Capacity Approaching Run-Length-Limited Codes for Multilevel Recording Systems

Article *in* IEEE Transactions on Magnetics · February 2010

DOI: 10.1109/TMAG.2009.2028340 · Source: IEEE Xplore

4 authors, including:

Hsin-Yi Chen
Taipei Medical University

Hong-Fu Chou
Auckland University of Technology

Shih-Kai Lee
Yuan Ze University

Capacity Approaching Run-Length-Limited Codes for Multilevel Recording Systems

Hsin-Yi Chen[1], Hong-Fu Chou[1], Mao-Chao Lin[1], and Shih-Kai Lee[2]

[1]Graduate Institute of Communication Engineering, National Taiwan University, Taipei 10617, Taiwan, R.O.C.
[2]Department of Communications Engineering, Yuan Ze University, Chungli, Taiwan, R.O.C.

We propose two constructions for multilevel run-length-limited (RLL) block codes for which the rates are very close to the capacity. For each code construction, we propose a variation that has the advantage of low complexity of encoding and decoding. We conducted a simulation to see the combined effect of channel coding and our proposed RLL coding over an optical recording channel.

***Index Terms*—Constrained codes, multilevel run-length-limited codes, optical recording channel.**

I. Introduction

IN recording systems, run-length-limited (RLL) coding [1]–[3] is usually used to avoid the adverse effect of inter-symbol interference (ISI) and to facilitate the operation of synchronization. In particular, binary RLL coding has been extensively investigated [4]–[12]. With the increased demand for higher capacity, M-level recording systems with $M > 2$ have attracted much attention recently.

A multilevel run-length-limited sequence is a sequence with the run length of zeros between two consecutive nonzero symbols being under a given constraint. For example, an (M, d, k) sequence [13]–[17] is an M-level (or M-ary) RLL sequence with symbol alphabet, $\{0, 1, \ldots, M-1\}$, where $2 \le M < \infty$, for which the number of zeros between two consecutive nonzero symbols is at least d and at most k. A multilevel RLL code is a collection of multilevel RLL sequences. An M-level RLL (M, d, k) code has a code rate $R = Q/n$ in bits/symbol, if it contains 2^Q n-symbol sequences satisfying the M-level (d, k) constraint. The capacity of the M-level RLL (M, d, k) code [18], denoted by $C_{(M,d,k)}$, is given by

$$C_{(M,d,k)} = \log_2 \lambda \tag{1}$$

where λ is the largest real root of the characteristic equation

$$Z^{k+2} - Z^{k+1} - (M-1)Z^{k-d+1} + M - 1 = 0. \tag{2}$$

The code efficiency of an M-level RLL (M, d, k) code is expressed as $\eta = R/C_{(M,d,k)}$.

By now, some M-level RLL (M, d, k) codes using finite-state machine encoders have been designed in [19]–[21]. Instead of the sequential design as in [19]–[21], we focus on $(M, 0, k)$ block coding in this paper, where $M \ge 4$. We look for efficient $(M, 0, k, k_l, k_r)$ codes of finite length, where an $(M, 0, k, k_l, k_r)$ sequence is an M-level $(0, k)$-constrained RLL sequence which has at most k_l leading zeros and at most k_r trailing zeros. The condition of $k_l + k_r \le k$ is required such that concatenating any two $(M, 0, k, k_l, k_r)$ sequences can still meet the $(M, 0, k)$ constraint.

In this paper, we propose two code constructions. Using Construction I, we can obtain $(M, 0, k, k_l, k_r)$ RLL codes of length $n = k_l + k_r + 2k$, for which the number of code sequences is $M^{n-1}(M-1)$. The encoding is implemented by first mapping a message P into a sequence $\bar{u} = (u_0, \ldots, u_{n-1})$, where $P = 0, 1, \ldots, M^{n-1}(M-1) - 1$, $u_i \in \{0, 1, \ldots, M-1\}$, $i = 0, 1, \ldots, n-2$ and $u_{n-1} \in \{1, 2, \ldots, M-1\}$, and then $\bar{u}$ is encoded into a code sequence $\bar{v} = (v_0, \ldots, v_{n-1})$, where $v_i \in \{0, 1, \ldots, M-1\}$. For binary $(0, k)$ RLL codes of length n, some classes of rate $(n-1)/n$ code constructions have been reported [4]. In the sense of coding rate, the code obtained from Construction I may be regarded as an analogue of the rate $(n-1)/n$ binary $(0, k)$ RLL codes of length n. In Construction II, the condition of $u_{n-1} \ne 0$ is relaxed so that the code size can be greater than $M^{n-1}(M-1)$, where $n = k_l + k_r + 2k - 2$. Employing either Construction I or Construction II, the constructed code has a code rate close to the capacity given in (1). In case that the message P is in the form of a binary sequence and $\log_2 M$ is an integer, then we propose an extended design of Construction I, denoted as Construction $\mathrm{I}^{\mathrm{ext}}$, for which the complexity of encoding and decoding is low. Similarly, we modify Construction II into Construction $\mathrm{II}^{\mathrm{ext}}$.

The RLL decoding may result in significant error propagation if some of the critical symbols in the RLL sequence are in error. We will assess the likely error propagation of the proposed code construction. Then, we will investigate the joint effect of RLL coding and error-correcting coding (ECC). We consider the application of the proposed code constructions to two recording systems. The first is composed of a Reed-Solomon (RS) encoder, an RLL encoder, a precoder using pulse length modulation (PLM), an optical recording channel corrupted by additive white Gaussian noise (AWGN) and jitter noise, a maximum *a posteriori* (MAP) decoder for the combined PLM precoding and the recording channel, an RLL decoder and an RS decoder. The second is obtained by replacing the RS encoder and the RLL encoder by an RLL encoder followed by a nonbinary LDPC encoder and a uniform interleaver [9], and moreover replacing the RLL decoder and the RS decoder by a uniform deinterleaver, followed by a nonbinary LDPC decoder and an RLL decoder.

The rest of this paper is organized as follows. In Section II, two practical code constructions together with their variations

Manuscript received January 04, 2009; revised March 22, 2009 and June 15, 2009. Current version published December 23, 2009. Corresponding author: M.-C. Lin (e-mail: mclin@cc.ee.ntu.edu.tw).
Digital Object Identifier 10.1109/TMAG.2009.2028340

are presented. In Section III, possible error propagation of the proposed RLL coding is investigated. Moreover, simulation is implemented to see the combined effect of channel coding and the proposed $(M,0,k)$ RLL coding over an optical recording channel. Finally, conclusions are given in Section IV.

II. Practical Code Constructions

In this section, we propose code constructions for $(M,0,k,k_l,k_r)$ RLL codes of length n, for which each code sequence $\bar{v}=(v_0,v_1,\ldots,v_{n-1})$ is chosen based on the sequence $\bar{u}=(u_0,u_1,\ldots,u_{n-1})$, where $v_i, u_i \in \{0,1,\ldots,M-1\}$ for $i=0,1,\ldots,n-1$. A nonzero u_{n-1} is frequently used. In such a case, we denote u_{n-1} by $\underline{u}_{n-1}$.

A. Construction I

Let $M \geq 4$ and $k \geq 2$. We can construct an RLL code V consisting of $(M,0,k,k_l,k_r)$ sequences of length $n=k_l+k_r+2k$, which are divided into four disjoint groups, denoted G_1, G_2, G_3, and G_4 respectively, based on $\bar{u}$.

The details of the four groups are given as follows.

G_1: In case of $(u_{k_l-1},u_{k_l}) \neq (0,0)$ and $(u_{k_l+2k-2},u_{k_l+2k-1}) \neq (0,0)$, we set $v_i=u_i$ for $0 \leq i \leq k_l+k-2$, $v_{k_l+k-1}=\underline{u}_{n-1}$, and $v_i=u_{i-1}$ for $k_l+k \leq i \leq k_l+k_r+2k-1=n-1$.

G_2: In case of $(u_{k_l-1},u_{k_l})=(0,0)$ and $(u_{k_l+2k-2}, u_{k_l+2k-1}) \neq (0,0)$, we set $v_i=u_i$ for $0 \leq i \leq k_l-2$ and $k_l \geq 2$, $v_{k_l-1}=u_{k_l+1}$, $v_{k_l}=\underline{u}_{n-1}$, $v_i=u_{i+1}$ for $k_l+1 \leq i \leq k_l+k-2$, $v_{k_l+k-1}=0$, $v_{k_l+k}=1$, $v_i=u_{i-1}$ for $k_l+k+1 \leq i \leq k_l+k_r+2k-1$.

G_3: In case of $(u_{k_l-1},u_{k_l}) \neq (0,0)$ and $(u_{k_l+2k-2}, u_{k_l+2k-1})=(0,0)$, we set $v_i=u_i$ for $0 \leq i \leq k_l+k-2$, $v_{k_l+k-1}=0$, $v_{k_l+k}=2$, $v_i=u_{i-2}$ for $k_l+k+1 \leq i \leq k_l+2k-1$, $v_{k_l+2k}=\underline{u}_{n-1}$, $v_i=u_{i-1}$ for $k_l+2k+1 \leq i \leq k_l+2k+k_r-1$.

G_4: In case of $(u_{k_l-1},u_{k_l})=(0,0)$ and $(u_{k_l+2k-2}, u_{k_l+2k-1})=(0,0)$, we set $v_i=u_i$ for $0 \leq i \leq k_l-2$ and $k_l \geq 2$, $v_{k_l-1}=u_{k_l+1}$, $v_{k_l}=\underline{u}_{n-1}$, $v_i=u_{i+1}$ for $k_l+1 \leq i \leq k_l+k-2$, $v_{k_l+k-1}=0$, $v_{k_l+k}=3$, $v_i=u_{i-1}$ for $k_l+k+1 \leq i \leq k_l+2k-2$, $v_{k_l+2k-1}=1$, $v_{k_l+2k}=1$, $v_i=u_{i-1}$ for $k_l+2k+1 \leq i \leq k_l+2k+k_r-1$.

Theorem 1: The code V consists of $M^{n-1}(M-1)$ code sequences which satisfy the $(M,0,k,k_l,k_r)$ constraint. The coding rate in bits/symbol is

$$R=\frac{\log_2[M^{n-1}(M-1)]}{n}. \quad (3)$$

Proof: For G_1, we have $v_{k_l+k-1}=\underline{u}_{n-1} \neq 0$. For G_2, we have $v_{k_l+k-1}=0$ and $v_{k_l+k}=1$. For G_3, we have $v_{k_l+k-1}=0$ and $v_{k_l+k}=2$. For G_4, we have $v_{k_l+k-1}=0$ and $v_{k_l+k}=3$. Based on these characteristics, we see that these four groups are disjoint. In the following, we will show that the sequences in each group satisfy the $(M,0,k,k_l,k_r)$ constraint and will compute the number of sequences in each group.

G_1: (i) For $k \geq 2$, we have $v_i=u_i$ for $0 \leq i \leq k_l+k-2 \leq k_l$. Hence, $(u_{k_l-1},u_{k_l})=(v_{k_l-1},v_{k_l}) \neq (0,0)$. In addition, $v_{k_l+k-1}=\underline{u}_{n-1} \neq 0$. Hence, there will be at most k_l zeros from v_0 to v_{k_l} and at most $k-1$ zeros from v_{k_l-1} to v_{k_l+k-1}. (ii) Since $(u_{k_l+2k-2},u_{k_l+2k-1})=(v_{k_l+2k-1},v_{k_l+2k}) \neq (0,0)$, there will be at most k zeros from v_{k_l+k-1} to v_{k_l+2k} and at most k_r zeros from v_{k_l+2k-1} to v_{n-1}. (iii) In G_1, we encode $\bar{u}$ into $\bar{v}$, where $u_{n-1} \neq 0$, $(u_{k_l-1},u_{k_l}) \neq (0,0)$ and $(u_{k_l+2k-2},u_{k_l+2k-1}) \neq (0,0)$. Hence, the number of sequences in G_1 is $M^{n-1}(M-1)-2M^{n-3}(M-1)+M^{n-5}(M-1)$.

G_2: (i) Since $v_{k_l}=\underline{u}_{n-1} \neq 0$ and $v_{k_l+k}=1$, there will be at most k_l zeros from v_0 to v_{k_l} and at most $k-1$ zeros from v_{k_l} to v_{k_l+k}. (ii) Since $(u_{k_l+2k-2},u_{k_l+2k-1})=(v_{k_l+2k-1},v_{k_l+2k}) \neq (0,0)$, there will be at most $k-1$ zeros from v_{k_l+k} to v_{k_l+2k} and at most k_r zeros from v_{k_l+2k-1} to v_{n-1}. (iii) In G_2, we encode $\bar{u}$ into $\bar{v}$, where $u_{n-1} \neq 0$, $(u_{k_l-1},u_{k_l})=(0,0)$ and $(u_{k_l+2k-2},u_{k_l+2k-1}) \neq (0,0)$. Hence, the number of sequences in G_2 is $M^{n-3}(M-1)-M^{n-5}(M-1)$.

G_3: (i) Since $(u_{k_l-1},u_{k_l})=(v_{k_l-1},v_{k_l}) \neq (0,0)$ and $v_{k_l+k}=2$, there will be at most k_l zeros from v_0 to v_{k_l} and at most k zeros from v_{k_l-1} to v_{k_l+k}. (ii) Since $v_{k_l+2k}=\underline{u}_{n-1} \neq 0$, there will be at most $k-1$ zeros from v_{k_l+k} to v_{k_l+2k} and at most k_r-1 zeros from v_{k_l+2k} to v_{n-1}. (iii) In G_3, we encode $\bar{u}$ into $\bar{v}$, where $u_{n-1} \neq 0$, $(u_{k_l-1},u_{k_l}) \neq (0,0)$ and $(u_{k_l+2k-2},u_{k_l+2k-1})=(0,0)$. The number of sequences in G_3 is $M^{n-3}(M-1)-M^{n-5}(M-1)$.

G_4: (i) Since $v_{k_l}=\underline{u}_{n-1}$ and $v_{k_l+k}=3$, there will be at most k_l zeros from v_0 to v_{k_l} and at most $k-1$ zeros from v_{k_l} to v_{k_l+k}. (ii) Since $v_{k_l+2k}=1$, there will be at most $k-1$ zeros from v_{k_l+k} to v_{k_l+2k} and at most k_r-1 zeros from v_{k_l+2k} to v_{n-1}. (iii) In G_4, we encode $\bar{u}$ into $\bar{v}$, where $u_{n-1} \neq 0, (u_{k_l-1},u_{k_l})=(u_{k_l+2k-2},u_{k_l+2k-1})=(0,0)$. The number of sequences in G_4 is $M^{n-5}(M-1)$.

Taking the logarithm of the sum of the numbers of sequences in the four groups and dividing the result by n, we have the rate of V given in (3). □

Example 1: We can construct an RLL code V consisting of $(M,0,k,k_l,k_r)=(4,0,7,4,3)$ sequences of length $n=21$ using Construction I. Let $\bar{u}=(u_0u_1u_2u_3u_4u_5u_6u_7u_8u_9u_{10}u_{11}u_{12}u_{13}u_{14}u_{15}u_{16}u_{17}u_{18}u_{19}\underline{u}_{20})$.

The four groups are given as follows.

G_1: For $(u_3u_4) \neq (0,0)$ and $(u_{16}u_{17}) \neq (0,0)$, $\bar{v}=(u_0u_1u_2u_3u_4\ u_5u_6u_7u_8u_9\underline{u}_{20}u_{10}u_{11}u_{12}\ u_{13}u_{14}u_{15}u_{16}u_{17}u_{18}u_{19})$.

G_2: For $(u_3u_4)=(0,0)$ and $(u_{16}u_{17}) \neq (0,0)$, $\bar{v}=(u_0u_1u_2u_5\underline{u}_{20}\ u_6u_7u_8u_9u_{10}01u_{11}u_{12}u_{13}\ u_{14}u_{15}u_{16}u_{17}u_{18}u_{19})$.

G_3: For $(u_3u_4) \neq (0,0)$ and $(u_{16}u_{17})=(0,0)$, $\bar{v}=(u_0u_1u_2u_3u_4u_5u_6u_7u_8u_902u_{10}u_{11}u_{12}u_{13}u_{14}u_{15}\underline{u}_{20}u_{18}u_{19})$.

G_4: For $(u_3u_4)=(0,0)$ and $(u_{16}u_{17})=(0,0), \bar{v}=(u_0u_1u_2u_5\underline{u}_{20}u_6u_7u_8u_9u_{10}03u_{11}u_{12}u_{13}u_{14}u_{15}11u_{18}u_{19})$.

There are $4^{20}\cdot 3-2\cdot 4^{18}\cdot 3+4^{16}\cdot 3$ distinct sequences in G_1, $4^{18}\cdot 3-4^{16}\cdot 3$ distinct sequences in G_2, $4^{18}\cdot 3-4^{16}\cdot 3$ distinct sequences in G_3, and $4^{16}\cdot 3$ distinct sequences in G_4. The total number of sequences in this RLL code is $4^{20}\cdot 3$. Hence, the code rate $R=log_2(4^{20}\cdot 3)/21 \approx 1.980236$ bits/symbol. The efficiency is $1.980236/1.999983 \approx 99.01264\%$. □

Encoding and Decoding: Let P be a message to be encoded, where $P = 0, 1, \ldots, M^{n-1}(M-1) - 1$. The encoding of V is implemented by the following procedure.

1) Let $P' = P + M^{n-1}$. Then, $P' \in \{M^{n-1}, M^{n-1} + 1, \ldots, M^n - 1\}$.
2) Find the representation of P' in the base of M, i.e.,

$$P' = u_0 + u_1 \cdot M + u_2 \cdot M^2 + \cdots + u_{n-1} \cdot M^{n-1}. \quad (4)$$

Note that u_{n-1} is nonzero and can be denoted by $\underline{u}_{n-1}$.

3) Map $\bar{u}$ into $\bar{v}$ according to the conditions of G_1, G_2, G_3 and G_4 for Construction I.

The decoding of $\bar{v}$ into the message P is implemented by the following procedure.

1) If $v_{k_l+k-1} \neq 0$, then $\bar{v}$ is in G_1. If $v_{k_l+k-1} = 0$ and $v_{k_l+k} = i$, then $\bar{v}$ is in G_{i+1}, $i = 1, 2$, and 3.
2) According to rule of G_i, find the associated $\bar{u}$ from $\bar{v}$.
3) Obtain P' by (4). The message is $P = P' - M^{n-1}$.

B. Construction I^{ext}

For Construction I, the message P is an integer in the range of $\{0, 1, \ldots, M^{n-1}(M-1) - 1\}$. In case that the message is in the form of a binary sequence, then the mapping from the message to $\bar{u}$ can not be implemented. Hence, we need to modify Construction I to Construction $\mathrm{I}^{\mathrm{ext}}$.

Let V be a code of length n designed according to Construction I. Let V^L be the code of length Ln, in which each code sequence is in the form of $(\bar{v}_0, \ldots, \bar{v}_\ell, \ldots, \bar{v}_{L-1})$, where $\bar{v}_\ell \in V$. There are $[M^{n-1}(M-1)]^L$ code sequences in V^L. For encoding from binary message sequences, we consider a subset of V^L, denoted V^{ext}, which consists of 2^Q sequences, where $Q = \lfloor log_2 [M^{n-1}(M-1)]^L \rfloor$. The coding rate will be reduced to

$$R = \frac{\lfloor \log_2 [M^{n-1}(M-1)]^L \rfloor}{Ln}. \quad (5)$$

In general, the decoding/encoding complexity of V^{ext} may be high. However, in case that $\log_2 M = p$ is an integer, the encoding/decoding complexity can be very low. In this case, $Q = L(n-1)p + q$, where $q = \lfloor \log_2 (M-1)^L \rfloor$, and the coding rate of V^{ext} is

$$R = \frac{L(n-1)p + q}{Ln}. \quad (6)$$

The message P^{ext} for encoding and decoding is represented by $L(n-1)p + q$ message bits, i.e.,

$$P^{\mathrm{ext}} = (b_0, b_1, \ldots, b_{L(n-1)p-1}, b_{L(n-1)p}, \ldots, b_{L(n-1)p+q-1}).$$

Encoding and Decoding: We map P^{ext} into

$$\bar{u}^{\mathrm{ext}} = (\bar{u}_0, \bar{u}_1, \ldots, \bar{u}_\ell, \ldots, \bar{u}_{L-1}) \quad (7)$$

where $\bar{u}_\ell = (u_{\ell,0}, u_{\ell,1}, \ldots, u_{\ell,n-2}, \underline{u}_{\ell,n-1})$. Then, $\bar{u}^{\mathrm{ext}}$ is encoded into an $(M, 0, k, k_l, k_r)$ sequence of length Ln represented by

$$\bar{v}^{\mathrm{ext}} = (\bar{v}_0, \bar{v}_1, \ldots, \bar{v}_\ell, \ldots, \bar{v}_{L-1}) \quad (8)$$

TABLE I
THE 6-BIT INPUT/4-SYMBOL OUTPUT LOOKUP TABLE FOR EXAMPLE 2

Binary						Quaternary			
$b_{0,40}$	$b_{0,41}$	$b_{1,40}$	$b_{1,41}$	$b_{2,40}$	$b_{3,40}$	$\underline{u}_{0,20}$	$\underline{u}_{1,20}$	$\underline{u}_{2,20}$	$\underline{u}_{3,20}$
0	0	0	0	0	0	1	1	1	1
0	0	0	0	0	1	1	1	1	2
0	0	0	0	1	0	1	1	1	3
0	0	0	0	1	1	1	1	2	1
0	0	0	1	0	0	1	1	2	2
0	0	0	1	0	1	1	1	2	3
≀						≀			
1	1	1	0	1	0	3	3	2	1
1	1	1	0	1	1	3	3	2	2
1	1	1	1	0	0	3	3	2	3
1	1	1	1	0	1	3	3	3	1
1	1	1	1	1	0	3	3	3	2
1	1	1	1	1	1	3	3	3	3

where $\bar{v}_\ell = (v_{\ell,0}, v_{\ell,1}, \ldots, v_{\ell,n-1})$ is a codeword of V encoded from $\bar{u}_\ell$. The mapping from P^{ext} to $\bar{u}^{\mathrm{ext}}$ is implemented as follows.

1) For $0 \leq \ell \leq L-1$, $(u_{\ell,0}, u_{\ell,1}, \ldots, u_{\ell,n-2}) = (b_{(n-1)\ell p}, b_{(n-1)\ell p+1}, \ldots, b_{(n-1)(\ell+1)p-1})$.
2) We use a lookup table to map the binary q-tuple $(b_{L(n-1)p}, \ldots, b_{L(n-1)p+q-1})$ into a nonbinary L-tuple $(\underline{u}_{0,n-1}, \underline{u}_{1,n-1}, \ldots, \underline{u}_{\ell,n-1}, \ldots, \underline{u}_{(L-1),n-1})$. The size of the lookup table is q by L (q-bit input and L-symbol output).

The decoding of V^{ext} can be implemented by simply reversing the encoding procedure.

Example 2: Let V be the 4-level RLL code designed by Construction I, which consists of $(M, 0, k, k_l, k_r) = (4, 0, 7, 4, 3)$ sequences. Then, $\log_2 M = p = 2$. Let $n = k_l + k_r + 2k = 21$ and $L = 4$. Then, $\lfloor \log_2 (M-1)^L \rfloor = \lfloor \log_2 3^4 \rfloor = q = 6$. We will encode a message P^{ext} represented by $L(n-1)p + q = 166$ message bits into a (4,0,7,4,3) sequence $\bar{v}^{\mathrm{ext}}$ of $Ln = 84$. We need a lookup table with 6-bit input and 4-symbol output as shown in Table I. The code rate of the resulting code V^{ext} is $(40 \cdot 4 + 6)/(21 \cdot 4) \approx 1.976190$ bits/symbol. The efficiency of V^{ext} is $1.976190/1.999983 \approx 98.81033\%$. Note that the code rate of V is approximately 1.980236. □

If we place $M = 2$ in (3), we have $R = (n-1)/n$. Since some binary $(0, k)$ RLL codes of rate $(n-1)/n$ have already been reported in the literature [4], there may be the impression that practical $(M, 0, k, k_l, k_r)$ codes can be constructed in ways which are nothing more than generalizations of binary $(0, k)$ codes. In fact, a larger M allows more room for RLL code construction. In the following, we will show an $(M, 0, k, k_l, k_r)$ RLL code construction for $M \geq 4$, which can achieve better efficiency.

C. Construction II

Let $M \geq 4$, $k_l \leq 4$, $k_r \leq 3$ and $k \geq 3$. We can construct an RLL code V consisting of $(M, 0, k, k_l, k_r)$ sequences of length $n = k_l + k_r + 2k - 2$, which are divided into nine disjoint groups, denoted $G_{A1}, G_{A2}, G_{A3}, G_{B1}, G_{B2}, G_{B3}, G_{C1}, G_{C2}$, and G_{C3} respectively, based on $\bar{u}$. The details of the nine groups are given as follows.

G_{A1}: In case of $u_{n-1} = \underline{u}_{n-1}$, $(u_1, \ldots, u_{k_l}) \neq (0, \ldots, 0)$ and $(u_{n-2-k_r}, \ldots, u_{n-3}) \neq (0, \ldots, 0)$, set $v_i = u_i$ for

$0 \le i \le k_l + k - 3$, $v_{k_l+k-2} = \underline{u}_{n-1}$, and $v_i = u_{i-1}$ for $k_l + k - 1 \le i \le k_l + k_r + 2k - 3 = n - 1$.

G_{A2}: In case of $u_{n-1} = \underline{u}_{n-1}$, $u_1 = \cdots = u_{k_l} = 0$, set $v_0 = u_0$ for $k_l \ge 1$, $v_i = u_{i+k_l}$ for $1 \le i \le k_l - 1$ and $k_l \ge 1$, $v_{k_l} = 1$, $v_i = u_{i+k_l-1}$ for $k_l+1 \le i \le k_l+k-3$, $v_{k_l+k-2} = 0$, $v_{k_l+k-1} = 0$, $v_{k_l+k} = \underline{u}_{n-1}$, $v_i = u_i$ for $k_l+k+1 \le i \le k_l+2k-4$, $v_{k_l+2k-3} = 1$, and $v_i = u_{i-1}$ for $k_l + 2k - 2 \le i \le k_l + k_r + 2k - 3 = n - 1$.

G_{A3}: In case of $u_{n-1} = \underline{u}_{n-1}$, $(u_1, \ldots, u_{k_l}) \ne (0, \ldots, 0)$ and $u_{n-2-k_r} = \cdots = u_{n-3} = 0$, set $v_i = u_i$ for $0 \le i \le k_l - 1$ and $k_l \ge 1$, $v_{k_l} = 1$, $v_i = u_{i-1}$ for $k_l + 1 \le i \le k_l + k - 3$, $v_{k_l+k-2} = 0$, $v_{k_l+k-1} = 1$, $v_i = u_{i-3}$ for $k_l + k \le i \le k_l + 2k - 4$, $v_{k_l+2k-3} = \underline{u}_{n-1}$, $v_i = u_{i-4}$ for $k_l + 2k - 2 \le i \le k_l + 2k - 1$, and $v_i = u_{i-1}$ for $k_l + 2k \le i \le k_l + k_r + 2k - 3 = n - 1$.

G_{B1}: In case of $u_{n-2} = u_{n-1} = 0$, $(u_2, \ldots, u_{k_l}) \ne (0, \ldots, 0)$ and $(u_{n-3-k_r}, \ldots, u_{n-5}) \ne (0, \ldots, 0)$, set $v_i = u_i$ for $0 \le i \le k_l+k-3$, $v_{k_l+k-2} = 0$, $v_{k_l+k-1} = 2$ and $v_i = u_{i-2}$ for $k_l + k \le i \le k_l + k_r + 2k - 3 = n - 1$.

G_{B2}: In case of $u_{n-2} = u_{n-1} = 0$, $u_2 = \cdots = u_{k_l} = 0$, set $v_0 = u_0$ and $v_1 = u_1$ for $k_l \ge 1$, $v_i = u_{i+k_l-1}$ for $2 \le i \le k_l - 1$ and $k_l \ge 1$, $v_{k_l} = 2$, $v_i = u_{i+k_l-2}$ for $k_l + 1 \le i \le k_l + k - 3$, $v_{k_l+k-2} = 0$, $v_{k_l+k-1} = 0$, $v_{k_l+k} = 2$, $v_i = u_{i-1}$ for $k_l + k + 1 \le i \le k_l + 2k - 4$, $v_{k_l+2k-3} = 2$, and $v_i = u_{i-2}$ for $k_l + 2k - 2 \le i \le k_l + k_r + 2k - 3 = n - 1$.

G_{B3}: In case of $u_{n-2} = u_{n-1} = 0$, $(u_2, \ldots, u_{k_l}) \ne (0, \ldots, 0)$ and $u_{n-3-k_r} = \cdots = u_{n-5} = 0$, set $v_i = u_i$ for $0 \le i \le k_l - 1$ and $k_l \ge 1$, $v_{k_l} = 2$, $v_i = u_{i-1}$ for $k_l + 1 \le i \le k_l + k - 3$, $v_{k_l+k-2} = 0$, $v_{k_l+k-1} = 1$, $v_i = u_{i-3}$ for $k_l + k \le i \le k_l + 2k - 4$, $v_{k_l+2k-3} = 2$, $v_i = u_{i-4}$ for $i = k_l + 2k - 2$, and $v_i = u_{i-2}$ for $k_l + 2k - 1 \le i \le k_l + k_r + 2k - 3 = n - 1$.

G_{C1}: In case of $u_{n-2} = 1$, $u_{n-1} = 0$, $(u_2, \ldots, u_{k_l}) \ne (0, \ldots, 0)$ and $(u_{n-3-k_r}, \ldots, u_{n-5}) \ne (0, \ldots, 0)$, set $v_i = u_i$ for $0 \le i \le k_l+k-3$, $v_{k_l+k-2} = 0$, $v_{k_l+k-1} = 3$ and $v_i = u_{i-2}$ for $k_l + k \le i \le k_l + k_r + 2k - 3 = n - 1$.

G_{C2}: In case of $u_{n-2} = 1$, $u_{n-1} = 0$, $u_2 = \cdots = u_{k_l} = 0$, set $v_0 = u_0$ and $v_1 = u_1$ for $k_l \ge 1$, $v_i = u_{i+k_l-1}$ for $2 \le i \le k_l - 1$ and $k_l \ge 1$, $v_{k_l} = 3$, $v_i = u_{i+k_l-2}$ for $k_l + 1 \le i \le k_l + k - 3$, $v_{k_l+k-2} = 0$, $v_{k_l+k-1} = 0$, $v_{k_l+k} = 3$, $v_i = u_{i-1}$ for $k_l + k + 1 \le i \le k_l + 2k - 4$, $v_{k_l+2k-3} = 3$, and $v_i = u_{i-2}$ for $k_l + 2k - 2 \le i \le k_l + k_r + 2k - 3 = n - 1$.

G_{C3}: In case of $u_{n-2} = 1$, $u_{n-1} = 0$, $(u_2, \ldots, u_{k_l}) \ne (0, \ldots, 0)$ and $u_{n-3-k_r} = \cdots = u_{n-5} = 0$, set $v_i = u_i$ for $0 \le i \le k_l - 1$ and $k_l \ge 1$, $v_{k_l} = 3$, $v_i = u_{i-1}$ for $k_l + 1 \le i \le k_l + k - 3$, $v_{k_l+k-2} = 0$, $v_{k_l+k-1} = 1$, $v_i = u_{i-3}$ for $k_l + k \le i \le k_l + 2k - 4$, $v_{k_l+2k-3} = 3$, $v_i = u_{i-4}$ for $i = k_l + 2k - 2$, and $v_i = u_{i-2}$ for $k_l + 2k - 1 \le i \le k_l + k_r + 2k - 3 = n - 1$.

Theorem 2: The code V in Construction II consists of $M^{n-2}[M(M-1)+2]$ code sequences which satisfy the $(M, 0, k, k_l, k_r)$ constraint. The coding rate in bits/symbol is

$$R = \frac{\log_2\{M^{n-2}[M(M-1)+2]\}}{n}. \tag{9}$$

Proof: For G_{A1}, we have $v_{k_l+k-2} = \underline{u}_{n-1} \ne 0$. For G_{A2}, we have $v_{k_l} = 1$, $v_{k_l+k-2} = 0$, and $v_{k_l+k-1} = 0$. For G_{A3}, we have $v_{k_l} = 1$, $v_{k_l+k-2} = 0$, and $v_{k_l+k-1} = 1$. For G_{B1}, we have $v_{k_l+k-2} = 0$ and $v_{k_l+k-1} = 2$. For G_{B2}, we have $v_{k_l} = 2$, $v_{k_l+k-2} = 0$, and $v_{k_l+k-1} = 0$. For G_{B3}, we have $v_{k_l} = 2$, $v_{k_l+k-2} = 0$, and $v_{k_l+k-1} = 1$. For G_{C1}, we have $v_{k_l+k-2} = 0$ and $v_{k_l+k-1} = 3$. For G_{C2}, we have $v_{k_l} = 3$, $v_{k_l+k-2} = 0$ and $v_{k_l+k-1} = 0$. For G_{C3}, we have $v_{k_l} = 3$, $v_{k_l+k-2} = 0$ and $v_{k_l+k-1} = 1$. Based on these characteristics, we see that these nine groups are disjoint. In the following, we will show that the sequences in each group satisfy the $(M, 0, k, k_l, k_r)$ constraint and will compute the number of sequences in each group.

G_{A1}: (i) For $k \ge 3$, we have $v_i = u_i$ for $0 \le i \le k_l + k - 3 \le k_l$. Hence, we have $(v_1, \ldots, v_{k_l}) = (u_1, \ldots, u_{k_l}) \ne (0, \ldots, 0)$. Hence, there are at most k_l zeros from v_0 to v_{k_l}. Since $v_{k_l+k-2} = \underline{u}_{n-1}$, there are at most k zeros from v_1 to v_{k_l+k-2} for $k_l \le 4$. (ii) For $k \ge 2$, we have $k_l + k - 2 \le n - 2 - k_r$ and hence $(v_{n-1-k_r}, \ldots, v_{n-2}) = (u_{n-2-k_r}, \ldots, u_{n-3}) \ne (0, \ldots, 0)$. There will be at most k zeros from v_{k_l+k-2} to v_{n-2} (for $k_r \le 3$) and at most k_r zeros from v_{n-1-k_r} to v_{n-1}. (iii) In G_{A1}, we encode $\bar{u}$ into $\bar{v}$, where $u_{n-1} \ne 0$, $(u_1, \ldots, u_{k_l}) \ne (0, \ldots, 0)$ and $(u_{n-2-k_r}, \ldots, u_{n-3}) \ne (0, \ldots, 0)$. Hence, the number of sequences in G_{A1} is $M^{n-1}(M-1)$ - $M^{n-1-k_l}(M-1)$ - $M^{n-1-k_r}(M-1) + M^{n-1-k_l-k_r}(M-1)$.

G_{A2}: (i) Since $v_{k_l} = 1$ and $v_{k_l+k} = \underline{u}_{n-1}$, there will be at most k_l zeros from v_0 to v_{k_l} and at most $k - 1$ zeros from v_{k_l} to v_{k_l+k}. (ii) Since $v_{k_l+2k-3} = 1$, there will be at most $k - 4$ zeros from v_{k_l+k} to v_{k_l+2k-3} and at most k_r zeros from v_{k_l+2k-3} to v_{n-1}. (iii) In G_{A2}, we encode $\bar{u}$ into $\bar{v}$, where $u_{n-1} \ne 0$, $(u_1, \ldots, u_{k_l}) = (0, \ldots, 0)$. Hence, the number of sequences in G_{A2} is $M^{n-1-k_l}(M-1)$.

G_{A3}: (i) Since $v_{k_l} = 1$ and $v_{k_l+k-1} = 1$, there will be at most k_l zeros from v_0 to v_{k_l} and at most $k - 2$ zeros from v_{k_l} to v_{k_l+k-1}. (ii) Since $v_{k_l+2k-3} = \underline{u}_{n-1}$, there will be at most $k - 3$ zeros from v_{k_l+k-1} to v_{k_l+2k-3} and at most k_r zeros from v_{k_l+2k-3} to v_{n-1}. (iii) In G_{A3}, we encode $\bar{u}$ into $\bar{v}$, where $u_{n-1} \ne 0$, $(u_1, \ldots, u_{k_l}) \ne (0, \ldots, 0)$ and $(u_{n-2-k_r}, \ldots, u_{n-3}) = (0, \ldots, 0)$. Hence, the number of sequences in G_{A3} is $M^{n-1-k_r}(M-1) - M^{n-1-k_l-k_r}(M-1)$.

G_{B1}: (i) Since $(v_2, \ldots, v_{k_l}) = (u_2, \ldots, u_{k_l}) \ne (0, \ldots, 0)$ and $v_{k_l+k-1} = 2$, there will be at most k_l zeros from v_0 to v_{k_l} and at most k zeros from v_2 to v_{k_l+k-1} for $k_l \le 4$. (ii) Since $(v_{n-1-k_r}, \ldots, v_{n-3}) = (u_{n-3-k_r}, \ldots, u_{n-5}) \ne (0, \ldots, 0)$, there will be at most k zeros from v_{k_l+k-1} to v_{n-3} (for $k_r \le 5$) and no more than k_r zeros from v_{n-1-k_r} to v_{n-1}. (iii) In G_{B1}, we encode $\bar{u}$ into $\bar{v}$, where $u_{n-2} = u_{n-1} = 0$, $(u_2, \ldots, u_{k_l}) \ne (0, \ldots, 0)$ and $(u_{n-3-k_r}, \ldots, u_{n-5}) \ne (0, \ldots, 0)$. Hence, the number of sequences in G_{B1} is $M^{n-2} - M^{n-1-k_l} - M^{n-1-k_r} + M^{n-k_l-k_r}$.

G_{B2}: (i) Since $v_{k_l} = 2$ and $v_{k_l+k} = 2$, there will be at most k_l zeros from v_0 to v_{k_l} and at most $k - 1$ zeros from v_{k_l} to v_{k_l+k}. (ii) Since $v_{k_l+2k-3} = 2$, there will be at most $k - 4$ zeros from v_{k_l+k} to v_{k_l+2k-3} and at most k_r zeros from v_{k_l+2k-3} to v_{n-1}. (iii) In G_{B2}, we encode $\bar{u}$ into $\bar{v}$, where $u_{n-2} = u_{n-1} = 0$, $(u_2,\ldots,u_{k_l}) = (0,\ldots,0)$. Hence, the number of sequences in G_{B2} is M^{n-1-k_l}.

G_{B3}: (i) Since $v_{k_l} = 2$ and $v_{k_l+k-1} = 1$, there will be at most k_l zeros from v_0 to v_{k_l} and at most $k-2$ zeros from v_{k_l} to v_{k_l+k-1}. (ii) Since $v_{k_l+2k-3} = 2$, there will be at most $k-3$ zeros from v_{k_l+k-1} to v_{k_l+2k-3} and at most k_r zeros from v_{k_l+2k-3} to v_{n-1}. (iii) In G_{B3}, we encode $\bar{u}$ into $\bar{v}$, where $u_{n-2} = u_{n-1} = 0$, $(u_2,\ldots,u_{k_l}) \neq (0,\ldots,0)$ and $(u_{n-3-k_r},\ldots,u_{n-5}) = (0,\ldots,0)$. Hence, the number of sequences in G_{B2} is $M^{n-1-k_r} - M^{n-k_l-k_r}$.

G_{C1}: The proof for this group is similar to that for group G_{B1}. The number of sequences in G_{C1} is $M^{n-2} - M^{n-1-k_l} - M^{n-1-k_r} + M^{n-k_l-k_r}$.

G_{C2}: The proof for this group is similar to that for group G_{B2}. The number of sequences in G_{C2} is M^{n-1-k_l}.

G_{C3}: The proof for this group is similar to that for group G_{B3}. The number of sequences in G_{C3} is $M^{n-1-k_r} - M^{n-k_l-k_r}$.

Summing up the numbers of sequences in G_{A1}, G_{A2} and G_{A3} respectively, we see that there are $M^{n-1}(M-1)$ sequences in $G_{A1} \cup G_{A2} \cup G_{A3}$. Similarly, there are M^{n-2} sequences in $G_{B1} \cup G_{B2} \cup G_{B3}$ and there are M^{n-2} sequences in $G_{C1} \cup G_{C2} \cup G_{C3}$. Hence, there are a total of $M^{n-2}[M(M-1)+2]$ in V. Taking the logarithm of this number and dividing the result by n, we have the rate given in (9). □

For $G_{A1} \cup G_{A2} \cup G_{A3}$, we encode $\bar{u}$ into $\bar{v}$, where $u_{n-1} \neq 0$. For $G_{B1} \cup G_{B2} \cup G_{B3}$, we encode $\bar{u}$ into $\bar{v}$, where $u_{n-2} = u_{n-1} = 0$. For $G_{C1} \cup G_{C2} \cup G_{C3}$, we encode $\bar{u}$ into $\bar{v}$, where $u_{n-2} = 1$, $u_{n-1} = 0$. For Construction II, we can employ $\bar{u}$ with either $u_{n-1} \neq 0$ or $u_{n-1} = 0$. In contrast, for Construction I, $u_{n-1} \neq 0$ is required. This is the key point that makes the coding rate of Construction II to be better than that of Construction I.

Encoding and Decoding: The encoding of V is designed to map a message P into one of the $M^{n-2}[M(M-1)+2]$ codewords in V, where $P = 0, 1, \ldots, M^{n-2}[M(M-1)+2]-1$. The encoding is implemented by the following procedure.

1) Suppose that $0 \leq P \leq M^{n-2}-1$. (i) Find the representation of P in the base of M as in (4). (ii) Map $\bar{u}$ into $\bar{v}$ according to the rule of $G_{B1} \cup G_{B2} \cup G_{B3}$.
2) Suppose that $M^{n-2} \leq P \leq 2M^{n-2}-1$. (i) Find the representation of P in the base of M as in (4). (ii) Map $\bar{u}$ into $\bar{v}$ according to the rule of $G_{C1} \cup G_{C2} \cup G_{C3}$.
3) Suppose that $2M^{n-2} \leq P \leq M^{n-2}[M(M-1)+2]-1$. (i) Let $P' = P + M^{n-1} - 2M^{n-2}$, where $P' \in \{M^{n-1},\ldots,M^n-1\}$. (ii) Find the representation of P' in the base of M as in (4). (iii) Map $\bar{u}$ into $\bar{v}$ according to the rule of $G_{A1} \cup G_{A2} \cup G_{A3}$.

The decoding of $\bar{v}$ into the message P is implemented by the following procedure.

1) According to v_{k_l}, v_{k_l+k-2}, and v_{k_l+k-1}, classify whether $\bar{v}$ is in $G_{A1} \cup G_{A2} \cup G_{A3}$ or in $G_{B1} \cup G_{B2} \cup G_{B3}$ or in $G_{C1} \cup G_{C2} \cup G_{C3}$.
2) Now that we find the union of groups $G_{i1} \cup G_{i2} \cup G_{i3}$, $i = A, B, C$ for $\bar{v}$. According to rule of $G_{i1} \cup G_{i2} \cup G_{i3}$, we can find the associated $\bar{u}$ and the associated P.

Example 3: We can design a 4-level RLL code V consisting of $(M,0,k,k_l,k_r) = (4,0,7,4,3)$ of length $n = 19$ using in Construction II. Let $\bar{u} = (u_0u_1u_2u_3u_4u_5 u_6u_7u_8u_9u_{10}u_{11}u_{12}u_{13}u_{14}u_{15}u_{16}u_{17}u_{18})$. The nine groups are as follows.

G_{A1}: For $(u_1u_2u_3u_4) \neq (0,0,0,0)$ and $(u_{14}u_{15}u_{16}) \neq (0,0,0)$, $u_{18} = \underline{u}_{18} \neq 0$, set $\bar{v} = (u_0u_1u_2u_3u_4u_5u_6 u_7u_8\underline{u}_{18}u_9u_{10}u_{11}u_{12}u_{13}u_{14}u_{15}u_{16}u_{17})$.

G_{A2}: For $(u_1u_2u_3u_4) = (0,0,0,0)$, $u_{18} = \underline{u}_{18} \neq 0$, set $\bar{v} = (u_0u_5u_6u_71u_8u_9u_{10}u_{11}00\underline{u}_{18}u_{12}u_{13}u_{14}1u_{15}u_{16}u_{17})$.

G_{A3}: For $(u_1u_2u_3u_4) \neq (0,0,0,0)$ and $(u_{14}u_{15}u_{16}) = (0,0,0)$, $u_{18} = \underline{u}_{18} \neq 0$, set $\bar{v} = (u_0u_1u_2u_31u_4u_5u_6 u_701u_8u_9u_{10}u_{11}\underline{u}_{18}u_{12}u_{13}u_{17})$.

G_{B1}: For $(u_2u_3u_4) \neq (0,0,0)$ and $(u_{13}u_{14}) \neq (0,0)$, $u_{17} = u_{18} = 0$, set $\bar{v} = (u_0u_1u_2u_3u_4u_5u_6u_7u_802u_9u_{10}u_{11} u_{12}u_{13}u_{14}u_{15}u_{16})$.

G_{B2}: For $(u_2u_3u_4) = (0,0,0)$, $u_{17} = u_{18} = 0$, set $\bar{v} = (u_0u_1 u_5u_62u_7u_8u_9u_{10}002u_{11}u_{12}u_{13}2u_{14}u_{15}u_{16})$.

G_{B3}: For $(u_2u_3u_4) \neq (0,0,0)$ and $(u_{13}u_{14}) = (0,0)$, $u_{17} = u_{18} = 0$, set $\bar{v} = (u_0u_1u_2u_32u_4u_5u_6u_701u_8u_9u_{10} u_{11}2u_{12}u_{15}u_{16})$.

G_{C1}: For $(u_2u_3u_4) \neq (0,0,0)$ and $u_{13}u_{14} \neq (0,0)$, $u_{17} = 1$, $u_{18} = 0$, set $\bar{v} = (u_0u_1u_2u_3u_4u_5u_6u_7u_803u_9 u_{10}u_{11}u_{12}u_{13}u_{14}u_{15}u_{16})$.

G_{C2}: For $(u_2u_3u_4) = (0,0,0)$, $u_{17} = 1$, $u_{18} = 0$, set $\bar{v} = (u_0u_1u_5u_63u_7u_8u_9u_{10}003u_{11}u_{12}u_{13}3u_{14}u_{15}u_{16})$.

G_{C3}: For $(u_2u_3u_4) \neq (0,0,0)$ and $(u_{13}u_{14}) = (0,0)$, $u_{17} = 1$, $u_{18} = 0$, set $\bar{v} = (u_0u_1u_2u_33u_4u_5u_6u_701u_8u_9 u_{10}u_{11}3u_{12}u_{15}u_{16})$.

There are $4^{18} \cdot 3$ sequences in $G_{A1} \cup G_{A2} \cup G_{A3}$; 4^{17} sequences in $G_{B1} \cup G_{B2} \cup G_{B3}$; 4^{17} sequences in $G_{C1} \cup G_{C2} \cup G_{C3}$. In total, there are $4^{17} \cdot 14$ sequences of length 19. The rate is $R = \log_2(4^{17} \cdot 14)/19 \approx 1.989861$. The efficiency is $1.989861/1.999983 \approx 99.49389\%$. □

D. Construction II^{ext}

In case that the message is in the form of a binary sequence, we need to modify Construction II to Construction II$^{\text{ext}}$.

Let V be a code of length n designed according to Construction II. Let V^L be the code of length Ln, in which each code sequence is in the form of $(\bar{v}_0,\ldots,\bar{v}_\ell,\ldots,\bar{v}_{L-1})$, where $\bar{v}_\ell \in V$. There are $\{M^{n-2}[M(M-1)+2]\}^L$ code sequences in V^L. For encoding from binary message sequences, we need to pick a subset of V^L, denoted V^{ext}, which consists of 2^Q sequences, where $Q = \lfloor \log_2\{M^{n-2}[M(M-1)+2]\}^L \rfloor$. The coding rate will be reduced to

$$R = \frac{\lfloor \log_2\{M^{n-2}[M(M-1)+2]\}^L \rfloor}{Ln}. \tag{10}$$

In case that $\log_2 M = p$ is an integer, the encoding/decoding complexity can be very low. In this case, $Q = L(n-2)p + q$ where $q = \lfloor \log_2[M(M-1)+2]^L \rfloor$, and the coding rate is

$$R = \frac{L(n-2)p + q}{Ln}. \tag{11}$$

The message P^{ext} for encoding and decoding is represented by $L(n-2)p + q$ message bits, i.e.,

$$P^{\text{ext}} = (b_0, b_1, \ldots, b_{L(n-2)p-1},\ b_{L(n-2)p}, \ldots, b_{L(n-2)p+q-1}).$$

Encoding and Decoding: We need to map P^{ext} into $\bar{u}^{\text{ext}} = (\bar{u}_0, \bar{u}_1, \ldots, \bar{u}_\ell, \ldots, \bar{u}_{L-1})$, which is then encoded into a codeword $\bar{v}^{\text{ext}} = (\bar{v}_0, \bar{v}_1, \ldots, \bar{v}_\ell, \ldots, \bar{v}_{L-1})$, where $\bar{u}_\ell = (u_{\ell,0}, u_{\ell,1}, \ldots, u_{\ell,n-2}, u_{\ell,n-1})$ and $\bar{v}_\ell = (v_{\ell,0}, v_{\ell,1}, \ldots, v_{\ell,n-2}, v_{\ell,n-1})$. The details are as follows.

For $0 \le \ell \le L-1, (u_{\ell,0}, u_{\ell,1}, \ldots, u_{\ell,n-3}) = (b_{(n-2)\ell p}, b_{(n-2)\ell p+1}, \ldots, b_{(n-2)(\ell+1)p-1})$.

We need an efficient mapping from the binary q-tuple, $(b_{L(n-2)p}, \ldots, b_{L(n-2)p+q-1})$ to the nonbinary $2L$-tuple $(u_{0,n-2}, u_{0,n-1}, u_{1,n-2}, u_{1,n-1}, \ldots, u_{\ell,n-2}, u_{\ell,n-1}, \ldots, u_{(L-1),n-2}, u_{(L-1),n-1})$. Let IBR be the integer with the binary representation $(b_{L(n-2)p}, \ldots, b_{L(n-2)p+q-1})$. That means

$$\text{IBR} = \sum_{j=0}^{q-1} b_{L(n-2)p+j} 2^j. \tag{12}$$

Then, we find the representation of IBR in the base of $M(M-1)+2$. Thus

$$\text{IBR} = \sum_{\ell=0}^{L-1} W_\ell [M(M-1)+2]^\ell \tag{13}$$

where $0 \le W_\ell \le M(M-1)+1$. For each W_ℓ, three cases will be considered.

a) Suppose that $0 \le W_\ell < M(M-1)$. Let $W'_\ell = W_\ell + M$. Then, $M \le W'_\ell < M^2$. Write W'_ℓ in the form of base M, i.e., $W'_\ell = \underline{u}_{\ell,n-1} M + u_{\ell,n-2}$, where $\underline{u}_{\ell,n-1} \neq 0$. Then, we have $(u_{\ell,n-2}, \underline{u}_{\ell,n-1})$. According to the rule in $G_{A1} \cup G_{A2} \cup G_{A3}$, we encode $\bar{u}_\ell = (u_{\ell,0}, u_{\ell,1}, \ldots, u_{\ell,n-2}, \underline{u}_{\ell,n-1})$ into $\bar{v}_\ell$.

b) Suppose that $W_\ell = M(M-1)$. Set $u_{\ell,n-2} = u_{\ell,n-1} = 0$. Then, we encode $\bar{u}_\ell$ into $\bar{v}_\ell$ according to the rule of $G_{B1} \cup G_{B2} \cup G_{B3}$.

c) Suppose that $W_\ell = M(M-1)+1$. Set $u_{\ell,n-2} = 1, u_{\ell,n-1} = 0$. Then, we encode $\bar{u}_\ell$ into $\bar{v}_\ell$ according to the rule of $G_{C1} \cup G_{C2} \cup G_{C3}$.

Then, we have $\bar{v}^{\text{ext}} = (\bar{v}_0, \bar{v}_1, \ldots, \bar{v}_\ell, \ldots, \bar{v}_{L-1})$ as the RLL sequence encoded from $(b_0, b_1, \ldots, b_{L(n-2)p+q-1})$.

The decoding is a simple reverse operation of the encoding. Let $\bar{v}^{\text{ext}}$ be the RLL sequence for decoding. By checking v_{k_l}, v_{k_l+k-2}, and v_{k_l+k-1}, we can tell the group (one of the nine groups, i.e., $G_{A1}, G_{A2}, G_{A3}, G_{B1}, G_{B2}, G_{B3}, G_{C1}, G_{C2}, G_{C3}$) to which $\bar{v}_\ell$ belongs.

i) If $\bar{v}_\ell$ is in $G_{C1} \cup G_{C2} \cup G_{C3}$, then $W_\ell = M(M-1)+1$.

ii) If $\bar{v}_\ell$ is in $G_{B1} \cup G_{B2} \cup G_{B3}$, then $W_\ell = M(M-1)$.

iii) If $\bar{v}_\ell$ is in $G_{A1} \cup G_{A2} \cup G_{A3}$, then we can retrieve $(u_{\ell,n-2}, \underline{u}_{\ell,n-1})$ from $\bar{v}_\ell$. We then have $W'_\ell = \underline{u}_{\ell,n-1} M + u_{\ell,n-2}$ and $W_\ell = W'_\ell - M$.

With $W_\ell, 0 \le \ell \le L-1$, we can compute IBR according to (13). Then, we can recover $(b_{L(n-2)p}, \ldots, b_{L(n-2)p+q-1})$, the binary representation of IBR. With the knowledge of the group to which $\bar{v}_\ell$ belongs, we can also recover $(u_{\ell,0}, u_{\ell,1}, \ldots, u_{\ell,n-3})$ for $0 \le \ell \le L-1$. Then, we can find $(b_0, b_1, \ldots, b_{L(n-2)p-1})$.

TABLE II
$(M, d, k) = (4, 0, k)$ RLL CODES OBTAINED BY CONSTRUCTION II

(M,d,k)	*Length*	*Rate* : R	*Capacity* : C	$\eta = R/C$
(4,0,3)	7	1.9724	1.9957164944207	98.832%
(4,0,4)	10	1.9807	1.9989390561406	99.089%
(4,0,5)	13	1.9851	1.9997355196971	99.269%
(4,0,6)	16	1.9879	1.9999339363403	99.399%
(4,0,7)	19	1.9898	1.9999834880591	99.494%
(4,0,8)	22	1.9912	1.9999958723044	99.564%
(4,0,9)	25	1.9922	1.9999989680965	99.614%
(4,0,10)	28	1.9931	1.9999997420255	99.659%

TABLE III
$(M, d, k) = (4, 0, k)$ RLL CODES OBTAINED BY CONSTRUCTION II^{ext}

(M,d,k)	*Length*	*Rate* : R	*Capacity* : C	$\eta = R/C$
(4,0,3)	7	1.9642	1.9957164944207	98.421%
(4,0,4)	10	1.9750	1.9989390561406	98.804%
(4,0,5)	13	1.9807	1.9997355196971	99.049%
(4,0,6)	16	1.9843	1.9999339363403	99.219%
(4,0,7)	19	1.9868	1.9999834880591	99.344%
(4,0,8)	22	1.9886	1.9999958723044	99.434%
(4,0,9)	25	1.9900	1.9999989680965	99.504%
(4,0,10)	28	1.9910	1.9999997420255	99.554%

Note that no lookup table is needed in the encoding and decoding.

Example 4: Suppose that we employ Example 3 and use $L = 4$ blocks for encoding. Remember that $n = 19$. In addition, $\log_2 M = p = 2$, and $\lfloor \log_2 \{M(M-1)+2\}^L \rfloor = q = 15$. For the integer IBR, we have $0 \le \text{IBR} \le 2^{15}-1$. According to (13), the integer $\text{IBR} = W_3 \cdot 14^3 + W_2 \cdot 14^2 + W_1 \cdot 14 + W_0$, where $0 \le W_\ell \le 13$, for $\ell = 0, 1, 2$ and $0 \le W_3 \le 11$. The code rate is $(34 \cdot 4 + 15)/(19 \cdot 4) \approx 1.986842$ (bits/symbol), the efficiency is $1.986842/1.999983 \approx 99.34294\%$. We see that this practical realization of (4, 0, 7, 4, 3) code with low encoding/decoding complexity has coding rate approaching the capacity and the efficiency is very close to 1. □

The coding rates of some $(M, 0, k, k_l, k_r)$ codes obtained from Construction II and Construction II^{ext} respectively, with $M = 4$ and $n = 2k + k_l + k_r - 2$ for various k are listed in Tables II and III respectively. We can see that the proposed $(M, 0, k)$ code designs are efficient for a wide range of k. Note that for binary $(0, k)$ code, its capacity can not exceed 1 bit/symbol. Hence, advantage of the proposed $(M, 0, k)$ block code over the binary $(0, k)$ block code is obvious.

III. System Performance Over an Optical Recording Channel

The dependency of symbols within the RLL sequence will incur the phenomenon of error propagation for the decoded message bits. Sometimes, only one symbol of the RLL sequence in error may result in many erroneous message bits after decoding. For the RLL code of length Ln designed by Construction II^{ext}, errors in positions of $v_{\ell,k_l}, v_{\ell,k_l+k-2}$, and v_{ℓ,k_l+k-1} may result in erroneous bits in $(u_{\ell,1}, u_{\ell,2}, \ldots, u_{\ell,n-3})$ and in $(b_{L(n-2)p}, \ldots, b_{L(n-2)p+q-1})$ respectively. The erroneous bits in $(u_{\ell,1}, u_{\ell,2}, \ldots, u_{\ell,n-3})$ may form a burst, denoted as αBURST. The worst case of αBURST has a length up to $(\log_2 M)(n-4)$ bits, which occurs when a sequence in G_{A1} is mistaken for a sequence in G_{B2}. The erroneous bits in $(b_{L(n-2)p}, \ldots, b_{L(n-2)p+q-1})$ may form an error burst of length up to q, denoted as βBURST. However, errors for other positions in $\bar{v}_\ell$ will not cause serious error propagation. For the (M, d, k) codes of length n listed in Table III, we see that increasing k will result in an increased n and hence an increased αBURST.

We need to use an efficient error-correcting code (ECC) to combat the combination of random errors and bursts. In this section, we consider the joint effect of ECC and the proposed M-level RLL code over a precoded optical recording channel. We consider two systems. The first is the classical ECC-RLL system, for which the ECC encoding is followed by the RLL coding. The second is the RLL-ECC system [22], [4], [8], [9], for which the RLL encoding is followed by the ECC encoding. For each system, the RLL code obtained from Construction II$^{\text{ext}}$ with $L = 4, n = 2k + k_l + k_r - 2$ is used. We use an M-level PLM precoder, which uses $\{v_i\}$ as input and $\{z_i\}$ as output, where i is the time index, the relation between $\{v_i\}$ and $\{z_i\}$ is given by

$$z_i = v_i + z_{i-1} \pmod{M}. \tag{14}$$

In each system, there is a signal mapper which maps the M-level (M-ary) symbol z_i into an M-level pulse amplitude modulation (PAM) signal $\tilde{z}_i$. Assuming that M is an even integer, we have $\tilde{z}_i \in \{-(M-1)A, -(M-3)A, \ldots, (M-3)A, (M-1)A\}$, where A is a constant.

We consider the optical recording channel as presented in [23], for which the impulse response is

$$f(t) = \frac{2}{ST\sqrt{\pi}} \exp\left\{-\left(\frac{2t}{ST}\right)^2\right\} \tag{15}$$

and the transition response $g(t)$ is the convolution of $\prod(t)$ and $f(t)$, where $\prod(t)$ equals 1 for $0 \le t \le T$ and equals 0 otherwise. The output signal of the optical recording channel is

$$y(t) = \sum_{i=-\infty}^{\infty} \tilde{z}_i g(t - (i + \Delta_i)T) + n(t) \tag{16}$$

where T is the symbol duration, $\{\Delta_i\}$ is an independently identically distributed random process representing the jitter effect, $n(t)$ is the additive white Gaussian (AWGN) noise with one-sided power spectral density N_0. In the simulation, we assume that each Δ_i is a zero mean Gaussian random variable with variance σ_Δ^2. In addition, we assume that $|\sigma_\Delta|T$ is small and $g(t)$ is sufficiently bandlimited [24] so that

$$g(t - (i + \Delta_i)T) \approx g(t - iT) - \Delta_i T f(t - iT). \tag{17}$$

Hence, we have

$$y(t) \approx \sum_{i=-\infty}^{\infty} \tilde{z}_i g(t - iT) - \sum_{i=-\infty}^{\infty} \tilde{z}_i \Delta_i T f(t - iT) + n(t) \tag{18}$$

where $\sum_{i=-\infty}^{\infty} \tilde{z}_i \Delta_i T f(t - iT)$ is the jitter noise and $n(t)$ is the AWGN noise. The average energy per code symbol at the output of the optical channel is

$$\begin{aligned} Es &= E\left\{\tilde{z}_i^2\right\} E\left\{\sum_{j=-\infty}^{\infty} g(jT)^2\right\} \\ &= 2\frac{\sum_{l=0}^{\frac{M}{2}-1}[M-(2l+1)]^2}{M} \\ &\quad \times A^2 E\left\{\sum_{j=-\infty}^{\infty} g(jT)^2\right\} \end{aligned} \tag{19}$$

where $E\{X\}$ denotes the expected value of X. The average energy per message bit is

$$E_b = \frac{E_s}{R_{\text{rll}} R_{\text{ecc}}} \tag{20}$$

where R_{rll} and R_{ecc} are the rates of the RLL code and the ECC code respectively. The average energy of jitter noise for each received code symbol is

$$\begin{aligned} M_0 &= E\left\{\tilde{z}_i^2\right\} E\left\{\Delta_i^2\right\} E\left\{\sum_{j=-\infty}^{\infty} [Tf(jT)]^2\right\} \\ &= 2\frac{\sum_{l=0}^{\frac{M}{2}-1}[M-(2l+1)]^2}{M} \\ &\quad \times A^2 \sigma_\Delta^2 E\left\{\sum_{j=-\infty}^{\infty} [Tf(jT)]^2\right\}. \end{aligned} \tag{21}$$

We use $S = 3.8$ [25]. Then, $Tf(-2T) = Tf(2T) = 0.0981, Tf(-T) = Tf(T) = 0.2251, Tf(0) = 0.2969$. We also have $g(T) = g(4T) = 0.1601, g(2T) = g(3T) = 0.2727$. For simplicity, we set $f(iT) = 0$, for $|i| > 2$ and $g(iT) = 0$ for $i \le 0$ and $i \ge 5$. Hence, in the simulation, an MAP detector with 4^3 states will be used for this optical recording channel, where the state of precoder is integrated into the state of the channel represented by $g(iT)$. Furthermore, we write $M_0 = \beta N_0$. In the simulation, we consider $\beta = 0$ and 0.15 respectively. Note that for $\beta = 0$, there is no jitter noise.

A. ECC-RLL System

The system under consideration is shown in Fig. 1. A sequence of information bits is first encoded by the Reed-Solomon (RS) encoder. Then, its output is sequentially processed by an M-level RLL encoder, an M-level pulse length modulation (PLM) precoder, an M-level signal mapper to result in M-level signals which will be used as input to the optical recording channel. The output of the channel is processed by an M-level MAP (maximum a posteriori) detector for the precoded optical recording channel, an M-level RLL decoder, and finally an RS decoder to retrieve the information bits.

For the RLL code, we consider $(M, d, k) = (4, 0, 4)$ code with $n = 10$ and rate $R \approx 1.9750$ bits/symbol, $(M, d, k) = (4, 0, 7)$ code with $n = 19$ and rate $R \approx 1.9868$ bits/symbol respectively, which are given in Table III. The largest lengths of the associated αBURST's are 12 bits and 30 bits respectively, while the largest lengths of the associated βBURST's are both equal to $q = 15$.

To provide the error-correcting capability to tackle the error burst in the M-level RLL code of length Ln, we consider the

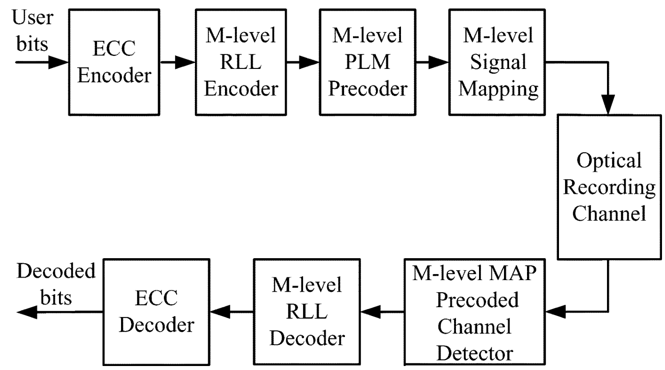

Fig. 1. An ECC-RLL multilevel optical recording system.

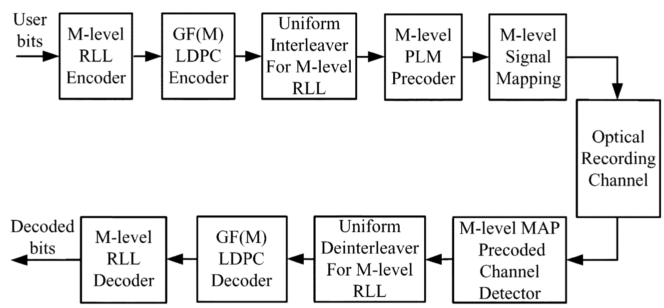

Fig. 2. An RLL-ECC multilevel optical recording system.

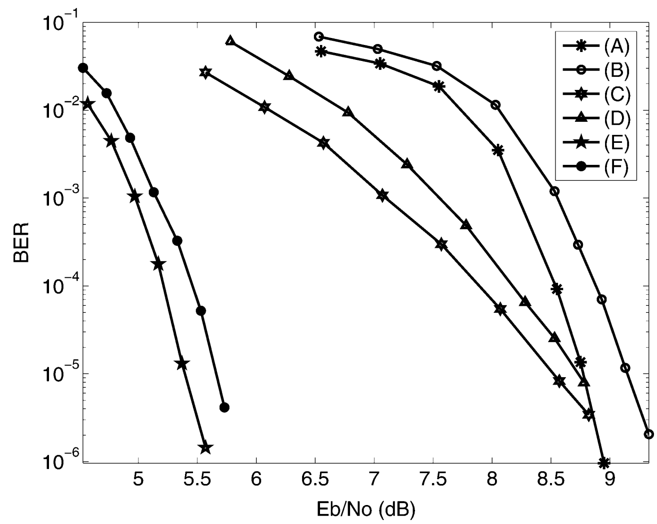

Fig. 3. Simulation results of an optical recording system with AGWN noise only (i.e., $\beta = 0$). (A) RS Code + 4-level (0, 4) RLL Code; (B) RS Code + 4-level (0, 7) RLL Code; (C) 4-level (0, 3) RLL Code + LDPC Code with hard decoding, (0, 4) constraint; (D) 4-level (0, 6) RLL Code + LDPC Code with hard decoding, (0, 7) constraint; (E) 4-level (0, 3) RLL Code + LDPC Code with soft decoding, (0, 4) constraint, 10 iterations; (F) 4-level (0, 6) RLL Code + LDPC Code with soft decoding, (0, 7) constraint, 10 iterations.

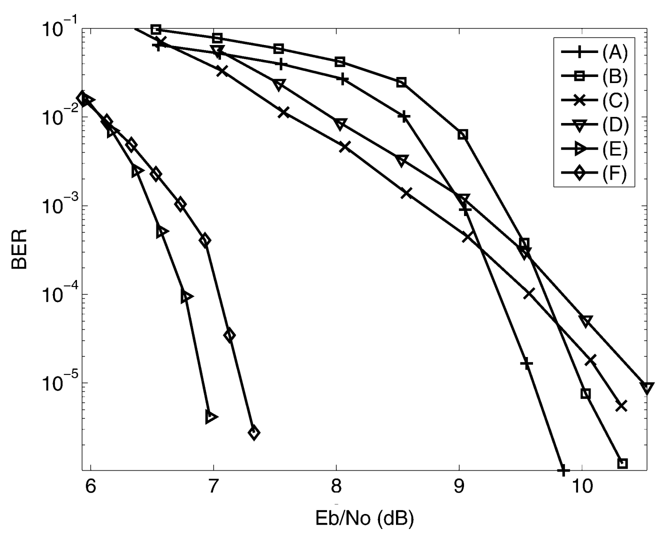

Fig. 4. Simulation results of an optical recording system with both AGWN and jitter noise ($\beta = 0.15$). (A) RS Code + 4-level (0, 4) RLL Code; (B) RS Code + 4-level (0, 7) RLL Code; (C) 4-level (0, 3) RLL Code + LDPC Code with hard decoding, (0, 4) constraint; (D) 4-level (0, 6) RLL Code + LDPC Code with hard decoding, (0, 7) constraint; (E) 4-level (0, 3) RLL Code + LDPC Code with soft decoding, (0, 4) constraint, 10 iterations; (F) 4-level (0, 6) RLL Code + LDPC Code with soft decoding, (0, 7) constraint, 10 iterations.

(511, 455) Reed-Solomon code based on $\mathrm{GF}(2^9)$, which is equivalently a (4599, 4095) binary code with code rate approximately 0.8904. This RS code is capable of correcting 28 9-bit erroneous symbols. For the $(M,d,k) = (4,0,4)$ code using Construction $\mathrm{II}^{\mathrm{ext}}$ with $n = 10$, each $\alpha\mathrm{BURST}$ may result in up to two erroneous 9-bit symbols in the RS code, while each $\beta\mathrm{BURST}$ may also result in up to two erroneous 9-bit symbols. Hence, the RS code can correct 7 longest $\alpha\mathrm{BURST}$'s plus 7 longest $\beta\mathrm{BURST}$'s for $(M,d,k) = (4,0,4)$ code with $n = 10$, and 4 longest $\alpha\mathrm{BURST}$'s plus 6 longest $\beta\mathrm{BURST}$'s for $(M,d,k) = (4,0,7)$ code with $n = 19$. This code is also capable of correcting a combination of some bursts and some random errors for each k.

The simulation results of bit error rate (BER) versus E_b/N_0 based on β of 0 and of 0.15 are plotted in Figs. 3 and 4, respectively. From Fig. 3 and Fig. 4, we see that using RS code can somewhat alleviate the effect of channel errors and the error propagation phenomenon of the RLL codes. For either $\beta = 0$ or of 0.15, the BER curves for the $(M,0,k) = (4,0,4)$ code are better than the curves for the (4, 0, 7) code. This results fits our expectation since the (4, 0, 4) code has less serious error propagation.

B. RLL-ECC System

For the second scheme shown in Fig. 2, we apply the concept of the RLL-ECC concatenation scheme [9] to uniformly interleaves the nonbinary ECC codeword to satisfy the desired multilevel RLL constraint. In order to impose the M-level $(0,k)$ RLL constraint to the interleaved nonbinary ECC codeword, we need to employ a slightly tighter M-level $(0,k-h)$ RLL constraint to the symbol sequence prior to the M-level ECC encoder, where h is the number of inserted symbol from nonbinary ECC parity symbols. Here, we use the LDPC code over $\mathrm{GF}(M)$ as the nonbinary ECC.

The sequence of information bits is first encoded by an M-level RLL encoder. Then, its output is sequentially processed by a nonbinary systematic LDPC encoder, a uniform interleaver, an M-level PLM precoder, an M-level signal mapper to result in M-level signals which will be used as input to the optical recording channel. The output of the channel is processed by an MAP detector for the precoded optical recording channel, a uniform deinterleaver, a nonbinary LDPC decoder, and an M-level RLL decoder to retrieve the recorded information bits. The RLL-ECC system is expected to significantly reduce the error propagation from the RLL decoder to the ECC decoder. However, the error propagation from the input of the RLL decoder to its output still exists.

From Table III, we have (4, 0, 3) code with $n = 7$ and rate $R \approx$ 1.9642 bits/symbol, (4, 0, 6) code with $n = 16$ and rate $R \approx$ 1.9843 bits/symbol. With $M = 4, k = 4, h = 1$ and $k - h = 3$, each (4, 0, 3) sequence will be converted into a sequence satisfying the (4, 0, 4) constraint after passing through the systematic ECC and the uniform interleaver. With $M = 4, k = 7, h = 1$ and $k - h = 6$, each (4, 0, 6) sequence will be converted into a sequence satisfying the (4, 0, 7) constraint after passing through the systematic ECC and the uniform interleaver [9]. Following the progressive edge-growth LDPC code construction [26] and nonbinary LDPC [27] code construction, a PEG nonbinary LDPC code is obtained for this system. Specifically, we have the (2297, 2048) LDPC code over $\mathrm{GF}(4)$. The overall system rate of (4, 0, 4) and (4, 0, 7) are $1.9642 * (2048/2297) \approx 1.7512$ and $1.9843 * (2048/2297) \approx 1.7691$. Note that for the ECC-RLL which uses the (511, 455) Reed-Solomon code over $\mathrm{GF}(2^9)$ concatenated with the $(M, d, k) = (4, 0, 4)$ and $(M, d, k) = (4, 0, 7)$ RLL codes have rate of $1.9750 * (455/511) \approx 1.7585$ and $1.9868 * (455/511) \approx 1.7690$, respectively. We consider hard decoding and soft decoding respectively for the nonbinary LDPC code.

From the simulation results given in Figs. 3 and 4, we see that the BER curves of the RLL-ECC system with hard decoding have flatter slopes than those comparable curves in the ECC-RLL system. This slope of the BER curve for a ECC reflects its error-correcting capability. However, at low E_b/N_0, the better BER performance of the RLL-ECC system verifies its advantage of eliminating the error propagation from RLL to ECC which may occurs in the ECC-RLL system.

Note that using the nonbinary LDPC code with soft decoding in the RLL-ECC system can provide very excellent error performance for either $\beta = 0$ or $\beta = 0.15$, where the soft output of the MAP precoded channel detector is required and 10 iterations are used for the decoding. This result shows the powerful error-correcting capability of the nonbinary LDPC code with soft decoding even if the noise is not restricted to the AWGN noise.

For this RLL-ECC system, the BER curves for the $(M, 0, k) = (4, 0, 4)$ constraint are better than the curves for the (4, 0, 7) constraint. Note that, in the RLL encoding/decoding for this system, $(M, 0, k) = (4, 0, 3)$ code will be used for the (4, 0, 4) constraint and (4, 0, 6) code will be used for the (4, 0, 7) constraint. The largest lengths of the associated $\alpha\mathrm{BURST}$'s are 6 bits and 24 bits respectively, while the largest lengths of the associated $\beta\mathrm{BURST}$'s are both equal to $q = 15$. The less serious error propagation of the (4, 0, 3) RLL code as compared to the (4, 0, 6) RLL code may result in the better BER performance for the case of the (4, 0, 4) RLL constraint as compared to the case of the (4, 0, 7) RLL constraint.

IV. Conclusion

We propose two constructions for $(M, 0, k, k_l, k_r)$ run-length-limited codes of finite length. The code obtained from any of the proposed constructions has coding rate very close to the capacity. In addition, we propose variations of the proposed constructions to achieve the merit of low complexity of encoding and decoding.

When an error-corrupted RLL sequence is applied to the input of an RLL decoder, errors at some positions may cause serious error propagation. We analyze the possible error propagation of the $(M, 0, k)$ codes proposed in Construction $\mathrm{II}^{\mathrm{ext}}$. In particular, the numerical values of error propagation of the $(M, 0, k) = (4, 0, 3), (4, 0, 4), (4, 0, 6)$ and the (4, 0, 7) RLL codes are provided. The error propagation of the $(M, 0, k)$ RLL code is more serious than that of the $(M, 0, k')$ RLL code for $k > k'$. We consider the application of the constructed RLL codes to the optical recording channel corrupted by the jitter noise and the AWGN noise. Simulation results show that, in the ECC-RLL system, using the (4, 0, 4) code can achieve better BER as compared to using the (4, 0, 7) code. Simulation results also show that, in the RLL-ECC system, using the (4, 0, 3) RLL code for the (4, 0, 4) constraint can achieve better BER as compared to using the (4, 0, 6) RLL code for the (4, 0, 7) constraint.

Acknowledgment

This work was supported by the National Science Council of the R.O.C. under Grants NSC 97-2219-E-002-001 and NSC 97-2221-E-002-145-MY3. The authors are grateful to the anonymous reviewers who provided valuable comments and suggestions which significantly enhanced the quality of this paper.

References

[1] B. H. Marcus, P. H. Siegel, and J. K. Wolf, "Finite-state modulation codes for data storage," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 1, pp. 5–37, Jan. 1992.
[2] K. A. S. Immink, *Coding Techniques for Digital Recorders*. Englewood Cliffs, NJ: Prentice Hall, 1991.
[3] K. A. S. Immink, *Codes for Mass Data Storage Systems*. Eindhoven, The Netherlands: Shannon Foundation, 2004.
[4] A. J. van Wijngaarden and K. A. S. Immink, "Maximum runlength limited codes with error control capabilities," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 4, pp. 602–611, Apr. 2001.
[5] K. A. S. Immink and A. J. van Wijngaarden, "Simple high-rate constrained codes," *Electron. Lett.*, vol. 32, no. 20, pp. 1877–, Sep. 1996.
[6] A. Kunisa, "Runlength violation of weakly constrained code," *IEEE Trans. Commun.*, vol. 50, no. 1, pp. 1–6, Jan. 2002.
[7] I. J. Fair, W. D. Gover, W. A. Krzymien, and R. I. MacDonald, "Guided scrambling: A new line coding technique for high bit rate fiber optic transmission systems," *IEEE Trans. Commun.*, vol. 39, pp. 289–297, Feb. 1991.
[8] Y. Han and W. E. Ryan, "Concatenating a structured LDPC code and a constrained code to preserve soft-decoding, structure, and burst correction," *IEEE Trans. Magn.*, vol. 42, no. 10, pp. 2558–2560, Oct. 2006.
[9] J. Lu and K. G. Boyer, "Novel RLL-ECC concatenation scheme for high-density magnetic recording," *IEEE Trans. Magn.*, vol. 43, no. 6, pp. 2271–2273, Jun. 2007.
[10] B. Vasic and K. Pedagani, "Run-length-limited low-density parity check codes based on deliberate error insertion," *IEEE Trans. Magn.*, vol. 40, no. 3, pp. 1738–1743, May 2004.
[11] Z. Li and B. V. K. V. Kumar, "Low-density parity-check codes with run length limited (RLL) constraints," *IEEE Trans. Magn.*, vol. 42, no. 2, pp. 344–349, Feb. 2006.
[12] H.-Y. Chen, M.-C. Lin, and Y.-L. Ueng, "Low-density parity-check coded recording systems with run-length-limited constraints," *IEEE Trans. Magn.*, vol. 44, no. 9, pp. 2235–2242, Sep. 2008.
[13] C. French, G. Dixon, and J. Wolf, "Results involving d,k-constrained M-ary codes," *IEEE Trans. Magn.*, vol. MAG-23, no. 5, pp. 3678–3680, Sep. 1987.
[14] D. T. Tang and L. R. Bahl, "Block codes for a class of constrained noiseless channels," *Inf. Control*, vol. 17, pp. 436–461, 1970.
[15] G. Jacoby, "Ternary 3PM magnetic recording code and system," *IEEE Trans. Magn.*, vol. MAG-17, no. 6, pp. 3326–3328, Nov. 1981.

[16] A. Earman, "Optical data storage with electron trapping materials using M-ary data channel coding," in *Optical Data Storage 92; Proc. SPIE*, D. Carlin and D. Kay, Eds., 1992, vol. 1663, pp. 92–103.
[17] F. H. Lo, J. W. Kuo, N. H. Tseng, J. J. Ju, and D. Howe, "Recording of multi-level run-length-limited modulation signals on compact disc/digital versatile disc rewritable discs," *Jpn. J. Appl. Phys.*, vol. 43, no. 7B, pp. 4852–4855, Jul. 2004.
[18] C. A. French, G. S. Dixon, and J. Wolf, "Results involving (d,k)-constrained M-ary codes," *IEEE Trans. Magn.*, vol. 23, no. 5, pp. 3678–3680, Sep. 1987.
[19] S. W. McLaughlin, "Five runlength-limited codes for M-ary recording channels," *IEEE Trans. Magn.*, vol. 33, no. 3, pp. 2442–2450, May 1997.
[20] S. W. McLaughlin, J. Luo, and Q. Xie, "On the capacity of M-ary run-length-limited codes," *IEEE Trans. Inf. Theory*, vol. 41, pp. 1508–1511, Sep. 1995.
[21] S. W. McLaughlin, "The construction of M-ary (d, ∞) codes that achieve capacity and have the fewest number of encoder states," *IEEE Trans. Inf. Theory*, vol. 43, pp. 699–703, Mar. 1997.
[22] J. L. Fan and A. R. Calderbank, "A modified concatenated coding scheme, with applications to magnetic storage," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1565–1574, Jul. 1998.
[23] J. W. M. Bergmans*, Digital Baseband Transmission and Recording*. Boston, MA: Kluwer, 1996.
[24] X. Yang, Maligono, Leihuang, Pingyan, and H. Zhang, "Performance of run-length limited (4, 18) code for optical storage systems," *Opt. Quantum Electron.*, vol. 36, pp. 1079–1088, Aug. 2004.
[25] J. Pei, H. Hu, L. Pan, Q. Shen, H. Hu, and D. Xu, "Constrained code and partial-response maximum-likelihood detection for high density multi-level optical recording channels," *Jpn. J. Appl. Phys.*, vol. 46, no. 6B, pp. 3771–3774, Jun. 2007.
[26] X. Hu, E. Eleftheriou, and D. Arnold, "Progressive edge-growth Tanner graphs, global telecommunications conference," in *Proc. IEEE GLOBECOM*, Nov. 25–29, 2001, vol. 2, pp. 995–1001.
[27] W. Chang and J. R. Cruz, "Performance and decoding complexity of nonbinary LDPC codes for magnetic recording," *IEEE Trans. Magn.*, vol. 44, no. 1, pp. 211–216, Jan. 2008.

Hsin-Yi Chen was born in Chiayi County, Taiwan, R.O.C., on November 6, 1969. He received the Ph.D. degree in communication engineering from National Taiwan University, Taipei, Taiwan, in 2009.

Since 1998, he has served as an Assistant Researcher in the Telecommunication Laboratories of Chunghwa Telecom Co., Ltd., Taipei, Taiwan. His research interests are in the area of coding theory and constrained coding for recoding systems.

Hong-Fu Chou received the B.S. degree in electrical engineering from National Central University, Jhongli, Taiwan, R.O.C., in 2004 and the M.S. degree in communication engineering from National Taiwan University, Taipei, Taiwan, in 2006. He is working toward the Ph.D. degree at the Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan. His research interests include communication IC, coding theory, and coding for recoding systems.

Mao-Chao Lin (M'87) was born in Taipei, Taiwan, R.O.C., on December 24, 1954. He received the bachelor and master degrees, both in electrical engineering, from National Taiwan University in 1977 and 1979, respectively, and the Ph.D. degree in electrical engineering from University of Hawaii in 1986.

From 1979 to 1982, he was an Assistant Scientist, Chung-Shan Institute of Science and Technology at Lung-Tan, Taiwan. He is currently a Professor in the Department of Electrical Engineering, National Taiwan University. His research interests are in the area of coding theory and its applications.

Shih-Kai Lee (M'08) was born in Taipei, Taiwan, R.O.C., on July 4, 1967. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from National Taiwan University (NTU) in 1989, 1991, and 1998 respectively.

From 1998 to 2004, he worked as an R&D engineer in the Computer and Communication Laboratory (CCL) of Industrial Technology and Research Institute (ITRI), Hsinchu, Taiwan. During this period, He had applied several inventions in the category of wireless communications for U.S. patents and R.O.C. patents. He had also received several ITRI institute awards for his good performance in WLAN and SDR projects. He is currently an Assistant Professor in the Department of Communication Engineering, Yuan Ze University (YZU), Taoyuan, Taiwan. His research efforts are focused on technical issues such as multi-carrier synchronization and equalization, channel coding and ARQ protocols in the physical (PHY) and medium-access-control (MAC) layers of wireless communications.