} = \frac{\max\left[|x(n)|^2\right]}{E\left\{|x(n)|^2\right\}}, \tag{2}$$

where $E\{|x(n)|^2\}$ denotes the average power of $x(n)$. As mentioned in [22], if the elements of $\mathbf{X}$ are statistically independent and identically distributed random variables, the real and imaginary parts of $x(n)$ are Gaussian random variables with zero mean and same variance $\sigma^2$ when $N > 64$. Consequently, the amplitude of $x(n)$, that is, $|x(n)|$, is a Rayleigh random variable and, sometimes, the peak power of $x(n)$ is much larger than its average power. In other words, it results in high PAPR of the OFDM signal, which degrades the system's performance.

*2.2. CCM Scheme.* The CCM scheme is a compressive coded modulation for seamless rate adaptation. Its core is a special random projection code (RPC) and the corresponding decoding algorithm. Due to the embedding of source compressive into modulation, the CCM scheme brings significant throughput gain when source is sparse. Moreover, the decoding algorithm performs joint decoding based on received RP symbols, and the number of received RP symbols can be adjusted in fine granularity. Then, it is usually used for seamless rate adaption. The key process of the CCM scheme is as follows:

(i) Design the desired RPC.

(ii) Use the RPC and source bits to generate the RP symbols.

(iii) Map every two consecutive RP symbols into a wireless symbol with I and Q components.

(iv) Transmit the wireless symbols through carrier modulation and digital-to-analog converter (DAC).

(v) Recover the transmitted source bits through the RPC-Belief Propagation (BP) [21] decoding algorithm in the receiver.

Specifically, unlike traditional bit-to-symbol mappings, the RP can implement the bit-to-symbol mapping and channel protection at the same time. We present the schematic diagram of RP in Figure 2.

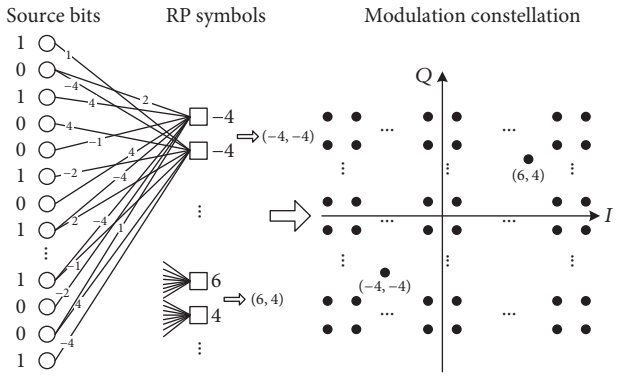



FIGURE 2: Illustration of random projection.

As illustrated in Figure 2, source bits can be converted to RP symbols by weighted sum operation. Then, every two RP symbols are mapped into one wireless symbol. Each constellation point is used to represent one wireless symbol. RPC bit-to-symbol mapping converts the source vector $\mathbf{b}$ into a series of RP symbols $s_1, s_2, \ldots, s_R$ by weighted sum operation. Define $w = \{w_1, w_2, \ldots, w_l\} \in \mathbb{Z}$, $l = P$, as weight set; the $r$th RP symbol can be generated by

$$s_r = \sum_{l=1}^{P} w_l \cdot b_{m_{rl}}, \tag{3}$$

where $b_{m_{rl}}$ is the element of the source vector $\mathbf{b}$ and $m_{rl}$ is the index of the bit weighted by $w_l$ for generating symbol $s_r$. Based on (3), the RPC can be regarded as a low-density generator matrix $\mathbf{G}$. There are only $P$ nonzero entries in every row of matrix $\mathbf{G}$, and $w_l$ are located in the $P$ columns randomly. Denote $\mathbf{s} = (s_1, s_2, \ldots, s_R)^{\mathrm{T}}$ as a symbol vector, and then the RP symbols can be formulated as

$$\mathbf{s} = \mathbf{G} \cdot \mathbf{b}. \tag{4}$$

Generally, in order to achieve better transmission rate for CCM scheme, the key problem of weight set selection is that the entropy of the RP symbols generated by the weighted sum operation is always large. In addition, the low-density characteristic of $\mathbf{G}$ also should be guaranteed. The specific design principles of $w$ and $\mathbf{G}$ can be found in [21].

To generate the waveform for RF transmitter, the RP symbols have to be mapped to the amplitude of sinusoid signals. Let $V_{\min}$ and $V_{\max}$ be the minimum and maximum value in $\mathbf{s}$. Suppose that $A$ is the maximum amplitude of sinusoid signals. The mapping is thus a linear projection from $[V_{\min}, V_{\max}]$ to $[-A, A]$, and the actual amplitude after mapping is

$$x_r = -A + \frac{2A}{V_{\max} - V_{\min}} \left(s_r - V_{\min}\right). \tag{5}$$

Eventually, the wireless symbols are transmitted to wireless channel through carrier modulation and DAC, while each wireless symbol is composed of two consecutive $x_r$, that is, $x_r + j \cdot x_{r+1}$.

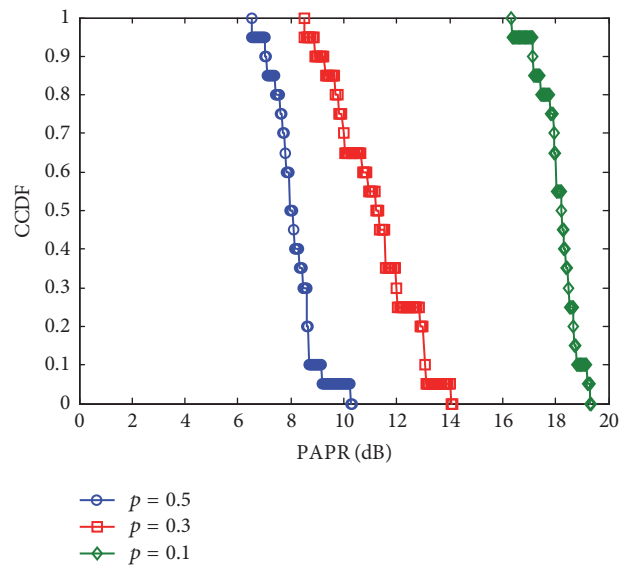



FIGURE 3: PAPRs for different $p$ with QPSK-based OFDM and $L = 8$.

In the receiver, through the demodulation and inverse projection from $[V_{\min}, V_{\max}]$ to $[-A, A]$, we can obtain the noisy RP symbols from wireless channel. Furthermore, using the RPC-BP decoding algorithm, we can recover the source bits from the noisy RP symbols.

## 3. Proposed CCM-Based OFDM System

From Section 2.1, we know that the OFDM signal is the summation of the products of entries in $\mathbf{X}$ and the multicarrier signals. According to (1), if some elements of $\mathbf{X}$ have the same sign and the carriers signals have the same phase, the signal obtained by summation will have a large peak amplitude; otherwise, the signal will have a peak amplitude around the average.

The source sparsity can be modelled by unequal probabilities on the appearance of 0's and 1's [23]. In this paper, we denote the probability $P(b = 1) = p$ as the source sparsity, that is, the probability on the appearance of 1. In the traditional OFDM system, if the source sparsity $p$ is small, that is, the source bits contain a large portion of 0's, the traditional bit-to-symbol mapping operations, for example, MPSK and MQAM, may generate a lot of identical nonzero symbols. This may cause higher PAPR than that of nonsparse source. For instance, Figure 3 shows the PAPR complementary cumulative distribution function (CCDF) [10] curves with different source sparsity. The CCDF of the PAPR denotes the probability that a PAPR exceeds a certain value. It is shown that the PAPR increases as the source becomes sparse.

The above analysis shows that the high PAPR for sparse source in traditional OFDM system is caused by the identical nonzero symbols generated by conventional bit-to-symbol mapping operations. Therefore, we would like to use new bit-to-symbol mapping that produces less identical nonzero

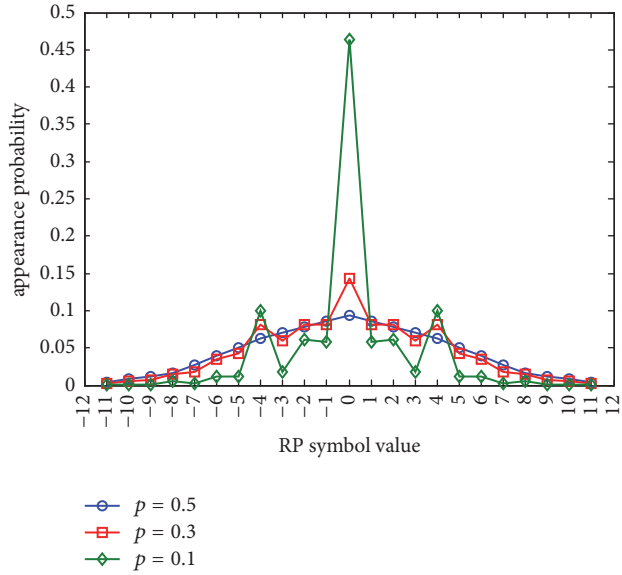



Figure 4: RP symbols distribution for different $p$.

symbols for sparse source. Later, we will verify that the CCM produces more zero symbols but less identical nonzero symbols for sparse source. This feature of the CCM promises a low PAPR for sparse source.

According to (3), we know that all the generated RP symbols are integer including zero and distribute from the minimal value to the maximal value of the sum of the weights. Moreover, the distribution of the RP symbol $s_r$ is determined by the weight set and the source sparsity $p$. Considering $w = \{w_1, w_2, \ldots, w_l\} \in \mathbb{Z}$, $l = P$, and $w_1 = -w_2$, $w_3 = -w_4, \ldots, w_{l-1} = -w_l$, the RP symbol can be written as

$$\begin{aligned} s_r &= w_1 \cdot b_{m_{r1}} + w_2 \cdot b_{m_{r2}} + \cdots + w_{l-1} \cdot b_{m_{r(l-1)}} + w_l \cdot b_{m_{rl}} \\ &= w_1 \left(b_{m_{r1}} - b_{m_{r2}}\right) + w_3 \left(b_{m_{r3}} - b_{m_{r4}}\right) + \cdots \\ &\quad + w_{l-1} \left(b_{m_{r(l-1)}} - b_{m_{rl}}\right). \end{aligned} \tag{6}$$

Since the probability distribution function (PDF) of $b_{m_{rl}}$ is

$$\begin{aligned} P\left(b_{m_{rl}} = 0\right) &= 1 - p \\ P\left(b_{m_{rl}} = 1\right) &= p, \end{aligned} \tag{7}$$

let $b_{m_\Delta}$ denote $(b_{m_{l-1}} - b_{m_l})$; we can obtain the PDF of $b_{m_\Delta}$ as follows:

$$P\left(b_{m_\Delta}\right) = \begin{cases} p(1-p) & b_{m_\Delta} = -1 \\ p^2 + (1-p)^2 & b_{m_\Delta} = 0 \\ p(1-p) & b_{m_\Delta} = 1. \end{cases} \tag{8}$$

From (8), we can prove that $P(b_{m_\Delta} = 0)$ is always the maximal value among all values regardless of $p$, and the maximal value becomes large as $p$ becomes small. That is to say, when $P = 2$, the appearance probability of zero symbol, that is, $P(s_r = 0)$, dominates the PDF of $s_r$. Moreover, since the PDF of $s_r$ is a joint PDF of every $b_{m_\Delta}$, the PDF of $s_r$ would be similar to the case of $P = 2$ when $P$ becomes large. For instance, if $w = \{\pm 1, \pm 2, \pm 4, \pm 4\}$, the RP symbols take value from $[-11, +11]$, and they are generated by the equation as follows:

$$\begin{aligned} s_r &= (+1) \cdot b_{m_{r1}} + (-1) \cdot b_{m_{r2}} + (+2) \cdot b_{m_{r3}} + (-2) \cdot b_{m_{r4}} \\ &\quad + (+4) \cdot b_{m_{r5}} + (-4) \cdot b_{m_{r6}} + (+4) \cdot b_{m_{r7}} + (-4) \\ &\quad \cdot b_{m_{r8}}. \end{aligned} \tag{9}$$

In Figure 4, we illustrate the RP symbols distribution for different source sparsity $p$.

In Figure 4, we can see that the zero symbol always dominates the distribution regardless of source sparsity $p$ in the case of $w = \{\pm 1, \pm 2, \pm 4, \pm 4\}$. In the other cases of weight set, the same result can be obtained as Figure 4. This result also verifies the above analysis.

Therefore, the appearance probability of zero symbol always dominates the PDF of $s_r$ regardless of $p$, and if $p$ becomes small, $P(s_r = 0)$ will become large. That is to say, there are many zero symbols in RP symbols. Moreover, the sparser the source is, the more zero symbols there are.

Based on the above analysis, we use CCM together with SOICF to efficiently reduce the PAPR of OFDM for sparse source. Figure 5 illustrates the diagram of the CCM-based OFDM system.

As shown in Figure 5, we adopt the CCM scheme in the OFDM system. Then, the elements of $\mathbf{X}$ are the RP symbols. Because the values of the RP symbols are dominated by the zero value, the large peak amplitude of the summation

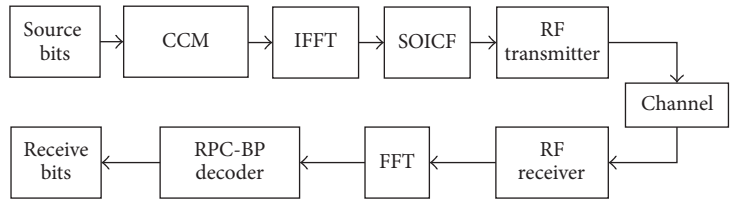



Figure 5: Diagram of the CCM-based OFDM system.

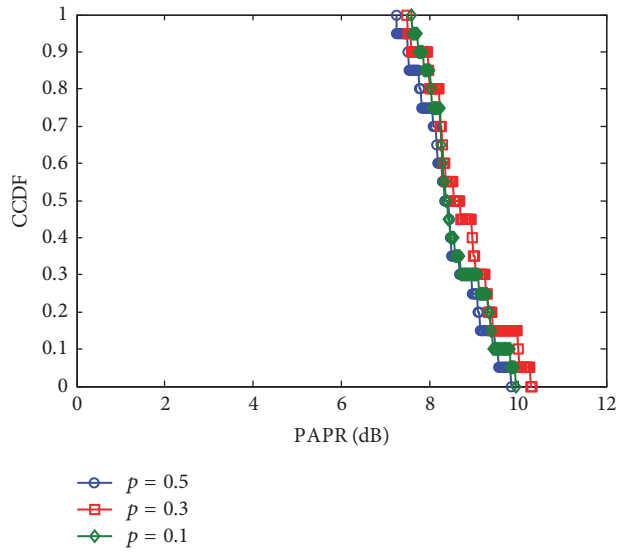



Figure 6: PAPRs of the CCM-based OFDM system for different $p$.

signal caused by many identical elements of $\mathbf{X}$ is efficiently restrained in the CCM-based OFDM system. As source sparsity becomes small, it will increase the number of the zero RP symbols. Then, in the case of sparse source, it does not cause high PAPR in the CCM-based OFDM system. Figure 6 depicts the PAPR in the CCM-based OFDM system for different $p$ with $w = \{\pm 1, \pm 2, \pm 4, \pm 4\}$, $M = 480$, $N = 128$, and $L = 8$.

From Figure 6, it can be seen that the PAPR of the CCM-based OFDM system is not sensitive to the source sparsity. As source becomes sparse, it will not bring additional PAPR increase.

## 4. Performance Evaluation

In the traditional OFDM system and the proposed CCM-based OFDM system, we adopt the SOICF method to evaluate the PAPR reduction performance for different source sparsity. Furthermore, the BER performance is evaluated for different source sparsity over the Additive White Gaussian Noise (AWGN) channel.

For a traditional OFDM system, we consider that it takes QPSK bit-to-symbol mapping operation and 128 subcarriers [20]. For the proposed CCM-based OFDM system, we consider that the size of its generator matrix $\mathbf{G}$ is 1920 rows and 480 columns [21]. Since the PAPRs with different weight sets are similar, we take the weight set $w = \{\pm 1, \pm 2, \pm 4, \pm 4\}$ with high achievable transmission rate to evaluate the PAPR reduction and BER performance. For source sparsity, we consider the source sparsity $p$ to be 0.5, 0.3, and 0.1 [21]. For the SOICF method, we define $A_{\text{th}}$ as the predefined threshold and $\gamma = A_{\text{th}}/\sqrt{P_{\text{av}}}$ as the clipped ratio and set $\gamma = 2.1$ dB, where $P_{\text{av}}$ means the average power of the signals before clipping. The oversampling factor $L = 8$ is able to provide a sufficiently accurate approximation of the PAPR [24], and thus we set $L = 8$ in our simulations.

The PAPR reduction performances are, respectively, shown in Figures 7(a), 7(b), and 7(c) for various source sparsity. Specifically, when the source is nonsparse, for example, $p = 0.5$, the PAPR reduction performances are almost identical in the traditional system and proposed system; when the source becomes sparse, for example, $p = 0.3$ and $p = 0.1$, due to the much higher PAPR for sparse source, the traditional OFDM system fails to restrain its PAPR. However, in the CCM-based OFDM system, because the PAPR of its signals has not been influenced by sparse source, it still can obtain much lower PAPR, and there is a performance gap around 9 dB between two systems when $p$ is 0.1.

Figure 8 shows the BER curves for different OFDM systems and source sparsity after 3 iterations of SOICF. As illustrated in Figure 8, the proposed OFDM system almost has the same BER performance as the traditional OFDM system with nonsparse source. Due to signal distortion caused by PAPR reduction, the BER performance of the traditional OFDM system begins to degrade as source becomes sparse, and the deterioration is about 2 dB at $p = 0.1$. However, in the proposed OFDM system, the BER performance is improved by the channel protection of CCM scheme, and the BER performance becomes better as source becomes sparser. Particularly, the proposed OFDM system outperforms about 4 dB over the traditional OFDM system, when source sparsity is 0.1. In addition, in the case of $w = \{\pm 1, \pm 2, \pm 4, \pm 4\}$, $M = 480$, and $p = 0.1$, the proposed system is able to achieve 8.5 bit/s/Hz average rate and about 12 bit/s/Hz peak rate for seamless rate adaptation [21].

## 5. Conclusion

In this paper, we have presented CCM-based OFDM system to reduce PAPR for sparse source. By using the CCM scheme in the traditional OFDM system, the generated frequency OFDM symbols are concentrated in zero, and the sparser

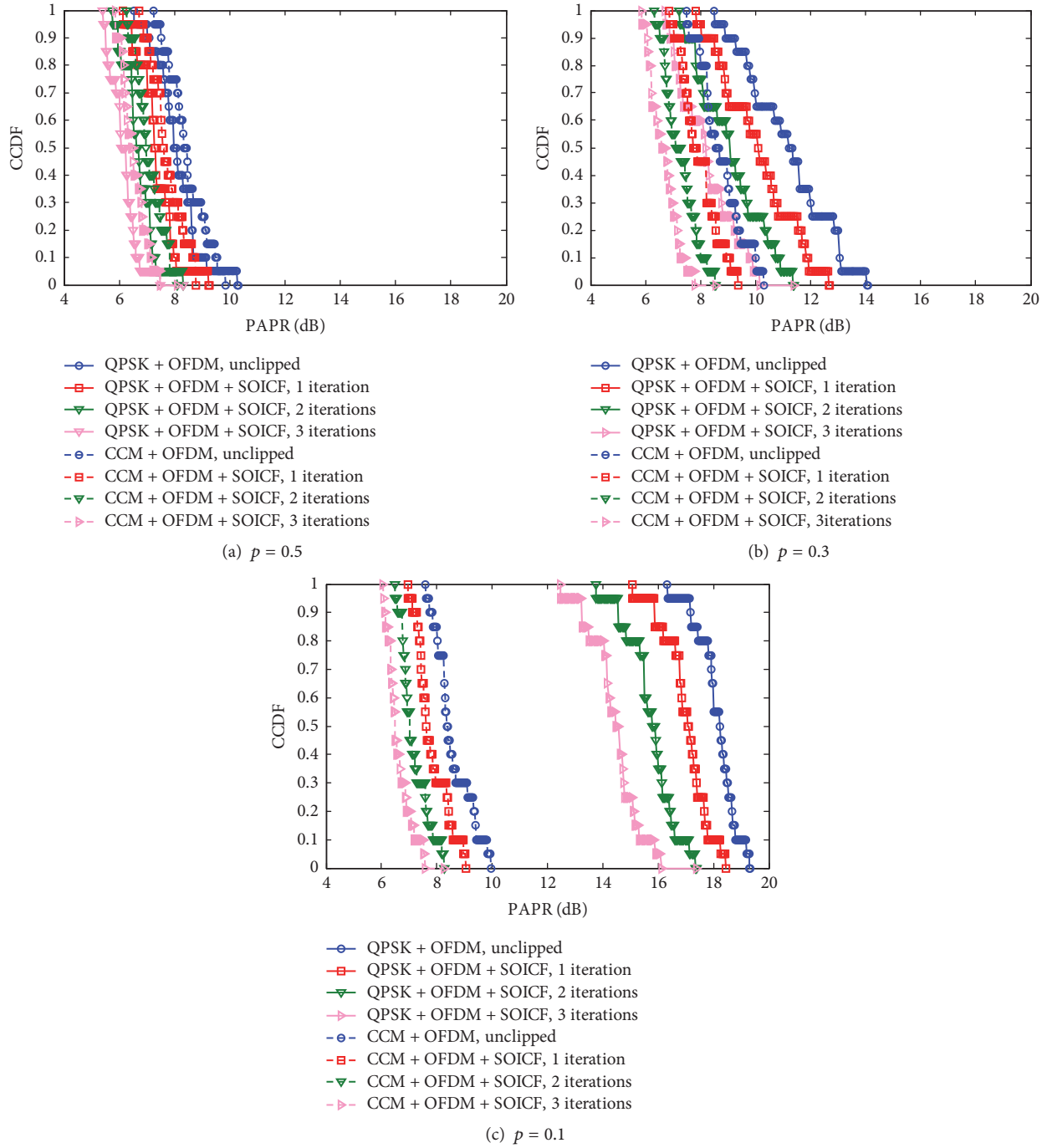



Figure 7: PAPR reduction with different source sparsity and iterations.

the source is, the more zero symbols there are. The zero symbols will not bring additional PAPR increase. Thus, the proposed CCM-based OFDM system, together with iterative clipping and filtering, can efficiently restrain the high PAPR for sparse source. Finally, the simulation results show that the proposed OFDM system almost has the same performance in terms of PAPR reduction as the traditional OFDM system regardless of source sparsity. Moreover, the BER performance of the proposed OFDM system is improved by the use of CCM scheme. Particularly, the proposed OFDM system outperforms about 4 dB over the traditional OFDM system when source sparsity is 0.1. In addition, the proposed CCM-based OFDM system also maintains the characteristic of seamless rate adaptation.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

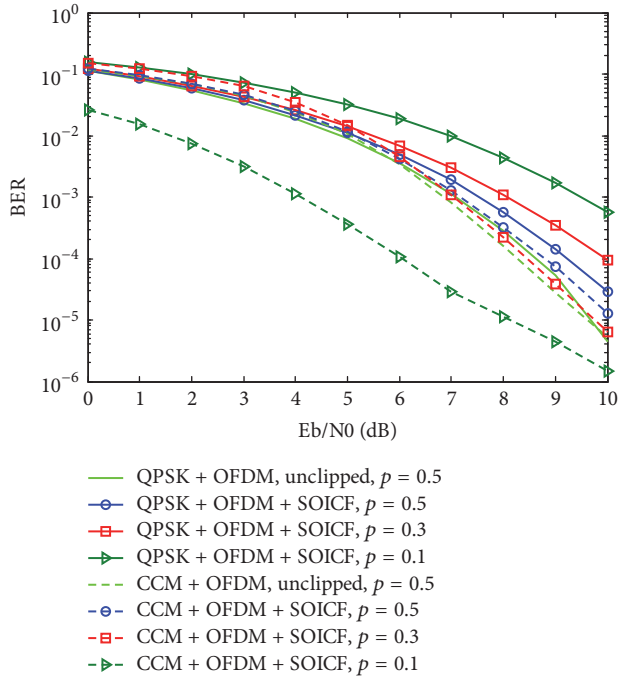



Figure 8: BER performance for different source sparsity.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant 61471022) and NSAF (Grant U1530117).

## References

[1] L. J. Cimini, "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Transactions on Communications*, vol. 33, no. 7, pp. 665–675, 1985.

[2] J. A. C. Bingham, "Multicarrier modulation for data transmission: an idea whose time has come," *IEEE Communications Magazine*, vol. 28, no. 5, pp. 5–14, 1990.

[3] C. Ru, L. Yin, J. Lu, and C. W. Chen, "UEP video transmission based on dynamic resource allocation in MIMO OFDM system," in *Proceedings of the 2007 IEEE Wireless Communications and Networking Conference, (WCNC '07)*, pp. 310–315, China, March 2007.

[4] I. Koffman and V. Roman, "Broadband wireless access solutions based on OFDM access in IEEE 802.16," *IEEE Communications Magazine*, vol. 40, no. 4, pp. 96–103, 2002.

[5] Z. Fei, C. Xing, N. Li, Y. Han, D. Danev, and J. Kuang, "Power allocation for OFDM-based cognitive heterogeneous networks," *Science China Information Sciences*, vol. 56, no. 4, pp. 1–10, 2013.

[6] Z. Chen, L. Yin, Y. Pei, and J. Lu, "CodeHop: physical layer error correction and encryption with LDPC-based code hopping," *Science China Information Sciences*, vol. 59, no. 10, Article ID 102309, 2016.

[7] P. Wang, L. Yin, and J. Lu, *An efficient helicopter-satellite communication scheme based on check-hybrid ldpc coding, Tsinghua science and technology*, 2018.

[8] J. Joung, C. K. Ho, K. Adachi, and S. Sun, "A survey on power-amplifier-centric techniques for spectrum- and energy-efficient wireless communications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 315–333, 2015.

[9] G. Wunder, R. F. H. Fischer, H. Boche, S. Litsyn, and J.-S. No, "The PAPR problem in OFDM transmission: new directions for a long-lasting problem," *IEEE Signal Processing Magazine*, vol. 30, no. 6, pp. 130–144, 2013.

[10] Y. Rahmatallah and S. Mohan, "Peak-to-average power ratio reduction in ofdm systems: a survey and taxonomy," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1567–1592, 2013.

[11] J. Ji, G. Ren, and H. Zhang, "PAPR reduction of sc-fdma signals via probabilistic pulse shaping," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 9, pp. 3999–4008, 2015.

[12] S.-H. Wang, W.-L. Lin, B.-R. Huang, and C.-P. Li, "PAPR reduction in OFDM systems using active constellation extension and subcarrier grouping techniques," *IEEE Communications Letters*, vol. 20, no. 12, pp. 2378–2381, 2016.

[13] H. Wang, X. Wang, L. Xu, and W. Du, "Hybrid PAPR reduction scheme for FBMC/OQAM systems based on multi data block PTS and TR methods," *IEEE Access*, vol. 4, no. 99, pp. 4761–4768, 2016.

[14] S. Shu, D. Qu, L. Li, and T. Jiang, "Invertible subset QC-LDPC codes for PAPR reduction of ofdm signals," *IEEE Transactions on Broadcasting*, vol. 61, no. 2, pp. 290–298, 2015.

[15] J. Armstrong, "Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering," *IEEE Electronics Letters*, vol. 38, no. 5, pp. 246-247, 2002.

[16] L. Wang and C. Tellambura, "A simplified clipping and filtering technique for PAR reduction in OFDM systems," *IEEE Signal Processing Letters*, vol. 12, no. 6, pp. 453–456, 2005.

[17] R. J. Baxley, C. Zhao, and G. T. Zhou, "Constrained clipping for crest factor reduction in OFDM," *IEEE Transactions on Broadcasting*, vol. 52, no. 4, pp. 570–575, 2006.

[18] J. Tong, L. Ping, Z. Zhang, and V. K. Bhargava, "Iterative soft compensation for OFDM systems with clipping and superposition coded modulation," *IEEE Transactions on Communications*, vol. 58, no. 10, pp. 2861–2870, 2010.

[19] Y.-C. Wang and Z.-Q. Luo, "Optimized iterative clipping and filtering for PAPR reduction of OFDM signals," *IEEE Transactions on Communications*, vol. 59, no. 1, pp. 33–37, 2011.

[20] X. Zhu, W. Pan, H. Li, and Y. Tang, "Simplified approach to optimized iterative clipping and filtering for PAPR reduction of OFDM signals," *IEEE Transactions on Communications*, vol. 61, no. 5, pp. 1891–1901, 2013.

[21] H. Cui, C. Luo, J. Wu, C. W. Chen, and F. Wu, "Compressive coded modulation for seamless rate adaptation," *IEEE Transactions on Wireless Communications*, vol. 12, no. 10, pp. 4892–4904, 2013.

[22] T. Jiang, M. Guizani, H.-H. Chen, W. Xiang, and Y. Wu, "Derivation of PAPR distribution for OFDM wireless systems based on extreme value theory," *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, pp. 1298–1305, 2008.

[23] G. Caire, S. Shamai, A. Shokrollahi, and S. Verdù, "Fountain codes for lossless data compression," in *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 68, pp. 1–18, 2005.

[24] C. Tellambura, "Computation of the continuous-time PAR of an OFDM signal with BPSK subcarriers," *IEEE Communications Letters*, vol. 5, no. 5, pp. 185–187, 2001.