IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 19, 2020, accepted November 4, 2020, date of publication November 11, 2020,
date of current version November 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3037169

Marker Codes Using the Decoding Based on Weighted Levenshtein Distance in the Presence of Insertions/Deletions

YUAN LIU, (Member, IEEE), YIWEI LU, YASHUO HE, XIAONAN ZHAO, AND CUIPING ZHANG
Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China
College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin 300387, China

Corresponding author: Yuan Liu (liuyuan@tjnu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61801327, in part by the Doctor Fund of Tianjin Normal University under Grant 043135202-XB1711, in part by the Natural Science Foundation of Tianjin City under Grant 18JCYBJC86400, and in part by the Tianjin Higher Education Creative Team Funds Program.

ABSTRACT A random marker code is inserted into the information sequences periodically, and a novel symbol-level decoding algorithm considering the weighted Levenshtein distance (WLD) is designed for correcting insertions, deletions, as well as substitutions in the received sequences. In this method, branch quantities in the decoding trellis are calculated by measuring the WLD, which is done using the dynamic programming. A simulation study is performed to demonstrate the effectiveness of the presented scheme in the practical system, especially for channels with weak synchronization problems.

INDEX TERMS Marker code, weighted Levenshtein distance, insertions/deletions, synchronization.

I. INTRODUCTION

Loss of synchronization due to the imperfect of the sampling clocks may cause catastrophic consequences with the variable length and enormous substitutions, which are of great interest in the communication systems [1]–[5]. Channels with errors caused by the loss synchronization have memory, and the techniques designed for memoryless channels can seldom be employed directly [6]–[12].

The DM construction proposed by Davey and MacKay is the most promising technique for recovering the synchronization. In this scheme, a watermark code is used to correct synchronization errors including insertions and deletions, and a non-binary low-density parity-check (LDPC) code is employed to correct the residual errors. The watermark as a known code is added modulo 2 to the LDPC code. At the receiver, the watermark decoder passes the decoding trellis and compares the received sequences with the known watermark bit-by-bit to output the log-likelihood ratios (LLRs).

In the recent past, on one hand, some modifications to the decoding algorithm of the DM construction have been made for the channel with synchronization errors to improve the performance. Briffa designed a symbol-level watermark decoding algorithm that takes the codebook of LDPC code into account, and reduced the block error rate significantly [13]. Subsequently, based on the scheme in [13], Jiao proposed an iterative decoding algorithm which allows soft a priori input, and further improved the synchronization-error-correcting capability of the system [14]. On the other hand, several new encoding schemes based on the DM construction are presented. In [15], marker code was used in the concatenated code, and provides easier synchronization. Marker can be viewed as an irregular version of the watermark. In terms of the decoding scheme in [15], it employs the bit-level forward-backward algorithm in order to identify the insertions and deletions.

In this paper, we focus on the communication over the binary insertions/deletions-substitutions channel model adopted in [16]. Particularly, we propose that the marker code as a known pattern is inserted at regular intervals. Furthermore, considering that the branch quantities in the computations of the forward/backward quantities can be replaced by the weighted Levenshtein distance (WLD), which is not constrained by the inner encoding method, a novel forward-backward decoding algorithm in the pure symbol-level is presented for the correction of synchronization errors. The branch quantities in the decoding trellis are efficiently calculated by taking into account the WLD, where the

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

WLD is used to measure the distance between the transmitted and received sequences having insertions/deletions. The forward-backward decoding algorithm based on WLD is powerful. Moreover, unlike previous algorithms, the presented algorithm can be used flexibly in the system when markers or watermarks are employed.

The rest of the paper is organized as follows. Section II introduces the system model and the encoding scheme. Section III describes the method for computing the WLD and the proposed forward-backward decoding scheme. In Section IV, results of simulations conducted to prove the effectivity of the proposed scheme. Finally, some conclusions are drawn in Section V.

II. SYSTEM MODEL AND ENCODER

In this paper, the binary insertion/deletion-substitution channel with random and independent bit errors is considered, which can be modeled the imperfect synchronization. A proposed reliable communication system working on this binary channel will be illustrated, and the marker encoding method will be described in detail in this section.

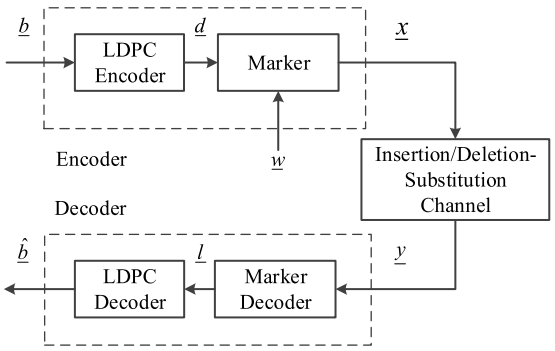

FIGURE 1. The system model.

A. SYSTEM MODEL

The proposed scheme depicted in Fig. 1 employs the known marker code as an inner coder, and uses the binary LDPC code as an outer code. The binary information $\underline{b}$ is first encoded by the outer LDPC code, and is mapped into the binary code $\underline{d}$ (N_L, K_L). The known marker code $\underline{w}$ of length N is then inserted into $\underline{d}$ uniformly, which producing the transmitted code $\underline{x}$ with the length of N_c. $\underline{x}$ is sent over the random binary insertion/deletion- substitution (IDS) channel [16].

The models of the IDS channel are shown in Fig. 2, where N_c^* is the length of the received sequences. For each bit of the sequences, there will be four situations that may occurs. Specifically, a random bit is inserted into the sequence with P_i, or the transmitted bit is deleted with P_d, or the bit is added (modulo 2) to the bit '1' with P_s, or the bit is transmitted correctly, where parameters P_i, P_d, and P_s denote insertion, deletion, substitution probabilities, respectively. The transmission probability $P_\mathrm{t} = 1 - P_\mathrm{i} - P_\mathrm{d}$. For each block, the probability of the channel making N_i insertions, N_d deletions,

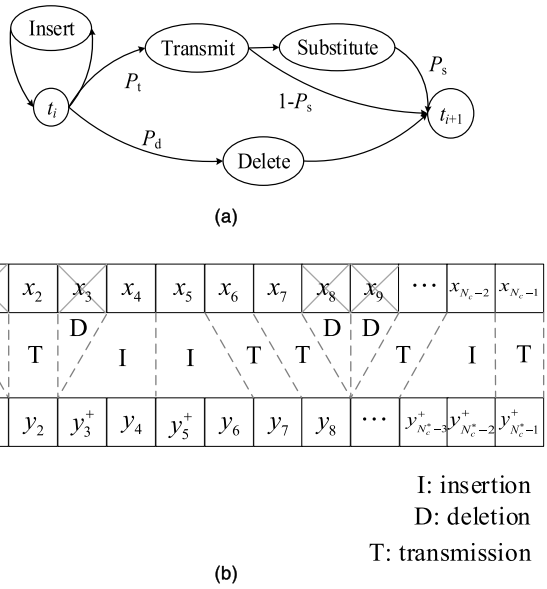

FIGURE 2. The IDS channel model.

and N_s substitutions is computed as follows [17].

$$P(N_\mathrm{i}, N_\mathrm{d}, N_\mathrm{s}) = P_\mathrm{i}^{N_\mathrm{i}} \cdot P_\mathrm{d}^{N_\mathrm{d}} (P_\mathrm{t} P_\mathrm{s})^{N_\mathrm{s}} (P_\mathrm{t}(1 - P_\mathrm{s}))^{N_c - N_\mathrm{d} - N_\mathrm{s}}. \tag{1}$$

B. ENCODING SCHEME

The outer encoder employs the binary LDPC encoding method due to its excellent error-correcting performance. Then, the LDPC code $\underline{d}$ is first divided into N_L/m symbols each of which having m-bits. The number of values q that the symbol taking is satisfied $q = 2^m$. In this paper, we choose the pseudo-random sequence as the marker code. The inner encoder allocates λ bits of the marker code to each sub-sequence of $\underline{d}$ in order to generate $\underline{x}$. The structure of the code $\underline{x}$ is illustrated in Fig. 3.

III. THE PROPOSED INNER DECODING METHOD USING THE WLD

In this section, the proposed novel forward-backward algorithm decoding on the symbol-level will be described in detail.

A. WLD

The distance between the transmitted and received sub-sequences is measured by computing the WLD, which is done using the dynamic programming [18]. The WLD is used to calculate the output probabilities which will be shown in sub-section B. Define the WLD between arbitrary two sequences $\underline{s}$ and $\underline{s}'$ as follows.

$$WLD(\underline{s}, \underline{s}') = \min_{N_\mathrm{d}, N_\mathrm{s}, N_\mathrm{i}} (\omega_\mathrm{d} N_\mathrm{d} + \omega_\mathrm{s} N_\mathrm{s} + \omega_\mathrm{i} N_\mathrm{i}), \tag{2}$$

where, ω_d, ω_s, and ω_i are weighting factors.

B. THE PROPOSED FORWARD-BACKWARD ALGORITHM

In order to recover the synchronization, based on the decoding trellis, the inner decoder identifies the insertions

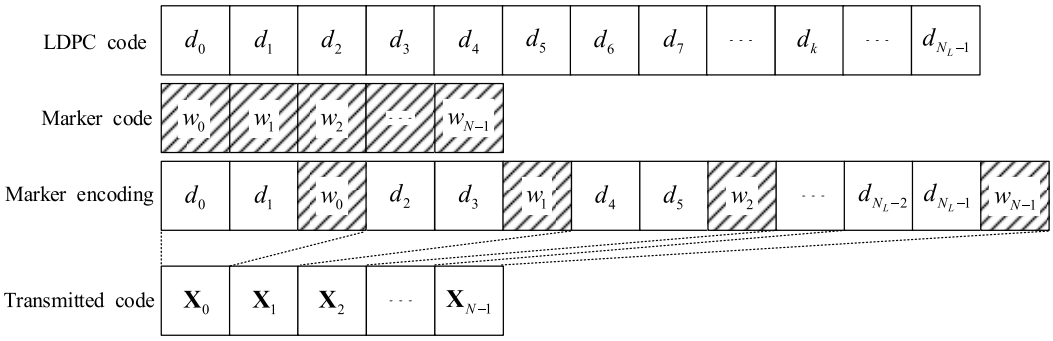

FIGURE 3. The structure of the concatenated code.

and deletions in the received sequences. By executing the symbol-level forward and backward passes, LLRs can be obtained, which initialize the LDPC decoder to correct the residual errors.

Since the symbol probabilities are equal, the LLR for i-th bit is computed in (3).

$$\begin{aligned} l_k &= \frac{P(d_k=1|\underline{y})}{P(d_k=0|\underline{y})} \\ &= \frac{P(\underline{y}|d_k=1)}{P(\underline{y}|d_k=0)} \\ &= \frac{\sum\limits_{\tilde{d}_i} P(\underline{y}|\tilde{d}_i=a, d_k=1)}{\sum\limits_{\tilde{d}_i} P(\underline{y}|\tilde{d}_i=a, d_k=0)}, \end{aligned} \tag{3}$$

where $0 \le k < N_L$, $0 \le i < (N_L/m)$, $\tilde{d}_i \to (d_{mi}, \cdots, d_{m(i+1)-1})$,

$$P(\underline{y}|\tilde{d}_i=a) = \sum_{t_i, t_{i+1}} F_i(t_i) M_{i+1}(\tilde{d}_i=a) B_{i+1}(t_{i+1}), \tag{4}$$

where $0 < a \le q-1, 0 \le i < (N_L/m)$, $\tilde{d}_i$ is the symbol value corresponding to the string having m-bits, the state t_i is the drift at the i-th position and $t_i = N_\mathrm{i} - N_\mathrm{d}$, $M(\cdot)$ denotes the middle quantity. If the bit x_i is not deleted then it will appear in the sequences as $x_{i+t_{i+1}}$. The maximum of the state $t_{\max}$ is set to $5\sqrt{N_\mathrm{c}P_\mathrm{d}/(1-P_\mathrm{d})}$, and the number of states at each time is $T = 2t_{\max}+1$.

The symbol-level forward probability that the drift t_i is τ and that the first $((m+\lambda)\times i+\tau-1)$ bits output by the channel are calculated as follows.

$$\begin{aligned} &F_i(t_i=\tau) \\ &= P(y_0, \cdots, y_{(m+\lambda)\times i+\tau-1}, t_i=\tau) \\ &= \sum_{c, \tilde{d}_{i-1}} F_{i-1}(t_{i-1}=c) P(\tilde{d}_{i-1}) P(\underline{y'}, t_i=\tau|t_{i-1}=c, \tilde{d}_{i-1}), \end{aligned} \tag{5}$$

where $\underline{y'} = (y_{(m+\lambda)\times(i-1)+c}, \cdots, y_{(m+\lambda)\times i+\tau-1})$, and the conditional probability $P(\underline{y'}, t_i=\tau|t_{i-1}=c, \tilde{d}_{i-1})$ is called the branch quantity.

Similarly, the symbol-level backward probability is calculated as follows.

$$\begin{aligned} &B_i(t_i=\tau) \\ &= P(y_{(m+\lambda)\times i+\tau}, \cdots|t_i=\tau) \\ &= \sum_{b, \tilde{d}_i} B_{i+1}(t_{i+1}=b) P(\tilde{d}_i) P(\underline{y''}, t_{i+1}=b|t_i=\tau, \tilde{d}_i). \end{aligned} \tag{6}$$

where, $\underline{y''} = (y_{(m+\lambda)\times i+\tau}, \cdots, y_{(m+\lambda)\times(i+1)+b-1})$.

The branch quantities in eq. (5) and eq. (6) are calculated by considering the WLD. We first deduce the log-likelihood function of the branch quantity under the condition that the marker is known for the receiver.

$$\begin{aligned} &\log P(\underline{y'}, t_i=\tau|t_{i-1}=c, \tilde{d}_{i-1}) \\ &\quad = N_\mathrm{i}\log P_\mathrm{i} + N_\mathrm{d}\log\left(\frac{P_\mathrm{d}}{P_\mathrm{t}(1-P_\mathrm{s})}\right) \\ &\qquad + N_\mathrm{s}\log\left(\frac{P_\mathrm{s}}{1-P_\mathrm{s}}\right) + (m+\lambda)\log(P_\mathrm{t}(1-P_\mathrm{s})). \end{aligned} \tag{7}$$

Note that, the above log-likelihood function can be written as $WLD(\underline{s}_{i-1}, \underline{y'}) + (m+\lambda)\log(P_\mathrm{t}(1-P_\mathrm{s}))$, where the length of $\underline{s}_{i-1}$ is $m+\lambda$, and

$$\begin{aligned} \omega_\mathrm{i} &= \log P_\mathrm{i}^{[17]}, \\ \omega_\mathrm{d} &= \log\left(\frac{P_\mathrm{d}}{P_\mathrm{t}(1-P_\mathrm{s})}\right)^{[17]}, \\ \omega_\mathrm{s} &= \log\left(\frac{P_\mathrm{s}}{1-P_\mathrm{s}}\right)^{[17]}. \end{aligned} \tag{8}$$

Thus, the branch quantity $P(\underline{y'}, t_i=\tau|t_{i-1}=c, \tilde{d}_{i-1}) = \exp\Big(WLD(\underline{s}_{i-1}, \underline{y'}) + (m+\lambda)\log(P_\mathrm{t}(1-P_\mathrm{s}))\Big)$.

The middle quantities $M_{i+1}(\tilde{d}_i = a)$ in eq. (4) denotes the probability $P(\underline{y^0}, t_{i+1}|t_i, \tilde{d}_i=a)$, where the sub-sequence $\underline{y^0} = (y_{(m+\lambda)\times i+t_i}, \cdots, y_{(m+\lambda)\times(i+1)+t_{i+1}-1})$. We write $M_{i+1}(\tilde{d}_i=a)$ as follows.

$$\begin{aligned} &M_{i+1}(\tilde{d}_i=a) \\ &\quad = P(\underline{y^0}, t_{i+1}|t_i, \tilde{d}_i=a) \\ &\quad = \exp\Big(WLD(\underline{s}_i, \underline{y^0}) + (m+\lambda)\log(P_\mathrm{t}(1-P_\mathrm{s}))\Big), \end{aligned} \tag{9}$$

where $\underline{s}_i \to (w_i, \tilde{d}_i)$.

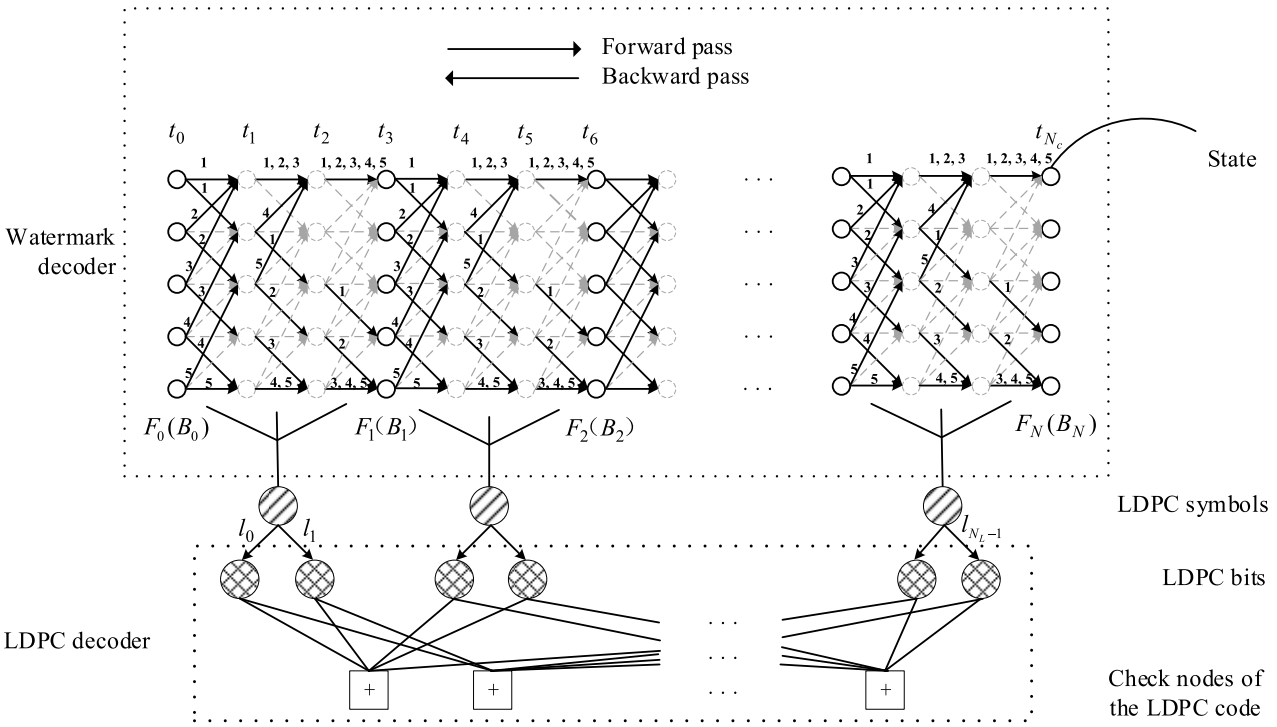

FIGURE 4. Illustration of the decoding procedure for the forward-backward scheme using the WLD. $\lambda = 1$, $m = 2$.

TABLE 1. Parameters of outer codes used in the concatenated codes.

Concatenated Code	Construction of the outer code	N_L (bits)	R_L
A	Irregular quasi-cyclic LDPC code [19]	576	1/2
B	Irregular quasi-cyclic LDPC code [19]	2304	1/2
C	Regular LDPC code constructed by the progressive edge growth algorithm [20]	4000	1/2

TABLE 2. Parameters of concatenated codes reported in this paper.

Concatenated Code	N (bits)	N_c (bits)	R_c
A	288	864	0.33
B	1152	3456	0.33
C	2000	6000	0.33

The diagrammatic sketch for the symbol-level forward and backward passes on the decoding trellis are shown in Fig. 4. In this figure, trellis with $T = 5$ is taken as an example, and the maximum insertions for each state at each bit $I_{\max}$ is set to 2. Specially, $1 \leq j \leq 5$ is considered. The path $j \to j \to j$ is the *maximum pass* for the j-th state, which produces the largest received sequence. The path $j' \to j' \to j'$ is the *minimum pass* for the j-th state, which produces the shortest received sequence. Furthermore, the case for five states is generalized to $T = 2t_{\max} + 1 = 10\sqrt{N_c P_d/(1-P_d)} + 1$. After a symbol-level pass, the j-th state at the i-th symbol can achieve states $\{j - (m+\lambda), \cdots, j + (m+\lambda) \times I_{\max}\}$ at the $(i+1)$-th symbol.

IV. SIMULATION RESULTS

In this section, some examples were used to evaluate the performance of the proposed scheme using the decoding based on the WLD. Table 1 shows the parameters of all outer codes used in the concatenated codes, and Table 2 gives the parameters of concatenated codes whose performance are reported in this paper. In Table 1, two irregular LDPC codes and a regular LDPC code, with different codelengths and different construction methods, were selected as the examples. A binary pseudorandom marker vector was created. $\lambda = 1$, $m = 2.I_{\max} = 5.P_i = P_d$. A belief-propagation algorithm in log-domain was used in the LDPC decoder and the maximum number of iterations was set to 20. Since the drift was set to zero at the start, the forward quantities for 0-th symbol in the first block are calculated as follows.

$$F_0(t) = \begin{cases} 1, & t = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

The backward quantities in each block are initialized with equal probabilities since the accurate block boundary is unknown at the receiver.

In order to demonstrate the efficiency of the proposed decoding scheme, a set of standard concatenated codes using binary LDPC codes were simulated to evaluate the performance of algorithms. As shown in the following three pictures, i.e., Fig. 5-7, all the block error rates (BERs) of code A-C decrease with the insertion/deletion (I/D) probabilities reducing. Furthermore, the BERs as a function of I/D

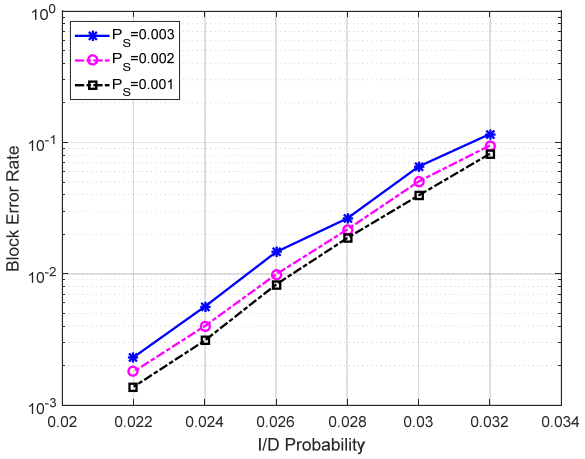

FIGURE 5. Performance of code A.

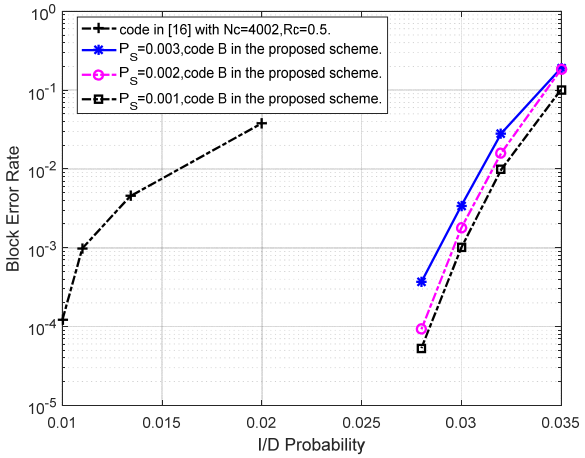

FIGURE 6. Performance of code B.

probabilities for several choices of the substitution probability are illustrated in the Fig. 5-7, respectively. Increasing P_s affects the performance smoothly. As a result, the degradation in the decoding performance with P_s increasing.

Firstly, the performance of code A for the different P_s is evaluated and illustrated in Fig. 5. From the Fig. 5, we notice that there is a graceful degradation in the performance for decreased values of (I/D) probabilities. Next, we simulate the decoding performance when using code B. It is worth mentioning that this comparison is not entirely fair as the inner codes consider in [16] is watermark code while in our scheme is marker code. This comparison is illustrated in Fig. 6. To make the comparison as fair as possible, we select the code B in Table 2 according to the parameters of codes considered in [16] with $N_c = 4002$ and $R_c = 0.5$. It is clear from Fig. 6 that an evident improvement in the error correction performance is achieved by using the proposed scheme. Furthermore, when code C is employed, three simulation curves with different substitution probabilities are shown in Fig. 7. As is exhibited in Fig. 7, the error correction performance is significantly improved with the decreasing of insertions and deletions. Therefore, the efficiency of

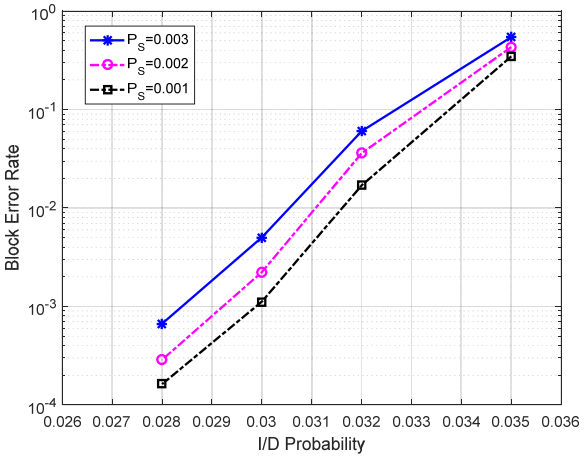

FIGURE 7. Performance of code C.

the proposed scheme for the given channel model is significant, and simulation results fully meet the performance requirements.

In detail, as is shown in Fig. 7, the BER of less than 10^{-3} was achieved for $P_i = P_d = 0.03$ and $P_s = 0.001$. At these noise levels there are an average of 360 synchronization and 6 substitution errors per block. Obviously, the performance of the proposed decoding scheme is effective.

V. CONCLUSION

In this paper, we proposed a decoding framework employing the random marker code inserted into the information sequences periodically, and an innovative symbol-level decoding scheme considering the WLD is designed for correcting insertions, deletions along with substitutions in the received sequences. The performance of the forward-backward decoding algorithm based on WLD is effective, which is demonstrated through simulation that conspicuous amount of insertions and deletions could be corrected by the proposed scheme. Moreover, unlike earlier forward-backward decoding algorithms, the presented algorithm can be used flexibly in the system where marker or watermark is used. Future work will investigate the construction of better encoding/decoding algorithms that are suited for the given channel model.

REFERENCES

[1] H. Mercier, V. K. Bhargava, and V. Tarokh, "A survey of error-correcting codes for channels with symbol synchronization errors," *IEEE Commun. Surveys Tuts.*, vol. 12, no. 1, pp. 87–94, 1st Quart., 2010.

[2] F. Wang, D. Fertonani, and T. M. Duman, "Marker code optimization and symbol-level synchronization for insertion/deletion channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 2010, pp. 1007–1011.

[3] Y. Liu, Y. He, X. Zhao, M. Xie, Y. Hong, and C. Zhang, "Irregular marker codes for insertion/deletion-AWGN channels," *IEEE Access*, vol. 8, pp. 50733–50739, 2020.

[4] M. Kovacevic and V. Y. F. Tan, "Coding for the permutation channel with insertions, deletions, substitutions, and erasures," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 1933–1937.

[5] M. Inoue and H. Kaneko, "Insertion/deletion/substitution error correction using adaptive inversion of synchronization marker," in *Proc. Int. Symp. Inf. Theory Appl.*, Honolulu, HI, USA, 2012, pp. 221–225.

[6] W. Chen and Y. Liu, "Efficient transmission schemes for correcting insertions/deletions in DPPM," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.
[7] V. Buttigieg and N. Farrugia, "Improved bit error rate performance of convolutional codes with synchronization errors," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 4077–4082.
[8] Q. Wang, S. Jaggi, M. Medard, V. R. Cadambe, and M. Schwartz, "File updates under random/arbitrary insertions and deletions," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6487–6513, Oct. 2017.
[9] A. S. J. Helberg and H. C. Ferreira, "On multiple insertion/deletion correcting codes," *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 305–308, Aug. 2002.
[10] T. Wu and M. A. Armand, "The davey-MacKay coding scheme for channels with dependent insertion, deletion, and substitution errors," *IEEE Trans. Magn.*, vol. 49, no. 1, pp. 489–495, Jan. 2013.
[11] R. Yazdani and M. Ardakani, "Reliable communication over non-binary insertion/deletion channels," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3597–3608, Dec. 2012.
[12] C. Schoeny, A. Wachter-Zeh, R. Gabrys, and E. Yaakobi, "Codes correcting a burst of deletions or insertions," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 1971–1985, Apr. 2017.
[13] J. A. Briffa, H. G. Schaathun, and S. Wesemeyer, "An improved decoding algorithm for the davey-MacKay construction," in *Proc. IEEE Int. Conf. Commun.*, Cape Town, South Africa, May 2010, pp. 1–5.
[14] X. Jiao and M. A. Armand, "Interleaved LDPC codes, reduced-complexity inner decoder and an iterative decoder for the Davey-Mackay construction," in *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, Russia, Jul. 2011, pp. 742–746.
[15] E. A. Ratzer, "Marker codes for channels with insertions and deletions," *Ann. Telecommun.*, vol. 60, no. 1, pp. 29–44, Feb. 2005.
[16] M. C. Davey and D. J. C. Mackay, "Reliable communication over channels with insertions, deletions, and substitutions," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 687–698, Feb. 2001.
[17] M. Mansour and A. Tewfik, "Convolutional decoding in the presence of synchronization errors," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 2, pp. 218–227, Feb. 2010.
[18] J. B. Kruskal, "An overview of sequence comparison: Time warps, string edits, and macromolecules," *SIAM Rev.*, vol. 25, no. 2, pp. 201–237, Apr. 1983.
[19] *IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1*, IEEE Standard 802.16e-2005, IEEE Computer Society and the IEEE Microwave Theory and Techniques Society, New York, NY, USA, Feb. 2006.
[20] S. Khazraie, R. Asvadi, and A. H. Banihashemi, "A PEG construction of finite-length LDPC codes with low error floor," *IEEE Commun. Lett.*, vol. 16, no. 8, pp. 1288–1291, Aug. 2012.