Transmission of Bernoulli Sources Using Convolutional LDGM Codes

Yixin Wang, Tingting Zhu and Xiao Ma, Member, IEEE

Abstract

We propose in this paper to exploit convolutional low density generator matrix (LDGM) codes for transmission of Bernoulli sources over binary-input output-symmetric (BIOS) channels. To this end, we present a new framework to prove the coding theorems for linear codes, which unifies the channel coding theorem, the source coding theorem and the joint source-channel coding (JSCC) theorem. In the presented framework, the systematic bits and the corresponding parity-check bits play different roles. Precisely, the noisy systematic bits are used to limit the list size of typical codewords, while the noisy parity-check bits are used to select from the list the maximum likelihood codeword. This new framework for linear codes allows that the systematic bits and the parity-check bits are transmitted in different ways and over different channels. With this framework, we prove that the Bernoulli generator matrix codes (BGMCs) are capacity-achieving over BIOS channels, entropy-achieving for Bernoulli sources, and also system-capacity-achieving for JSCC applications. A lower bound on the bit-error rate (BER) is derived for linear codes, which can be used to predict the error floors and hence serves as a simple tool to design the JSCC system. Numerical results show that the convolutional LDGM codes perform well in the waterfall region and match well with the derived error floors, which can be lowered down if required by simply increasing the encoding memory.

Index Terms

Coding theorem, convolutional low density generator (LDGM) codes, joint source-channel coding (JSCC), linear codes, partial error exponent, partial mutual information

Corresponding author: Xiao Ma. This work was supported by the National Key R&D Program of China (Grant No. 2021YFA1000500) and the NSF of China (No. 61971454). Part of this work is accepted by 2022 International Symposium on Information Theory.

The authors are with the Guangdong Key Laboratory of Information Security Technology, School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China (e-mail: wangyx58@mail2.sysu.edu.cn, zhutt@mail2.sysu.edu.cn, maxiao@mail.sysu.edu.cn)
I. Introduction

Shannon’s separation theorem [1] states that, with unbounded block length (delay) and complexity, the channel coding and the source coding can be separated without loss of optimality. Evidently, for most applications in the non-asymptotic regime with limited delay and complexity, a joint source-channel coding (JSCC) can be more attractive. A commonly accepted JSCC scheme consists of two component codes, one as the source code for compression and the other as the channel code for error correction, where the redundancy left by the source encoder can be exploited by the channel decoder to improve the system performance. Typical constructions include the double low-density parity-check (D-LDPC) JSCC and the double polar (D-Polar) JSCC. The D-LDPC JSCC was first proposed in [2, 3] and later evolved as double protograph LDPC (DP-LDPC) joint source-channel codes (JSCCs) [4] and double spatially coupled LDPC (D-SCLDPC) JSCCs [5] by taking either protograph LDPC codes or SC-LDPC codes as the component codes. The D-Polar JSCC was investigated in [6], where a turbo-like belief propagation (BP) decoder was proposed. In any case, the two component codes need to be designed jointly, which usually involves a complicated optimization algorithm such as the extended curve-fitting algorithm [7] and multi-objective differential evolution algorithm [8]. Since a concatenation of two linear codes is still linear, it is reasonable to develop a single linear code to achieve JSCC, as an attempt in [9]. In this paper, we will first prove in theory that one single linear code (even with low density generator matrix (LDGM)) is sufficient to achieve the JSCC limit and then provide concrete constructions for illustration. The most distinguished feature of the proposed system is that it does not require complicated optimization.

Linear codes play an important role in both the channel coding theory [10] and the source coding theory [11]. In [12], it was proved that the totally random linear code ensemble can achieve the capacity of binary-input output-symmetric (BIOS) channels. The same theorem was proved in [10] by deriving the error exponent. Systematic linear codes have the information bits in the codewords, which can benefit the encoding and decoding procedure compared with non-systematic linear codes. Of the same codeword length, systematic linear codes can have less operating steps in the coding procedure. More importantly, using systematic instead of non-systematic linear codes allows the decoder to obtain the decoded bits directly from the received sequences. However, most existing coding theorems are proved for non-systematic codes and direct proofs are rarely found for systematic codes.
In this paper, we propose to transmit reliably a Bernoulli source over a binary-input output symmetric (BIOS) channel by employing a single linear code, which integrates the source coding and the channel coding as a seamless system. To this end, we propose a new framework to prove the coding theorems for linear codes, which unifies the channel coding theorem, the source coding theorem and the JSCC theorem. In the presented framework, the systematic bits and the corresponding parity-check bits play different roles. Precisely, the noisy systematic bits are used to limit the list size of typical codewords, while the noisy parity-check bits are used to select from the list the maximum likelihood codeword. This new framework for linear codes allows that the systematic bits and the parity-check bits are transmitted in different ways and over different channels. With this framework, we prove that the Bernoulli generator matrix codes (BGMCs) are capacity-achieving over BIOS channels, entropy-achieving for Bernoulli sources, and also system-capacity-achieving for JSCC applications. A lower bound on the bit-error rate (BER) is also derived for linear codes, which can be used to predict the error floors and hence serves as a simple tool to design the JSCC system. For the simulations in JSCC, we consider a special class of linear codes called convolutional LDGM codes. Numerical results show that the convolutional LDGM codes are flexible to construct and have predictable error floors. The convolutional LDGM codes also perform well in the waterfall region with about one dB away from the system capacity.

The rest of the paper is organized as follows. In Section II, we describe the new framework and give an overall coding theorem for the framework. In Section III, we prove the coding theorem for the framework with BGMCs. In Section IV, the systematic capacity for transmitting Bernoulli sources is derived from the coding theorem and the performance lower bound for linear codes in JSCC scheme is also derived. We use the convolutional LDGM codes in the framework for transmitting Bernoulli sources. Numerical results for various parameter settings show that the convolutional LDGM codes have flexible construction, good performance in the waterfall region and match well with the lower bound in the error floor region. Finally, some concluding remarks are given in Section V.

In this paper, a random variable is denoted by an upper-case letter, say $X$, whose realization is denoted by the corresponding lowercase letter $x \in \mathcal{X}$. We use $P_X(x), x \in \mathcal{X}$ to represent the probability mass (or density) function of a random discrete (or continuous) variable. For a vector of length $m$, we represent it as $\mathbf{x} = (x_0, x_1, \cdots, x_{m-1})$. We also use $\mathbf{x}^m$ to emphasize the length of $\mathbf{x}$. We denote by $\mathbb{F}_2 = \{0, 1\}$ the binary field.
Fig. 1. A system model with linear coding. For the source coding, $V^k$ is punctured and the code rate is $m/k$, while for the channel coding, the code rate is defined as usual $k/(k + m)$.

II. System model and Problem Statement

A. Systematic Linear Codes

We consider a system model that is depicted in Fig. 1, where $U^k \in \mathbb{F}_2^k$ is referred to as the information bits to be transmitted, $G$ is a binary matrix of size $k \times m$ and $X^m = U^k G \in \mathbb{F}_2^m$ is referred to as parity-check bits corresponding to $U^k$. The two channels, channel 1 and channel 2 are independent and can be different. For the channel coding, channel 1 and channel 2 are typically the same. For the source coding, $V^k$ is not required and channel 1 can be viewed as an erasure channel with erasure probability one while channel 2 is a noiseless channel. The model can also be adapted to more general cases such as the JSCC. Even more generally, both channel 1 and channel 2 can be integrated with high-order modulations and demodulations. The task of the receiver is to recover $U^k$ from $(V^k, Y^m)$.

In this paper, we focus on Bernoulli sources and BIOS memoryless channels. A Bernoulli source (also referred to as a Bernoulli sequence or a Bernoulli process) is a sequence of independent and identically distributed (i.i.d.) binary random variables. A BIOS memoryless channel is characterized by the input $x \in \mathcal{X} = \mathbb{F}_2$, an output set $\mathcal{Y}$ (discrete or continuous), and a conditional probability mass (or density) function $\{P_{Y|X}(y|x), x \in \mathbb{F}_2, y \in \mathcal{Y}\}$ which satisfies the symmetric condition that $P_{Y|X}(y|1) = P_{Y|X}(\pi(y)|0)$ for some mapping $\pi : \mathcal{Y} \rightarrow \mathcal{Y}$ with $\pi^{-1}(\pi(y)) = y$. The channel is said to be memoryless if $P_{Y|X}(y|x) = \prod_{i=0}^{n-1} P_{Y|X}(y_i|x_i)$. For a Bernoulli source $U$, we can define the conditional entropy $H(U|V)$. For a Bernoulli input $X$, we can define the mutual information $I(X; Y)$. Obviously, the parity-check bits are necessary only when $H(U|V) > 0$.

1If the context is clear, we may omit the subscript of the probability mass (or density) function.
B. A General Coding Theorem for Linear Codes

Theorem 1: Let \( \mathbf{U} \) be a Bernoulli source and \( \mathbf{G} \) be a totally random matrix, whose elements are generated independently and identically according to the Bernoulli distribution with success probability \( \rho = 1/2 \). For any fixed nonnegative number \( r < I(X; Y)/H(U|V) \), there exists a matrix \( \mathbf{G} \) of sufficiently large size and a decoding algorithm such that \( k/m \geq r \) and the frame-error rate (FER) \( \Pr\{\hat{U} \neq U\} \) is arbitrarily small.

Outline Proof of Theorem 1: Let \( \epsilon \) be an arbitrarily small but fixed positive number. Upon receiving \((v, y)\), we conduct the following two-step decoding. First, list all sequences \( \tilde{u} \) such that \((\tilde{u}, v)\) are jointly typical. Second, find from the list a sequence \( \hat{u} \) such that \( P(y|\hat{u}\mathbf{G}) \) is maximized.

There are two types of errors. One is the case when \( u \) is not in the list and the other is the case when \( u \) is in the list but is not the most likely one. The former error event can have arbitrarily small probability since \((U, V)\) are jointly typical with arbitrarily high probability for sufficiently large \( k \). The latter error event can also have arbitrarily small probability as long as the list size (which is upper bounded by \( \exp[kH(U|V) + 2\epsilon] \)) is less than the capacity of channel 2. This can be fulfilled by choosing sufficiently large \( k \) and \( m \) such that \( k/m \geq r \) but \( k/m < I(X; Y)/H(U|V) \) since \( r < I(X; Y)/H(U|V) \) from the assumption of the theorem.

Remarks: The above theorem states that the ratio \( k/m \) can be arbitrarily close to \( I(X; Y)/H(U|V) \). Actually, we can prove that, for \( \Pr\{\hat{U} \neq U\} \to 0 \),

\[
\lim_{k \to \infty} \lim_{m \to \infty} \frac{k}{m} = \frac{C}{H(U|V)},
\]

where \( C \) is the channel capacity of channel 2.

- For the Bernoulli source with \( P(0) = P(1) = 1/2 \) and the same channel 1 and channel 2, the condition \( k/m < I(X; Y)/H(U|V) \) is equivalent to the conventional condition \( k/(m+k) < I(X; Y) \), where \( k/(k+m) \) is the code rate for channel coding. This can be verified by noting that \( I(X; Y) = 1 - H(X|Y) \) and \( H(U|V) = H(X|Y) \).

- For the source coding, channel 1 is an erasure channel and channel 2 is noiseless. Hence, the condition \( k/m < I(X; Y)/H(U|V) \) is equivalent to the conventional condition \( m/k > H(U) \), where \( m/k \) is the code rate for the source coding. This can be verified by noting that \( H(U|V) = H(U) \) and \( I(X; Y) = 1 \).
From the proof, we see that the systematic bits and the parity-check bits play different roles. Receiving noisy systematic bits provides us a list of the source output, while receiving noisy parity-check bits helps us to select the correct one from the list.

It is well-known that the linear codes defined by general matrices $G$ have no efficient decoding algorithms for large $k$ and $m$. Hence, we consider the following LDGM code ensemble [13], which is referred to as Bernoulli generator matrix code (BGMC).

**BGMC ensemble:** A BGMC transforms $u \in \mathbb{F}_2^k$ into $(u, x)$ by $x = uG \in \mathbb{F}_2^m$, where $G$ is a random matrix of size $k \times m$ with each element $G_{i,j}$ ($0 \leq i \leq k-1$, $0 \leq j \leq m-1$) being generated independently according to the Bernoulli distribution with success probability $\Pr\{G_{i,j} = 1\} = \rho \leq 1/2$. For $\rho \ll 1/2$, the BGMC ensemble is a class of LDGM codes.

It has been proved that, in terms of BER, the BGMC ensemble is capacity-achieving for BIOS memoryless channels [13, 14] and entropy-achieving for Bernoulli sources [15]. In this paper, we show that Theorem 1 also holds in terms of FER even for $\rho < 1/2$. The proof is definitely applicable to the case $\rho = 1/2$, indicating that the outline proof of Theorem 1 will be detailed in the next section.

### III. Coding Theorem for BGMCs

#### A. Partial Mutual Information

Let $P(1) = p$ and $P(0) = 1 - p$ be an input distribution of a BIOS memoryless channel. The mutual information between the input and the output is given by

$$I(p) = (1 - p)I_0(p) + pI_1(p),$$

where

$$I_0(p) = \sum_{y \in \mathcal{Y}} P(y|0) \log \frac{P(y|0)}{P(y)},$$

$$I_1(p) = \sum_{y \in \mathcal{Y}} P(y|1) \log \frac{P(y|1)}{P(y)},$$

and $P(y) = (1 - p)P(y|0) + pP(y|1)$. We define $I_0(p)$ (or $I_1(p)$) as partial mutual information. For a BIOS memoryless channel, we have $\max_{0 \leq p \leq 1} I(p) = I(1/2) = I_0(1/2) = I_1(1/2)$, which is the channel capacity. Notice that $I_0(p) > 0$ for $0 < p < 1$ as long as $\Pr\{y|P(y|0) \neq P(y|1)\} > 0$. This is a natural assumption in this paper.
**Lemma 1:** The partial mutual information $I_0(p)$ is continuous, differentiable and strictly increasing from $I_0(0) = 0$ to the capacity $I_0(1/2)$.

**Proof:** It can be easily seen that the partial mutual information is continuous and differentiable for $0 \leq p \leq 1/2$. By carrying out the differentiation, we can verify that partial mutual information is strictly increasing from $I_0(0) = 0$ to the capacity $I_0(1/2)$.

To prove the main theorem, we also need the following lemmas.

**Lemma 2:** For the BGMC ensemble, the parity-check vector corresponding to an information vector with weight $\omega$ is a Bernoulli sequence with success probability

$$\rho_\omega \triangleq \Pr\{X_j = 1|W_H(U) = \omega\} = \frac{1 - (1 - 2\rho)^\omega}{2},$$

where $W_H(\cdot)$ is the Hamming weight function. Furthermore, for any given positive integer $T \leq k$,

$$P(x|u) \triangleq \Pr\{X = x|U = u\} \leq P(0|u) \leq (1 - \rho_T)^m,$$

for all $u \in \mathbb{F}_2^k$ with $W_H(u) \geq T$ and $x \in \mathbb{F}_2^m$.

**Proof:** See the proof in [14, Lemma 1] and it is omitted here.

**Remark:** Lemma 2 states that, for information vector $u$ with high weight, the corresponding parity check vector is convergent in distribution to a Bernoulli process with success probability $1/2$, since $\rho_\omega \to 1/2$ as $\omega \to \infty$.

**Lemma 3:** Given a sequence $v$ of length $k$, for any $\epsilon > 0$, define $A_{\epsilon}^{(k)}(U|v)$ the set of $u$ sequences which are jointly $\epsilon$-typical with $v$. If $v$ is typical, then for sufficiently large $k$, the cardinality of $A_{\epsilon}^{(k)}(U|v)$ can be upper bounded by

$$|A_{\epsilon}^{(k)}(U|v)| \leq \exp\left[k(H(U|V) + 2\epsilon)\right].$$

**Proof:** See [16, Theorem 15.2.2] and it is omitted here.

B. Partial Error Exponent

The error exponent was derived to prove the channel coding theorem by assuming that all codewords are randomly generated according to an identical distribution, where the pair-wise independence between codewords is sufficient [10]. In this paper, we derive the partial error exponent for BIOS channels by assuming that the codeword 0 is transmitted. The derivation suggests that even the pair-wise independence is not required.
Lemma 4: Suppose that the codeword $0 \in \mathbb{F}_2^n$ is transmitted over a BIOS channel. Let $\mathcal{L} = \{x_1, x_2, \cdots, x_M\}$ be a random list, where $x_i \in \mathbb{F}_2^n$ is a segment of a Bernoulli process with success probability $p$. Then the probability that there exists some $i$ such that $x_i$ is more likely than $0$, denoted by $\Pr\{\text{error}|0\}$, can be upper bounded by

$$\Pr\{\text{error}|0\} \leq \exp\left(-nE(p, R)\right), \quad (8)$$

where

$$R = \frac{1}{n} \log M, \quad (9)$$

and

$$E(p, R) = \max_{0 \leq \gamma \leq 1} (E_0(\gamma, p) - \gamma R), \quad (10)$$

and

$$E_0(\gamma, p) = -\log \left\{ \sum_{y \in \mathcal{Y}} P(y|0)^{\frac{1}{1+\gamma}} \left[ (1-p)(P(y|0))^{\frac{1}{1+\gamma}} + p(P(y|1))^{\frac{1}{1+\gamma}} \right]^\gamma \right\}. \quad (11)$$

Furthermore, $E(p, R) > 0$ if $0 < R < I_0(p)$.

Proof: Denote by $E_i$ the event that $x_i$ is more likely than $0$ given a received sequence $y$.

For the decoding error, we have

$$\Pr[\text{error}|0] = \sum_{y \in \mathcal{Y}^n} P(y|0) \cdot \Pr\left( \bigcup_{i=1}^M E_i \right) \leq M^\gamma \sum_{y \in \mathcal{Y}^n} P(y|0) \left( \Pr\{P(y|x) \geq P(y|0)\} \right)^\gamma, \quad (12)$$

for any given $0 \leq \gamma \leq 1$. From Markov inequality, for $s = 1/(1+\gamma)$ and a given received vector $y$, the probability of a vector being more likely than $0$ is upper bounded by

$$\Pr\{P(y|x) \geq P(y|0)\} \leq \frac{\mathbb{E}\left[\left(P(y|x)\right)^s\right]}{\left(P(y|0)\right)^s} = \sum_x P(x) \frac{(P(y|x))^s}{(P(y|0))^s}. \quad (13)$$

Substituting this bound into (12), we have

$$\Pr[\text{error}|0] \leq M^\gamma \sum_{y \in \mathcal{Y}^n} P(y|0) \left[ \sum_x P(x) \frac{(P(y|x))^s}{(P(y|0))^s} \right]^\gamma \leq M^\gamma \prod_{i=0}^{n-1} \left\{ \sum_{y_i \in \mathcal{Y}} P(y_i|0)^{1-s}\gamma \left[ \sum_{x_i, y_i \in \mathbb{F}_2} P(x_i)(P(y_i|x_i))^s \right]^\gamma \right\}$$

$$\leq \exp\left(-nE(p, R)\right). \quad (14)$$
where the equality (*) follows from the memoryless channel assumption and the inequality (**) follows by recalling that $s = 1/(1 + \gamma)$ and denoting

$$E(p, R) = \max_{0 \leq \gamma \leq 1} (E_0(\gamma, p) - \gamma R),$$

and

$$E_0(\gamma, p) = -\log \left\{ \sum_{y \in \mathcal{Y}} P(y|0)^{\frac{1}{1+\gamma}} \left[ (1 - p)(P(y|0))^{\frac{1}{1+\gamma}} + p(P(y|1))^{\frac{1}{1+\gamma}} \right]^\gamma \right\}. \tag{16}$$

Considering $E_0(\gamma, p) - \gamma R$ for a given $p$, we have $E_0(0, p) - 0 = 0$ and

$$\frac{\partial E_0(\gamma, p)}{\partial \gamma} \bigg|_{\gamma=0} = I_0(p) - R. \tag{17}$$

Hence, $E(p, R) > 0$ if $R < I_0(p)$.

**Remark:** From the above proof, we see that the members in the list need to have the same distribution but the condition of pair-wise independence is not necessary, which is distinguished from the proof [10, Chapter 5].

**C. Coding Theorem for BGMCs**

**Theorem 2:** Consider the BGMC ensemble and let $U$ be a Bernoulli source. For any fixed nonnegative number $r < I(X; Y)/H(U|V)$, there exists a matrix $G$ of sufficiently large size and a decoding algorithm such that $k/m \geq r$ and the FER is arbitrarily small.

**Proof:** Suppose that $(u^k, u^kG)$ is transmitted. Let $\epsilon > 0$ be an arbitrarily small number. Upon receiving $(v^k, y^m)$, we use the following two-step decoding. First, list all sequences $\tilde{u}$ such that $(\tilde{u}, v)$ are jointly typical. Second, find from the list a sequence $\hat{u}$ such that $P(y|\hat{u}G)$ is maximized.

There are two types of errors. One is when $(u^k, v^k)$ are not jointly $\epsilon$—typical and hence $u^k$ is not in the list. This type of errors, from [16, Chapter 7], can have arbitrarily small probability as long as $k$ is sufficiently large.

The other case is that $u^k$ is in the list but is not the most likely one. In this case, denote the list as $\tilde{\mathcal{L}} = \{\tilde{u}_0 = u^k, \tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_M\}$. From Lemma 3, we have $M \leq \exp \left[ k(H(U|V) + 2\epsilon) \right]$. Given the received sequence $y$ and the list $\tilde{\mathcal{L}}$, the decoding output $\hat{U}$ is a random sequence over the code ensemble due to the randomness of $G$. Denote by $E_{u,i}$ the event that $\tilde{u}_iG$ is more likely than $u^kG$ given a received sequence $y$. We have

$$\Pr\{\text{error}|u^k\} \leq \epsilon + \sum_{x \in \mathcal{F}^m} \sum_{y \in \mathcal{Y}^m} \Pr\{u^kG = x\} \cdot P(y|x) \cdot \Pr\left\{\bigcup_{i=1}^M E_{u,i}\right\}. \tag{18}$$
Given \( y \in \mathcal{Y}^m \), \( x \in \mathbb{F}_2^m \), we define
\[
\pi^x(y_i) = \begin{cases} 
    y_i & \text{if } x_i = 0 \\
    \pi(y_i) & \text{if } x_i = 1
\end{cases}.
\] (19)

Then we have \( P(y|x) = P(\pi^x(y)|0) \) and \( P(y|\bar{u}G) \geq P(y|uG) \) is equivalent to \( P(\pi^x(y)|(u^k + \bar{u})G) \geq P(\pi^x(y)|0) \). Therefore, the performance can be analyzed by assuming that 0 is transmitted and competed with a list of codewords \( u_iG \) at the decoder, where \( u_i = u^k + \bar{u}_i \) for \( 1 \leq i \leq M \). With this equivalence, we have
\[
\sum_{x \in \mathbb{F}_2^m} \Pr\{u^kG = x\} \sum_{y \in \mathcal{Y}^m} P(y|x) \cdot \Pr\big\{ \bigcup_{i=1}^M E_{u,i}\big\}
\]
\[
= \sum_{y \in \mathcal{Y}^m} P(y|0) \cdot \Pr\big\{ \bigcup_{i=1}^M E_i\big\},
\]
where \( E_i \) is the event that \( u_iG \) is more likely than 0 given a received sequence \( y \).

We partition the list \( \mathcal{L} = \{0, u_1, u_2, \ldots, u_M\} \) according to the weight of \( u_i \) \( (0 \leq i \leq M) \) and denote by \( \mathcal{L}_\omega \) all the sequences of \( u_i \in \mathcal{L} \) with \( W_H(u_i) = \omega \). Thus, we have
\[
\mathcal{L} = \bigcup_{\omega=0}^k \mathcal{L}_\omega,
\]
and
\[
|\mathcal{L}_\omega| \leq \binom{k}{\omega}.
\]

For any positive integer \( T < k \), the error event can be split into two sub-events depending on whether \( W_H(U) \geq T \) or not. Thus, we have
\[
\Pr\{ \bigcup_{i=1}^M E_i \} \leq \sum_{\omega=1}^{T-1} \left( \sum_{u_i:W_H(u_i)=\omega} \Pr\{P(y|u_iG) \geq P(y|0)\} \right)^\gamma + \left( \sum_{u_i:W_H(u_i) \geq T} \Pr\{P(y|u_iG) \geq P(y|0)\} \right)^\gamma,
\]
for any \( 0 \leq \gamma \leq 1 \).

For \( \omega \geq 1 \), we define \( \text{FER}(\omega) \) as
\[
\text{FER}(\omega) = \sum_{y \in \mathcal{Y}^m} P(y|0) \cdot \left( \sum_{u_i:W_H(u_i)=\omega} \Pr\{P(y|u_iG) \geq P(y|0)\} \right)^\gamma.
\]
(24)
We know from Lemma 2 that, for any given $u_i$ with $W_H(u_i) = \omega$, the parity-check vector $x$ is a segment of Bernoulli process with success probability $\rho_\omega$. Hence, the conditional probability mass function $P(x|u_i)$ for $u_i \in \mathcal{L}_\omega$ is the same, denoted by $P_\omega(x)$. We have

$$\text{FER}(\omega) \leq \sum_{y \in \mathcal{Y}^m} P(y|0) \left[ \left( \gamma \right) \sum_{x \in \mathcal{F}_2^m} P_\omega(x) \frac{(P(y|x))^s}{(P(y|0))^s} \right]$$

$$\leq \exp \left( -mE(\rho_\omega, R_\omega) \right),$$

where the inequality (*) follows from (22), the proof of Lemma 4 and by denoting

$$R_\omega = \frac{1}{m} \log \left( \frac{k}{\omega} \right).$$

For $W_H(u_i) \geq T$, we have

$$\Pr\{P(y|u_i,G) \geq P(y|0)\} \leq \frac{\mathbb{E}[(P(y|u_i,G))^s]}{(P(y|0))^s}$$

$$= \sum_{x \in \mathcal{F}_2^n} P(x|u_i) \frac{(P(y|u_i,G))^s}{(P(y|0))^s}$$

$$\leq \left[ \frac{1 + (1 - 2\rho)^T}{2} \right]^m \sum_{x \in \mathcal{F}_2^n} (P(y|u_i,G))^s \frac{(P(y|0))^s}{(P(y|0))^s},$$

where the inequality follows from Lemma 2. Thus, we have

$$\sum_{y \in \mathcal{Y}^m} P(y|0) \left( \sum_{u_i:W_H(u_i) \geq T} \Pr\{P(y|u_i,G) \geq P(y|0)\} \right)$$

$$\leq \sum_{y \in \mathcal{Y}^m} P(y|0) \left( \sum_{\omega > T} \left| \mathcal{L}_\omega \right| \left[ \frac{1 + (1 - 2\rho)^T}{2} \right]^m \sum_{x \in \mathcal{F}_2^n} (P(y|x))^s \right)^{\gamma}$$

$$\leq \exp \left[ k\gamma (H(U|V) + 2\epsilon) \right] \left[ 1 + (1 - 2\rho)^T \right]^{m\gamma} \left( \sum_{y_i \in \mathcal{Y}} (P(y_i|0))^{1-s\gamma} \left( \sum_{x_i \in \mathcal{F}_2^n} \frac{1}{2} \cdot (P(y_i|x_i))^s \right)^{\gamma} \right)^m$$

$$\leq \exp \left( -mE\left( \frac{1}{2}, R_T \right) \right),$$

where the inequality (*) follows from Lemma 3 and the memoryless BIOS channel assumption, and the inequality (**) follows the proof of Lemma 4 and by denoting

$$R_T = \log[1 + (1 - 2\rho)^T] + \frac{k}{m} (H(U|V) + 2\epsilon).$$
Thus, we have
\[
\text{Pr}\{\text{error}|\mathbf{u}^k\} \leq \epsilon + \sum_{\omega=1}^{T} \exp \left( -mE(\rho_\omega, R_\omega) \right) + \exp \left( -mE\left( \frac{1}{2}, R_T \right) \right).
\]
(30)

For the information length $k$, we may choose the parity-check length $m = \lfloor k/r \rfloor$. Now letting $k \to \infty$ and $T \to \infty$, we have $R_T \to r(H(U|V) + 2\epsilon)$ since $\rho \leq 1/2$ and $k/m \to r$. We may choose sufficiently large $k$ (hence $m$) and $T$, such that $R_T < I(X;Y) \leq I_0(1/2)$ since $r < I(X;Y)/H(U|V)$ by assumption. We have from Lemma 4 that $E(1/2, R_T) > 0$ and the third term in the right hand side (RHS) of the inequality (30) can be made not greater than $\epsilon$. By fixing $T$ and for $\omega < T$, since $m$ increases linearly with $k$ but $\log^k(\omega)$ increases only logarithmically with $k$, we have $R_\omega \to 0$ as $k \to \infty$. Since $I(\rho_\omega) > 0$, we have $E(\rho_\omega, R_\omega) > 0$ for sufficiently large $k$ and hence $m$, implying that the second term in the RHS of (30) can also be made not greater than $\epsilon$. Now we have
\[
\text{Pr}\{\text{error}|\mathbf{u}^k\} \leq 3\epsilon.
\]
(31)

Therefore, $\text{Pr}\{\text{error}\} = \sum_{\mathbf{u}^k \in F_2^k} \text{Pr}\{\text{error}|\mathbf{u}^k\} \leq 3\epsilon$. \hfill \Box

IV. Simulation Results for JSCC

We have proved in Theorem 2 that, like the totally random linear codes, BGMCs can achieve the theoretical limits for binary sources and BIOS channels. For $\rho \ll 1/2$, the BGMCs are low density generator (LDGM) codes. However, even with this setup, the maximum-likelihood (ML) decoding is not implementable, so we turn to a very special class of convolutional LDGM codes [14] in this section.

A. Convolutional LDGM Codes

The generator matrix of a time-invariant convolutional LDGM code can be written as
\[
\mathbf{G} = \begin{bmatrix}
\mathbf{S}_0 & \mathbf{S}_1 & \cdots & \mathbf{S}_{\nu-1} & \mathbf{S}_\nu \\
\mathbf{S}_0 & \mathbf{S}_1 & \cdots & \mathbf{S}_{\nu-1} & \mathbf{S}_\nu \\
\vdots & \ddots & \ddots & \cdots & \ddots \\
\end{bmatrix}
\]
where \( \mathbf{S}_i \) is a random matrix of size $k \times m$ with each column drawn independently and uniformly from the collection of all binary column vectors of weight 1. That is to say, the column weights of \( \mathbf{S}_i \) are confined to be one and hence the generator matrix is very sparse for large $k$. The
corresponding encoder as shown in Fig. 2 is almost the same as that for Ensemble 3 presented in [13] and the slight difference lies in that the matrices $S_i$ are time-invariant without zero columns and the systematic bits are totally punctured. It is worth pointing out that, in the case when $m(\nu + 1) = k\gamma$ for some positive integer $\gamma$, the $\nu + 1$ matrices $S_i$, $0 \leq i \leq \nu$, can be constructed by partitioning equally $\gamma$ random permutation matrices $\Pi_i$, $1 \leq i \leq \gamma$, of order $k$. That is, $[S_0, S_1, \ldots, S_{\nu-1}, S_{\nu}]$ is another form of $[\Pi_1, \Pi_2, \ldots, \Pi_\gamma]$. With this setup, the construction is similar to that presented [17] but the encoder structure here is different from that presented in [17].

The input to the encoder is a sequence $u^{(0)}$, $u^{(1)}$, $\ldots$, where $u^{(i)} \in \mathbb{F}_2^k$. The encoding algorithm of the convolutional LDGM codes is described in Algorithm 1 (see Fig. 2 for reference).

**Algorithm 1** Encoding of convolutional LDGM codes

- **Initialization:** For $t < 0$, let $u^{(t)} = 0 \in \mathbb{F}_2^k$.
- **Iteration:** For $0 \leq t \leq L - 1$,
  - For $0 \leq i \leq \nu$, compute $w^{(t,i)} = u^{(t-i)}S_i \in \mathbb{F}_2^m$.
  - Compute $c^{(t)} = \sum_{i=0}^\nu w^{(t,i)}$.
  - Take $c^{(t)}$ as the encoding output block at time slot $t$.
- **Termination:** For $L \leq t \leq L + m - 1$, set $u^{(t)} = 0 \in \mathbb{F}_2^k$ and compute $c^{(t)}$ following Step Iteration.

For the decoding, we use an iterative sliding window decoding algorithm, which is slightly different from but similar to the decoding algorithm as described in [14, Algorithm 3]. The difference lies in that the systematic digits are totally erased but initialized according to the
source distribution.

It has been shown that the above convolutional LDGM codes perform well as source codes [15] and as channel codes [14]. We will show that the convolutional LDGM codes are also good as JSCCs. As shown in Fig. 3, we consider transmission of a Bernoulli source with binary phase shift keying (BPSK) signalling over an additive white Gaussian noise (AWGN) channels. The binary sparse source delivers $U^{(t)} \in \mathbb{F}_2^k$ at time slot $t$, which is an i.i.d Bernoulli source with success probability $\theta \triangleq \Pr\{U_i^{(t)} = 1\}$. The entropy of this source is given by

$$H(U) = -\theta \log \theta - (1 - \theta) \log(1 - \theta) \triangleq H(\theta). \quad (32)$$

The source block $U^{(t)}$ is encoded by the convolutional LDGM code with the generator matrix $G$ into $X^{(t)} \in \mathbb{F}_2^m$. Then $X^{(t)}$ is modulated using BPSK signalling with 0 and 1 mapped to $+1$ and $-1$, respectively, and transmitted over an AWGN channel characterized by the noise power $\sigma^2$, resulting in $Y^{(t)} \in \mathbb{R}^m$. The task of the receiver is to get the estimation of the source $\hat{U}^{(t)} \in \mathbb{F}_2^k$ from $Y^{(t)}$, $Y^{(t+1)}$, $\ldots$, $Y^{(t+d)}$ using the iterative sliding window decoding algorithm.

We define the code rate of the JSCC system as $R = k/m$ measured in source digits per channel use. From Theorem 1, the supremum of the code rate $R$ can be $C/H(\theta)$, which is measured in source digits per channel use and referred to as the system capacity to distinguish from the conventional channel capacity. Fig. 4 shows the system capacity with different $\theta$, where we can see that the sparser the source is, the more source digits can be carried on average by one channel use. In this paper, $E_s$ refers to the energy per source digit (rather than nonredundant information bit), the same as that in [9], and $N_0 = 2\sigma^2$ refers to the noise power spectral density and $\sigma^2$ is the variance of noise. Thus, the limits of $E_s/N_0$ in Fig. 4 can be less than $-1.59$ dB, which is the asymptotic limit for reliable transmission when the energy is measured per non-redundant information bit, denoted by $E_b$. Actually, we have

$$\frac{E_b}{N_0} = \frac{E_s}{N_0} - 10 \log_{10}(H(\theta)).$$
Fig. 4. System capacity with different $\theta$. Notice that the ordinate is measured in source digits per channel use instead of the commonly-adopted bits per channel use. Hence, the value of the ordinate can be greater than one for sparse sources. The abscissa is measured by $E_s/N_0$ where $E_s$ and $N_0$ refer to the energy per source digit and the noise power spectral density, respectively.

Specially, when the source is uniform, i.e., $\theta = 0.5$, we have $E_s/N_0 = E_b/N_0$.

**B. Performance Bound**

We define the BER which represents the source digit error rate and is referred to as the average Hamming distortion for biased sources, as $\mathbb{E}[W_H(\hat{U}^{(t)} + U^{(t)})]/k$, where $\mathbb{E}[-]$ denotes the expectation of the random variable, $W_H(\cdot)$ denotes the Hamming weight function. We first present a lower bound on the BER of the linear block codes, which is also applicable to the convolutional LDGM codes.

**Theorem 3:** For the BPSK-AWGN channel and sparse source, the BER of a linear block code with generator matrix $G$ of size $k \times m$ can be lower bounded by

$$\text{BER} \geq \frac{1}{k} \sum_{i=0}^{k-1} P_{\theta,W}(\theta, \omega_i),$$

(33)
where \( P_{\theta,W}(\theta,\omega_i) \) is the lower bound on the bit error probability for the \( i \)-th bit, given by

\[
P_{\theta,W}(\theta,\omega_i) = (1 - \theta)Q\left(\frac{\sqrt{\omega_i}}{\sigma} + \frac{\sigma}{2\sqrt{\omega_i}} \ln \frac{1 - \theta}{\theta} \right) + \theta \cdot Q\left(\frac{\sqrt{\omega_i}}{\sigma} - \frac{\sigma}{2\sqrt{\omega_i}} \ln \frac{1 - \theta}{\theta} \right).
\]  

(34)

Here, \( \theta \) is the success probability of the Bernoulli source, \( \omega_i \) is the Hamming weight of the \( i \)-th row of \( G \), \( \sigma^2 \) is the variance of the noise and \( Q(x) \) is the tail probability that the normalized Gaussian random variable takes a value not less than \( x \).

Proof: Let \( \omega_i \) be the Hamming weight of the \( i \)-th row of \( G \). The error probability of the \( i \)-th bit \( U_i \) is lower bounded by that with the genie-aided decoder [18], which assumes that all but \( U_i \) is known. Therefore, the lower bound is equal to the performance of repetition code with length \( \omega_i \). Without loss of generality, we assume that the receiving sequence corresponding to the repetition code is \( y = (y_0, \cdots, y_{\omega_i-1}) \). Upon receiving \( y \), the optimal decision with log-likelihood ratio (LLR) for the repetition code is given by

\[
\hat{U}_i = \begin{cases} 
0, & \text{if } \ln\left(\frac{1 - \theta}{\theta}\right) \sum_{j=0}^{\omega_i-1} \frac{2y_j}{\sigma^2} > 0 \\
1, & \text{if } \ln\left(\frac{1 - \theta}{\theta}\right) \sum_{j=0}^{\omega_i-1} \frac{2y_j}{\sigma^2} < 0
\end{cases}.
\]

(35)

For the optimal decision, the error probability for bit 0 is \( Q\left(\frac{\sqrt{\omega_i}}{\sigma} + \frac{\sigma}{2\sqrt{\omega_i}} \ln \left[\frac{(1 - \theta)}{\theta}\right] \right) \) and for bit 1 is \( Q\left(\frac{\sqrt{\omega_i}}{\sigma} - \frac{\sigma}{2\sqrt{\omega_i}} \ln \left[\frac{(1 - \theta)}{\theta}\right] \right) \). Hence, the error probability for the repetition code with length \( \omega_i \) is

\[
P_{\theta,W}(\theta,\omega_i) = (1 - \theta)Q\left(\frac{\sqrt{\omega_i}}{\sigma} + \frac{\sigma}{2\sqrt{\omega_i}} \ln \frac{1 - \theta}{\theta} \right) + \theta \cdot Q\left(\frac{\sqrt{\omega_i}}{\sigma} - \frac{\sigma}{2\sqrt{\omega_i}} \ln \frac{1 - \theta}{\theta} \right).
\]

(36)

We can get the result in the theorem by averaging the lower bound on BER of \( k \) bits. \( \square \)

Remark: For \( \theta = 1/2 \), the lower bound on BER for the convolutional LDGM codes is reduced to

\[
\text{BER} \geq \frac{1}{k} \sum_{i=0}^{k} Q\left(\frac{\sqrt{\omega_i}}{\sigma} \right),
\]

(37)

which is the same as the result derived in [14, Theorem 2].

We see that the limit performance is dominated by the row weights, which are in turn closely related to the encoding memory \( \nu \). This is confirmed by the following numerical examples, which also show the flexibility and universality of the construction. Particularly, the convolutional LDGM codes are definitely applicable to the case of the uniform source and the case of the noiseless channel, as already demonstrated in [19] and [15].
Fig. 5. The BER performance of convolutional LDGM codes with $k = 1024$, $m = 2048$, $\nu = 7$ in Example 1. The parameter $\theta$ is specified in the legends. The corresponding lower bounds and system capacities are also plotted.

C. Simulation results

Example 1 (Fixed $k$, $m$ and $\nu$, Increasing $\theta$): In this example, we fix $k = 1024$, $m = 2048$, $\nu = 7$ in Fig. 5, $k = 1024$, $m = 1024$, $\nu = 7$ in Fig. 6, $k = 1024$, $m = 512$, $\nu = 50$ in Fig. 7 and change the parameter $\theta$. The simulation results are shown in these figures, in which the system capacities and corresponding lower bounds for different $\theta$ are also plotted. From the figures, we can see that $\theta$ has little influence on the lower bounds for convolutional LDGM codes. Besides, the waterfall region for convolutional LDGM codes are about one dB away from the system capacity.

Example 2 (Fixed $m$ and $\theta$, Changing $k$ and $\nu$): In this example, we fix $m = 1024$ and $\theta = 0.125$. The BER performance with different values of $k$ and $\nu$ is shown in Fig. 8 and the code rates for the convolutional LDGM codes are $1/4$, $1/2$, $3/4$, $1$, $5/4$ and $3/2$. The corresponding lower bounds are also plotted. From the simulation results, we can see that the BER performance of the convolutional LDGM codes match well with the respective lower bounds in the low BER region for all considered code rates. We can also observe that the convolutional LDGM codes achieve the BER of $10^{-5}$ around one dB away from the system capacity for all
Fig. 6. The BER performance of convolutional LDGM codes with $k = 1024, m = 1024, \nu = 7$ in Example 1. The parameter $\theta$ is specified in the legends. The corresponding lower bounds and system capacities are also plotted.

Fig. 7. The BER performance of convolutional LDGM codes with $k = 1024, m = 512, \nu = 50$ in Example 1. The parameter $\theta$ is specified in the legends. The corresponding lower bounds and system capacities are also plotted.
Fig. 8. The BER performance of convolutional LDGM codes with $m = 1024, \theta = 0.125$ in Example 2. The parameter $k$ and $\nu$ are specified in the legends. The code rates for the convolutional LDGM codes are $1/4$, $1/2$, $3/4$, $1$, $5/4$ and $3/2$. The corresponding lower bounds are also plotted.

considered code rates, as shown in Fig. 9.

Example 3 (Fixed $k$ and $\theta$, Changing $m$ and $\nu$): In this example, we fix $k = 1024$ and $\theta = 0.125$. The BER performance with different values of $m$ and $\nu$ are shown in Fig. 10. The code rates for the convolutional LDGM codes are $1$, $4/5$, $1/2$ and $1/4$. The corresponding lower bounds are also plotted. From the simulation results, we can see that the performance of the convolutional LDGM codes match well with the respective lower bounds in the low BER region for all considered code rates.

Example 4 (Fixed $k$, $m$ and $\theta$, Increasing $\nu$): In the example, we consider the configuration of $k = 1024, m = 1024, \theta = 0.15$ for convolutional LDGM codes. The value of encoding memory $\nu$ is changed. The BER simulation results are shown in Fig. 11, in which the corresponding lower bounds are also plotted. We can observe from the figure that the error floor can be lowered down by increasing the encoding memory $\nu$.

Example 5 (Fixed code rate $k/m$, $\theta$ and $\nu$, Increasing $k$ and $m$): In this example, we fix the code rate $R = k/m$ to 1, the value of $\theta$ to 0.125 and encoding memory $\nu$ to 15. We change the
Fig. 9. The required SNR for the convolutional LDGM codes with $m = 1024$ to achieve the BER of $10^{-5}$ with BPSK signalling over AWGN channels.

Fig. 10. The BER performance of convolutional LDGM codes with $k = 1024, \theta = 0.125$ in Example 3. The parameter $m$ and $\nu$ are specified in the legends. The code rates for the convolutional LDGM codes are $1$, $4/5$, $1/2$ and $1/4$. The corresponding lower bounds are also plotted.
Fig. 11. The BER performance of convolutional LDGM codes with $k = 1024, m = 1024, \theta = 0.15$ in Example 4. The parameter $\nu$ is specified in the legends. The corresponding lower bounds are also plotted.

value of $k$ and $m$ in the simulations. The BER simulation results for the convolutional LDGM codes are shown in Fig. 12. We can see from the figure that the larger value of code length for the convolutional LDGM codes, the better performance we can get in the waterfall region.

D. Performance Comparison

In Fig. 13, we present the BER performance for the convolutional LDGM codes with $\theta = 0.04$, $k = 1024$, $m = 1024$, $\nu = 40$. We also plot the BER performance of AR3J-JSCC codes and AR4JA-JSCC codes in [9], $B_f^{opt, 4}$ codes in [8] and D-Polar in [6] for comparison. The system capacity and the lower bound for convolutional LDGM codes are also plotted. The results show that the convolutional LDGM codes have better waterfall region and error floor performance compared with AR3J-JSCC codes and D-Polar codes. Although no error floor down to BER of $10^{-6}$ has been observed with AR4JA-JSCC codes and $B_f^{opt, 4}$ codes, the convolutional LDGM codes can achieve better performance in the waterfall region, which is about one dB away from the system capacity. Notice that the error floor of convolutional codes can be further lowered down simply by increasing the encoding memory $\nu$, as shown in Fig. 11. At BER = $10^{-5}$ and $10^{-6}$, the convolutional LDGM codes have about one dB and 0.5 dB
Fig. 12. The BER performance of convolutional LDGM codes with code rate 1, $\theta = 0.125$ and $\nu = 15$ in Example 5. The parameter $k$ is specified in the legends.

Fig. 13. The BER performance of convolutional LDGM codes, AR3A-JSCC codes and AR4A-JSCC codes in [9], $B_j^{\text{opt},4}$ codes in [8] and D-Polar in [6] with $\theta = 0.04$. The system capacity for $\theta = 0.04$ and the lower bound for convolutional codes are also plotted.
gain compared to AR4JA-JSCC codes and $B_j^{opt}$-codes. It is worth to noting that the above comparison is not fair, since the convolutional LDG codes need large decoding window and have decoding delay of $d = 2\nu$. However, the convolutional LDG codes are stream-oriented and have a fix coding delay of $d$.

V. CONCLUSIONS

In this paper, we have presented a new framework with linear codes for transmission of Bernoulli sources and proved the coding theorems by deriving partial error exponents. This new framework allows us to unify the proofs of the lossless source coding theorem, the channel coding theorem and the source-channel coding theorem. We derive system capacity for JSCC from the coding theorem and the lower bound for linear codes for performance analysis. A special class of linear codes called convolutional LDG codes are considered in JSCC scheme for simulations. The simulation results show the flexibility of the construction and the predicable performance in error floor region of the convolutional LDG codes. To lower down the error floor, we can simply increase the encoding memory $\nu$, as shown in Example 4. Under iterative sliding window decoding algorithm, the convolutional LDG codes have good waterfall region performance and can achieve about one dB away from the system capacity in the waterfall region for various code rates.

VI. ACKNOWLEDGEMENT

The authors would like to thank Dr. Suihua Cai from Sun Yat-sen University and Prof. Kai Niu from Beijing University of Posts and Telecommunications for their helpful discussions.

REFERENCES

[1] C. E. Shannon, “A mathematical theory of communication,” The Bell Syst. Tech. J., vol. 27, pp. 379–423, 623–656, Jul./Oct. 1948.
[2] M. Fresia, F. Perez-Cruz, and H. V. Poor, “Optimized concatenated LDPC codes for joint source-channel coding,” in Proc. Int. Symp. Inf. Theor. (ISIT), Jun. 2009, pp. 2131–2135.
[3] M. Fresia, F. Peréz-Cruz, H. V. Poor, and S. Verdú, “Joint source and channel coding,” IEEE Signal Process Mag., vol. 27, no. 6, pp. 104–113, Nov. 2010.
[4] J. He, L. Wang, and P. Chen, “A joint source and channel coding scheme base on simple protograph structured codes,” in Proc. Int. Symp. Commun. Information Technol. (ISCIT), Oct. 2012, pp. 65–69.
[5] A. Golmohammadi and D. G. M. Mitchell, “Concatenated spatially coupled LDPC codes with sliding window decoding for joint source-channel coding,” IEEE Trans. Commun., vol. 70, no. 2, pp. 851–864, Feb. 2022.
[6] Y. Dong, K. Niu, J. Dai, S. Wang, and Y. Yuan, “Joint source and channel coding using double polar codes,” *IEEE Commu. Lett.*, vol. 25, no. 9, Sept. 2021.

[7] C. Chen, L. Wang, and F. C. M. Lau, “Joint optimization of protograph LDPC code pair for joint source and channel coding,” *IEEE Trans. Commun.*, vol. 66, no. 8, pp. 3255–3267, Aug. 2018.

[8] S. Liu, L. Wang, J. Chen, and S. Hong, “Joint component design for the JSCC system based on DP-LDPC codes,” *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5808–5818, Sept. 2020.

[9] F. C. M. Lau and J. Zhan, “Joint source-channel codes based on a single protograph,” in *Proc. 11th Int. Symp. Top. Coding (ISTC)*, Aug. 2021, pp. 1–5.

[10] R. Gallager, *Information Theory and Reliable Communication*. New York, NY: John Wiley and Sons, Inc., 1968.

[11] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems (Second edition)*. New York, NY: Cambridge University Press, 2011.

[12] P. Elias, “Coding for noisy channels,” *IRE Conv. Rec.*, vol. 4, pp. 37–46, 1955.

[13] X. Ma, “Coding theorem for systematic low density generator matrix codes,” in *Proc. 9th Int. Symp. Turbo Codes Iterative Inf. Process. (ISTC)*, Sept. 2016, pp. 11–15.

[14] S. Cai, W. Lin, X. Yao, B. Wei, and X. Ma, “Systematic convolutional low density generator matrix code,” *IEEE Trans. Inf. Theory*, vol. 67, no. 6, pp. 3752–3764, Jun. 2021.

[15] T. Zhu and X. Ma, “Near-lossless compression for sparse source using convolutional low density generator matrix codes,” in *Proc. Data Compression Conf. (DCC)*, Mar. 2021, pp. 323–332.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Second edition)*. John Wiley & Sons, Inc., Hoboken, New Jersey: Cambridge University Press, 2006.

[17] T. Zhu and X. Ma, “Block Markov superposition transmission: An approach to unifying source and channel coding,” *To be submitted to IEEE Trans. Inf. Theory*. 

[18] X. Ma, C. Liang, K. Huang, and Q. Zhuang, “Block markov superposition transmission: Construction of big convolutional codes from short codes,” *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3150–3163, Jun. 2015.

[19] X. Ma, Y. Wang, and T. Zhu, “A new framework for proving coding theorems for linear codes,” in *2022 Proc. Int. Symp. Inf. Theor. (ISIT)*, accepted.