IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 May 2022, accepted 10 June 2022, date of publication 15 June 2022, date of current version 23 June 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3183203

Asymptotically Tight MLD Bounds and Minimum-Variance Importance Sampling Estimator for Linear Block Codes Over BSCs

JINZHE PAN AND WAI HO MOW, (Senior Member, IEEE)
Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, SAR, China
Corresponding author: Jinzhe Pan (jpanad@ust.hk)

This work was supported by the General Research Fund of the Hong Kong Research Grants Council (project no. 16214422).

ABSTRACT In this paper, we re-examine the classical problem of efficiently evaluating the block and bit error rate performance of linear block codes over binary symmetric channels (BSCs). In communication systems, the maximum likelihood decoding (MLD) bounds are powerful tools to predict the error performance of the coded systems, especially in the asymptotic regime of low error probability (or high signal-to-noise ratio). Contrary to the conventional wisdom, we prove that for BSCs, all bounds based on Gallager's first bounding technique, including the famous union bound, are not asymptotically tight for all possible choices of the Gallager region. By proposing the so-called input demodulated-output weight enumerating function (IDWEF) of a code, asymptotically tight MLD upper and lower bounds for BSCs are then derived. In many practical scenarios where performance bounds are not applicable (e.g., due to the unavailability of the relevant coding parameters under a given decoder), the Monte Carlo simulation is commonly used despite its inefficiency, especially in the low error probability regime. We propose an efficient importance sampling (IS) estimator by deriving the optimal IS distribution of the Hamming weight of the error vector. In addition, the asymptotic relative saving on the required sample size of the proposed IS estimator over the state-of-the-art counterpart in the recent literature is characterized. Its accuracy in predicting the efficiency of the proposed IS estimator is verified by extensive computer simulation.

INDEX TERMS Linear block codes, importance sampling, Monte Carlo simulation, asymptotically tight bounds, binary symmetric channel.

I. INTRODUCTION

Linear block codes have been widely used in real-world communication systems for providing reliable data transmission over noisy channels [1], [2]. How to efficiently evaluate the performance of such coded systems, especially at a very low target bit (or word) error rate, is a long-standing problem. Useful performance bounds on the average error probability of the random code are available for any given blocklength and code rate [3]–[5]. However, efficient performance evaluation of a specific coding scheme with a specific decoder remains indispensable for practical code selection and verification of codec implementation. As a result, performance bound analysis and Monte Carlo (MC) performance simulation based on some specific properties of the target code and decoder receive ongoing research interests in the communication community.

Many known bounds on maximum likelihood decoding (MLD) performance have been reported in the literature [6], [7]. Many of them are derived using a general bounding technique developed by Gallager and later referred to as Gallager's first bounding technique (GFBT) by Divsalar [8]. The method introduces the so-called Gallager region around the transmitted codeword to divide the observation space and avoid over-estimating the error probability outside the region. The GFBT-based bounds resulted from different choices of the Gallager region include many well-known bounds as special cases. For instance, for the additive white Gaussian noise (AWGN) channel, the famous union bound can be obtained by setting the Gallager region to the whole observation space. The tangential bound of Berlekamp [9] tightens the union bound by determining the region as a boundary

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

of a plane. The sphere bound [10] derived by Herzberg and Poltyrev chooses the region as a sphere centered at the transmitted signal vector. The tangential sphere bound [11] selects the region as a circular cone whose central line passes through the origin and the transmitted signal vector.

Somewhat surprisingly, we discover that for binary symmetric channels (BSCs), the GFBT-based bounds are not asymptotically tight for all possible choices of the Gallager region. In particular, the knowledge of the minimum Hamming distance and the number of minimum-distance codewords of a code are insufficient for characterizing its asymptotic word error probability. To fix this issue, we propose the so-called input demodulated-output weight enumerating function (IDWEF) of a given linear block code and a given decoder. Intuitively speaking, the IDWEF represents the input-output weight relationship of the decoder, while the conventional input-output weight enumerating function (IOWEF) represents that of the encoder. Hence, unlike the IOWEF, the IDWEF characterizes the classification of all dominating error vectors according to which codewords they are decoded to and contains sufficient knowledge of the encoder-decoder pair to express asymptotically tight bounds.

For many practical coding schemes, the performance bound analysis is difficult to conduct. This is because the parameters (such as the minimum distance, IOWEF or IDWEF) of a given encoder-decoder pair required in the bounding expressions may not have been tabulated and made publicly available. Determining such coding parameters by theoretical analysis or by computer enumeration searches is in general a very challenging and costly (if not, intractable) task. Under such scenarios, MC simulation is a common numerical technique to carry out performance estimation. One drawback of this technique is the requirement of a sufficiently large number of generated samples in order to provide reliable estimates of the target bit error rate (BER) or word error rate (WER), which may be very low in the error rate region of practical interest. For example, in the modern Dense Wavelength Division Multiplexing (DWDM) optical communication systems, the transmission capacity per fiber can exceed a Tbps and the target BER is typical 10^{-12} and sometimes even as low as 10^{-15} [12]. At this level of error probability, performance evaluation by MC simulation is prohibitively complicated. It is noteworthy that due to the high circuit complexity and power consumption associated with the very high transmission rate requirement, hard-decision decoding, in place of soft-decision decoding, has been widely used in optical communication applications [12]. The same holds in the storage applications, such as nanoscale and flash memories [13], [14]. Therefore, a fast simulation method for accurately estimating a very low target error probability under hard-decision decoding (or the resulting BSCs) is highly desirable.

One of the most efficient ways to reduce the computational complexity of the MC method is to use the importance sampling (IS) technique. The idea is to use a biased distribution, called the IS distribution, to generate samples so that the frequency of the occurrence of the rare events can be increased. It can significantly enhance the efficiency in terms of the sample size for simulating the performance of communication systems. Although the optimal IS distribution is known from [15], it does not allow a practical sampling process to be implemented since it depends on the parameters to be estimated itself. Therefore, several sub-optimal but implementable schemes are presented. Many works in the literature choose the IS distribution from a parametric family of distributions like the mean translation [16], the variance scaling [15]. This kind of method is easy to implement but its efficiency in terms of the sample size is low if the parametric family is far away from the optimal IS distribution. To overcome this problem, a mixture of components is used in the multiple IS methods [17], [18], and an iterative adaptation of the IS distribution is applied in the adaptive IS methods [19], [20]. The adaptive IS is used in [21] for error probability estimation for multiple access systems, and a nested IS method that estimates the random-coding error probability of the coded-modulation is provided in [22]. There also exist methods based on non-parametric IS distributions, such as the dual adaptive IS method [23]. The method shows a significant improvement in reducing the sample size of the simulation. However, the empirical reliability measure of the estimated results makes the efficiency analysis of the method ambiguous compared to the MC method.

For BSCs, the method presented in [24] divides the sample space into different regions according to the error weight (i.e., Hamming weight of an error vector) and calculates the WER by manually assigning large enough samples for each weight to counting the number of errors and estimating the conditional error probability. Although the method is efficient in the sense that no additional computation cost is needed for different signal-to-noise ratios (SNRs), the number of samples allocated to each weight in this method is empirically chosen which can be further improved. A quasi-analytical simulation method is presented in [25], which estimates the boundary of the decision region and predicts the error performance. It is efficient under the assumption of the geodesic channel and the star domain decoder. The state-of-the-art IS method over BSCs, called the minimum-variance Bernoulli estimator, is introduced in [26]. It is assumed therein that the IS distribution is a parametric Bernoulli distribution and the optimal parameter is determined by minimizing the variance of the IS estimator. However, the optimal parameter requires the conditional error probabilities for each error weight, which need to be estimated themselves. A fast simulation algorithm, called the IS-MC basic, that estimates both the optimal parameter and the WER iteratively is therein presented.

We take the state-of-the-art IS method in [26] as the benchmark estimator in this paper. If we regard the error weight as a random variable, we find out that this estimator can be interpreted as sampling the error weight with a binomial

distribution as the IS distribution. Apparently, restricting the IS distribution of the error weight to the Binomial family is generally suboptimal. By relaxing the restriction and exploit the larger degree-of-freedom to optimize the IS distribution of the error weight, we propose a new IS estimator that can outperform this benchmark IS estimator [26] in terms of the sample size.

In this paper, we present new WER and BER evaluation methods including both performance bound analysis and IS simulation for linear block codes over BSCs. Our main contributions are summarized as follows.

1. We prove that the GFBT-based bounds for BSCs are not asymptotically tight for all possible choices of the Gallager region. By proposing the so-called IDWEF of a coding scheme, the asymptotically tight MLD upper and lower bounds for linear block codes over BSCs are proposed.
2. The optimal IS distribution among all the possible distributions of the error weight over BSCs is derived. A Hamming weight-based IS algorithm that can estimate the optimal IS distribution and the error probability iteratively without assuming any knowledge of the coding parameters is proposed[1],[2]
3. The asymptotic relative saving on the required sample size of the proposed IS estimator over the state-of-the-art counterpart [26] is derived. It only depends on the error-correcting capability of the code and can be used to predict the efficiency of the estimator before the simulation.

The rest of the paper is organized as follows. Section II briefly reviews some conventional GFBT-based bounds and some preliminaries of the IS method. In Section III, the definition of IDWEF for a linear block code and a given decoder is introduced and based on which, asymptotically tight upper and lower bounds are derived. In Section IV, the optimal IS distribution for error weight is given and a corresponding IS estimator is proposed. The asymptotic relative saving on the required sample size of the proposed IS estimator over the benchmark IS estimator [26] is derived in Section V. A comparison of the MC estimator, the benchmark IS estimator and the proposed IS estimator is presented in Section VI. Finally, some concluding remarks are given in Section VII.

II. PRELIMINARIES

For an (n, k) binary linear block code, a length-k information bit sequence is encoded into a length-n codeword $\mathbf{c} \in \mathcal{C} \subset \mathcal{X}^n \triangleq \{0, 1\}^n$, where $\mathcal{C}$ denotes an k-dimensional subspace (i.e., the code) of the n-dimensional Hamming space $\mathcal{X}^n$. After the codeword is transmitted over the BSC with cross-over probability p, a vector $\mathbf{z} \in \mathcal{X}^n$ is received. Since the code is linear and the channel is symmetric, without loss of generality, assume that the all-zero codeword $\mathbf{c}_0$ is transmitted. Then the received vector $\mathbf{z}$ can also be regarded as the error vector. Denote $\mathrm{wt}(\mathbf{z})$ as the Hamming weight of $\mathbf{z}$. Hence, the error vector $\mathbf{z}$ follows the multivariate Bernoulli distribution $\mathrm{Bern}(\mathbf{z}; p)$ with the probability mass function (p.m.f.) $f(\mathbf{z}) = p^{\mathrm{wt}(\mathbf{z})}(1-p)^{n-\mathrm{wt}(\mathbf{z})}$.

The word error rate (WER) P_e and the bit error rate (BER) P_b are widely used metrics to measure the error performance of a block code and the associated decoder. To avoid possible confusion between the similar notations for WER and BER, subscripts e and b are used to differentiate them. Denote $\mathcal{E}$ as the error region of the all-zero codeword and $I_e(\mathbf{z})$ as the indicator function that equals 1 when $\mathbf{z}$ leads to a decoding error and 0 otherwise. The WER can be expressed in terms of $I_e(\mathbf{z})$ as

$$P_e = \Pr(\mathbf{z} \in \mathcal{E}) = \sum_{\mathbf{z}\in\mathcal{X}^n} I_e(\mathbf{z}) f(\mathbf{z}), \tag{1}$$

Denote $I_b(\mathbf{z})$ as the ratio of the number of the non-zero information bits to k. The BER is defined as

$$P_b = \sum_{\mathbf{z}\in\mathcal{X}^n} I_b(\mathbf{z}) f(\mathbf{z}). \tag{2}$$

A. ERROR PROBABILITY BOUNDS FOR BSC

Error probability bounds are widely used to evaluate the MLD performance of a binary linear block code. One thing that plays an essential role in the derivation of many conventional bounds is the IOWEF [6]. It is defined as

$$A(X, Y) = \sum_{h,d} A_{h,d} X^h Y^d, \tag{3}$$

where X and Y are the input and output indeterminates, respectively, and $A_{h,d}$ represents the number of weight-d codewords generated from weight-h information bit sequences (i.e., the coefficient of the IOWEF with input weight h and output weight d with $0 \le h \le k$ and $0 \le d \le n$). By setting $X = 1$ in (3) we get the expression of the weight enumerating function (WEF).

For example, denote $\mathbf{Z}$ as a random error vector, the famous union bound can be expressed as

$$P_e \le \sum_{d=1}^{n} A_d \Pr(\mathbf{Z} \in \mathcal{V}_d) = \sum_{d=1}^{n} A_d \sum_{\ell=\lceil d/2 \rceil}^{d} \binom{d}{\ell} p^{\ell}(1-p)^{d-\ell}, \tag{4}$$

where the WEF coefficient $A_d \triangleq \sum_{h=0}^{k} A_{h,d}$, and $\mathcal{V}_d$ denotes the pairwise error region between a weight-d codeword and the all-zero codeword.

As we know, the union bound is loose when p is large. By limiting the usage of the union bound to an introduced Gallager region $\mathcal{R}$ around the transmitted codeword, a tighter bound can be derived by avoiding over-estimating the error

[1]Part of the contribution was presented in the conference version [27]. In this paper, we extend the method to the BER simulation and analyze the efficiency of the proposed IS estimator shown in the last contribution.

[2]We open-source our HW-IS algorithm in [28] and make it user-friendly for any linear block codes.

probability outside the Gallager region. Mathematically, the GFBT-based bound can be derived as

$$\begin{aligned} P_e &= \Pr(\mathbf{Z} \in \mathcal{E} \cap \mathcal{R}) + \Pr(\mathbf{Z} \notin \mathcal{E} \cap \mathcal{R}) \\ &\le \sum_{d=1}^{n} A_d \Pr\left(\mathbf{Z} \in \mathcal{V}_d \cap \mathcal{R}\right) + \Pr(\mathbf{Z} \notin \mathcal{R}). \end{aligned} \tag{5}$$

By choosing the region $\mathcal{R}$ as the Hamming sphere centered at the transmitted codeword and optimizing the radius to tighten the bound, the sphere bound is proposed by Poltyrev [11]. Based on the simpler formulation in [4], it can be expressed as

$$P_e \le \sum_{w=0}^{n} p^w (1-p)^{n-w} \min\left\{ \binom{n}{w}, \sum_{d=1}^{n} A_d B(w,d,n) \right\}, \tag{6}$$

where

$$B(w,d,n) = \sum_{\ell=\lceil d/2 \rceil}^{\min\{w,d\}} \binom{d}{\ell} \binom{n-d}{w-\ell} \tag{7}$$

represents the number of weight-w error vectors at the same distance or closer to a weight-d codeword compared with $\mathbf{c}_0$.

On the other hand, by assuming that a decoding error occurs only if the error vector falls inside any packing sphere centered at a codeword other than $\mathbf{c}_0$ [29, eqn. (20)], the union-of-packing-spheres (UPS) lower bound can be derived as

$$P_e \ge \sum_{w=0}^{n} p^w (1-p)^{n-w} \sum_{d=1}^{n} A_d \tilde{B}(w,d,n), \tag{8}$$

where

$$\tilde{B}(w,d,n) = \sum_{\ell=\lceil \frac{d+w-t}{2} \rceil}^{\min\{w,d\}} \binom{d}{\ell} \binom{n-d}{w-\ell} \tag{9}$$

is the number of weight-w vectors within the decoding sphere of a weight-d codeword and t is the error-correcting capability of the code.

It is noteworthy that if the WEF coefficient A_d in (5), (4) and (8) are replaced by $\sum_{h=1}^{k} \frac{h}{k} A_{h,d}$, the GFBT-based bound, the classical union bound and the UPS bound for BER are obtained.

B. IMPORTANCE SAMPLING AND RELATIVE ERROR

The MC estimator of the WER is unbiased and can be written as

$$\hat{P}_e^{\mathrm{MC}} = \frac{1}{N} \sum_{i=1}^{N} I_e(\mathbf{z}_i), \quad \mathbf{z}_i \sim f(\mathbf{z}), \tag{10}$$

where N is the total number of samples. Variance is commonly used as the reliability measure of the estimator. It is well-known that the variance of the MC estimator is

$$\mathrm{Var}\left[\hat{P}_e^{\mathrm{MC}}\right] = \frac{P_e(1-P_e)}{N}. \tag{11}$$

In IS simulation, the WER in (1) can be rewritten as

$$P_e = \sum_{\mathbf{z} \in \mathcal{X}^n} I_e(\mathbf{z}) \frac{f(\mathbf{z})}{f^*(\mathbf{z})} f^*(\mathbf{z}), \tag{12}$$

where the p.m.f. $f^*(\mathbf{z})$ represents the IS distribution. The IS estimator can be expressed as

$$\hat{P}_e^{\mathrm{IS}} = \frac{1}{N} \sum_{i=1}^{N} I_e(\mathbf{z}_i) \frac{f(\mathbf{z}_i)}{f^*(\mathbf{z}_i)}, \quad \mathbf{z}_i \sim f^*(\mathbf{z}). \tag{13}$$

The IS estimator is also unbiased and its variance can be derived as

$$\mathrm{Var}\left[\hat{P}_e^{\mathrm{IS}}\right] = \frac{1}{N} \left(\sum_{\mathbf{z} \in \mathcal{X}^n} I_e(\mathbf{z}) \frac{f^2(\mathbf{z})}{f^*(\mathbf{z})} - P_e^2 \right). \tag{14}$$

Relative error of the estimator is commonly used as the stopping criterion of the simulation, which is defined as [30]

$$\kappa \triangleq \frac{\sqrt{\mathrm{Var}\left[\hat{P}_e^{\mathrm{IS}}\right]}}{P_e}. \tag{15}$$

Specifically, if the IS distribution is same as the original sampling distribution, the IS estimator is equivalent to the MC estimator. When the error probability is small (i.e. $P_e \ll 1$), the relative error in the MC estimator is

$$\kappa = \sqrt{\frac{1-P_e}{NP_e}}.$$

The number of samples needed to achieve a given κ can be approximated as

$$N \approx \frac{1}{\kappa^2 P_e}, \tag{16}$$

which suggests that $N \approx 100/P_e$ samples are required in order to obtain a reliable estimation result with a relative error of 10%, i.e., $\kappa = 0.1$.

The BER counterparts of P_e, $\hat{P}_e^{\mathrm{MC}}$ and $\hat{P}_e^{\mathrm{IS}}$, denoted by P_b, $\hat{P}_b^{\mathrm{MC}}$ and $\hat{P}_b^{\mathrm{IS}}$, can be obtained by substituting I_b for I_e in (12), (10) and (13), respectively.

III. PROPOSED ASYMPTOTICALLY TIGHT BOUNDS

Many bounds in the literature have been derived to predict the error performance of the linear block codes for BSCs. Contrary to the conventional wisdom, we found that they are not asymptotically tight, as p tends to 0. For instance, as shown in Fig. 2 in Section VI, both the union bound and the sphere bound have a non-vanishing gap from the simulated WER performance for the example code considered therein. Consistent with the terminology in the literature, a WER bound $P_{e,\mathrm{bound}}$ is said to be *asymptotically tight* if

$$\lim_{p \to 0} \frac{P_{e,\mathrm{bound}}}{P_e} = 1$$

holds for *all* linear block codes. In other words, as long as there exists a code for which the WER bound cannot give the asymptotic exact performance, we shall not call it

an asymptotically tight bound in this work. The asymptotic tightness of a BER bound is defined in a similar manner. The following theorem formalizes our claim for the GFBT-based WER bounds.

Theorem 1: The WER bound (5) based on Gallager's first bounding technique for binary linear block codes over BSCs is not asymptotically tight for all possible choices of the Gallager region, as the cross-over probability p tends to 0.

Proof: Since only the asymptotic error probability is relevant, it is sufficient to only consider the weight-$(t+1)$ Hamming shell denoted by $\mathcal{Y}_{t+1} = \{\mathbf{z} \in \mathcal{X}^n : \mathrm{wt}(\mathbf{z}) = t+1\}$.

Denote the constrained pairwise error region $V_i = \{\mathbf{z} \in \mathcal{Y}_{t+1} : \mathrm{wt}(\mathbf{z}) \geq \mathrm{wt}(\mathbf{z} - \mathbf{c}^i)\}$ of the i-th codeword $\mathbf{c}^i$, $i = 1, 2, \ldots, m$, as the set of weight-$(t+1)$ vectors which are at the same distance or closer to $\mathbf{c}^i$ compared with $\mathbf{c}_0$. As $p \to 0$, a weight-$(t+1)$ vector may only be wrongly decoded to a weight-$(2t+1)$ or weight-$(2t+2)$ codeword. Let $m = A_{2t+1} + A_{2t+2}$ and index these codewords from 1 to m in a certain order. Obviously, the number of weight-$(t+1)$ vectors that are decoded wrongly is $|\bigcup_{i=1}^{m} V_i|$, where $|\cdot|$ represents the cardinality of a set. Therefore, the asymptotic WER is

$$P_e \sim \left|\bigcup_{i=1}^{m} V_i\right| p^{t+1}(1-p)^{n-t-1} \quad (p \to 0).$$

It follows that a WER bound is asymptotically tight if and only if it converges to the above value as p tends to 0.

Define $\mathcal{R}_{t+1} \triangleq \mathcal{R} \cap \mathcal{Y}_{t+1}$ and $\mathcal{R}'_{t+1} \triangleq \mathcal{R}^{\complement} \cap \mathcal{Y}_{t+1}$. As p tends to 0, the GFBT-based bound (5) can be rewritten as

$$\begin{aligned} P_{e,\mathrm{GFBT}} &\sim \sum_{d=2t+1}^{2t+2} A_d \Pr\left(\mathbf{z} \in \mathcal{V}_d \cap \mathcal{R}_{t+1}\right) + \Pr(\mathbf{z} \in \mathcal{R}'_{t+1}) \\ &= \sum_{i=1}^{m} \Pr\left(\mathbf{z} \in V_i \cap \mathcal{R}_{t+1}\right) + \Pr(\mathbf{z} \in \mathcal{R}'_{t+1}) \\ &= \left(\sum_{i=1}^{m} |V_i \cap \mathcal{R}_{t+1}| + |\mathcal{R}'_{t+1}|\right) p^{t+1}(1-p)^{n-t-1} \\ &\quad \times (p \to 0). \end{aligned}$$

On the other hand, we have

$$\begin{aligned} \left|\bigcup_{i=1}^{m} V_i\right| &= \left|\bigcup_{i=1}^{m} V_i \cap \mathcal{R}_{t+1}\right| + \left|\bigcup_{i=1}^{m} V_i \cap \mathcal{R}'_{t+1}\right| \\ &\leq \left|\bigcup_{i=1}^{m} V_i \cap \mathcal{R}_{t+1}\right| + |\mathcal{R}'_{t+1}| \\ &\leq \sum_{i=1}^{m} |V_i \cap \mathcal{R}_{t+1}| + |\mathcal{R}'_{t+1}|, \end{aligned}$$

where the second last equality holds if and only if $\mathcal{R}'_{t+1}$ is a subset of $\bigcup_{i=1}^{m} V_i$, i.e., the Gallager region $\mathcal{R}$ contains all weight-$(t+1)$ vectors that can be decoded correctly. In order to make the bound asymptotically tight, assume the chosen Gallager region meets this condition. The last equality holds if and only if the sets $V_i \cap \mathcal{R}_{t+1}$, $i = 1, 2, \cdots, m$, are mutually disjoint. According to the inclusion-exclusion principle, we have

$$\begin{aligned} \left|\bigcup_{i=1}^{m} V_i \cap \mathcal{R}_{t+1}\right| &= \sum_{i=1}^{m} |V_i \cap \mathcal{R}_{t+1}| \\ &\quad - \sum_{1 \leq i < j \leq m} |V_i \cap V_j \cap \mathcal{R}_{t+1}| \\ &\quad + \cdots + (-1)^{m-1} |V_i \cap \cdots \cap V_m \cap \mathcal{R}_{t+1}|. \end{aligned}$$

For general binary linear block codes, V_i's are clearly not disjoint. It is possible that the sets $V_i \cap \mathcal{R}_{t+1}$, $i = 1, 2, \cdots, m$, are disjoint with a suitable choice of the Gallager region, i.e., $V_i \cap V_j \subset \mathcal{R}'_{t+1}$ for all i, j. However, this requires the knowledge of the intersections of V_i and V_j, which involves the decoding error relationships among three codewords (including the all-zero codeword). This suggests that the conventional pairwise error weight distribution represented by the WEF coefficients A_d is insufficient for expressing an asymptotically tight WER bound for BSCs because, except the first term on the right-hand side, the other terms involve the error relationships among two or more non-zero codewords. Obviously, the issue remains for all possible choices of the Gallager region $\mathcal{R}$. Therefore, in general, there exists a non-vanishing gap between the GFBT-based bound and the asymptotic WER as p tends to 0. This completes the proof. □

TABLE 1. The codebook of a (6,2) example linear block code.

Message	Codeword
0 0	0 0 0 0 0 0
0 1	0 1 1 0 1 0
1 0	1 0 1 0 0 1
1 1	1 1 0 0 1 1

Take the (6, 2) binary linear block code, whose codebook is shown in Table 1, as an example. The error-correcting capability is $t = 1$. The WEF coefficients of the code are $A_0 = 1$, $A_3 = 2$, $A_4 = 1$. The total number of weight-$(t+1)$ vectors is $\binom{6}{2} = 15$. If we index the three non-all-zero codewords in Table 1 from top to bottom, the sets V_1, V_2 and V_3 can be written as

- $V_1 = \{[0\ 1\ 1\ 0\ 0\ 0], [0\ 1\ 0\ 0\ 1\ 0], [0\ 0\ 1\ 0\ 1\ 0]\}$;
- $V_2 = \{[1\ 0\ 1\ 0\ 0\ 0], [1\ 0\ 0\ 0\ 0\ 1], [0\ 0\ 1\ 0\ 0\ 1]\}$;
- $V_3 = \{[1\ 1\ 0\ 0\ 0\ 0], [1\ 0\ 0\ 0\ 1\ 0], [1\ 0\ 0\ 0\ 0\ 1], [0\ 1\ 0\ 0\ 1\ 0], [0\ 1\ 0\ 0\ 0\ 1], [0\ 0\ 0\ 0\ 1\ 1]\}$.

We can see that the vector [0 1 0 0 1 0] appears in both V_1 and V_3, and the vector [1 0 0 0 0 1] appears in both V_2 and V_3. The exact number of the wrongly decoded weight-$(t+1)$ vectors is $|V_1 \cup V_2 \cup V_3| = 10$. However, based only on the knowledge of WEF coefficients, the number of the wrongly decoded weight-$(t+1)$ vectors is overestimated as $A_3\binom{3}{1} + A_4\binom{4}{2} = 12 = |V_1| + |V_2| + |V_3|$ (as in the classical union bound and sphere bound). Note that the GFBT-based bound is asymptotically tight for this code by choosing the Gallager region to exclude the set of vectors [0 1 0 0 1 0] and

[1 0 0 0 0 1], which can be viewed an additional code-specific parameter that cannot be deduced from the WEF.

Although the GFBT-based WER bounds are not asymptotically tight for all codes, they can be tight for some specific codes. One example is that if the code satisfies $|\bigcup_{i=1}^{m} V_i| = \binom{n}{t+1}$ (i.e., all the weight-$(t+1)$ vectors will be decoded wrongly, such as the perfect code), the bounds can be asymptotically tight. Another example is that if all the sets V_i of the code are pairwise disjoint, the WER bounds can also be asymptotically tight. In particular, all V_i's satisfy the pairwise disjoint condition if $A_{2t+2} = 0$.

Note that some weight-$(t+1)$ vectors are wrongly decoded although they do not fall in the radius-t decoding spheres of any codewords. Thus, the UPS lower bound is not asymptotically tight.

The arguments behind the proof of Theorem 1 can be readily extended to assert that the GFBT-based BER bound for BSCs is not asymptotically tight for all possible choices of the Gallager region. Specifically, it can be deduced that the pairwise error weight distribution represented by the IOWEF is insufficient for representing an asymptotically tight BER bound for BSCs.

By far, we can see that in order to get an asymptotically tight bound for the WER, we need to get the exact number of weight-$(t+1)$ vectors that lead to an error. Furthermore, if the bounds for the BER are considered, we also need to know the weights of the information sequences that the weight-$(t+1)$ vectors wrongly decoded to. Therefore, we introduce the IDWEF defined as

$$S(D, Y) \triangleq \sum_{w=0}^{n} \sum_{h=0}^{k} S_{h,w} D^h Y^w, \tag{17}$$

where D and Y are the output and input indeterminates of the decoder, respectively, and $S_{h,w}$ is the number of weight-w vectors that will be decoded to an information sequence with weight h. It actually can be regarded as the IOWEF of the decoder as it focuses on the relationship between the weights of the error vectors and the information sequences. Consequently, the IDWEF is decoding algorithm dependent. Note that the sphere partitioning function defined in [29] is a special case of our IDWEF. It only shows the first term of the IDWEF and does not consider the weight of the information sequence. For a specific hard-decision decoder, the following proposition provides the exact WER and BER of a given code.

Proposition 1: Given the IDWEF (i.e., $S_{h,w}$ $\forall h, w$) of a binary linear block code and its decoder over BSCs,

$$P_e = \sum_{w=1}^{n} p^w (1-p)^{n-w} \sum_{h=1}^{k} S_{h,w} \tag{18}$$

$$P_b = \sum_{w=1}^{n} p^w (1-p)^{n-w} \sum_{h=1}^{k} \frac{h}{k} S_{h,w} \tag{19}$$

give the exact WER and BER expressions, respectively.

Unfortunately, the decision regions of a decoder are usually too complex for us to obtain the full details of the IDWEF. It is reasonable to use a truncated IDWEF and the existing bounds in order to acquire asymptotically tight bounds of the WER and BER. By replacing the $w = t+1$ term of the sphere bound (6) and UPS bound (8), we can get the proposed asymptotically tight upper and lower bounds for WER.

Proposition 2: Given the weight-$(t+1)$ IDWEF coefficients (i.e., $S_{h,t+1}$ $\forall h$) of a binary linear block code and its decoder,

$$P_e \le p^{t+1}(1-p)^{n-t-1} \sum_{h=1}^{k} S_{h,t+1} + \sum_{w=t+2}^{n} p^w (1-p)^{n-w} \cdot \min\left\{ \binom{n}{w}, \sum_{d=1}^{n} A_d B(w, d, n) \right\}, \tag{20}$$

$$P_e \ge p^{t+1}(1-p)^{n-t-1} \sum_{h=1}^{k} S_{h,t+1} + \sum_{w=t+2}^{n} p^w (1-p)^{n-w} \sum_{d=1}^{n} A_d \tilde{B}(w, d, n) \tag{21}$$

give asymptotically tight upper and lower bounds on WER for BSCs, respectively.

Similar modifications can be applied to the BER case, where we replace the first term of the union bound (4) and the UPS bound (8).

Proposition 3: Given the weight-$(t+1)$ IDWEF coefficients (i.e., $S_{h,t+1}$ $\forall h$) of a binary linear block code and its decoder,

$$P_b \le p^{t+1}(1-p)^{n-t-1} \sum_{h=1}^{k} \frac{h}{k} S_{h,t+1} + \sum_{w=t+2}^{n} p^w (1-p)^{n-w} \sum_{d=1}^{n} \sum_{h=0}^{k} \frac{h}{k} A_{h,d} B(w, d, n) \tag{22}$$

$$P_b \ge p^{t+1}(1-p)^{n-t-1} \sum_{h=1}^{k} \frac{h}{k} S_{h,t+1} + \sum_{w=t+2}^{n} p^w (1-p)^{n-w} \sum_{d=1}^{n} \sum_{h=0}^{k} \frac{h}{k} A_{h,d} \tilde{B}(w, d, n) \tag{23}$$

give asymptotically tight upper and lower bounds on BER for BSCs, respectively.

It is natural to think that the more terms in (4), (6) and (8) are replaced by those in (18) and (19), the tighter the final bounds can achieve. As the number of weight-w vectors increases dramatically w.r.t. the error-correcting capability of t (i.e., the value $\binom{w}{t+1}$ is huge for large t), the cost of the brute-force is already unaffordable with some short block-length codes. Efficient ways to acquire the IDWEF remains open. The simulation is still a practical way to estimate the error performance of long codes. The relative saving on the sample size that required for reliable estimation can be further improved if some fast simulation algorithms are applied.

IV. PROPOSED IMPORTANCE SAMPLING ESTIMATOR

In this section, we propose an IS estimator for error probability evaluation of linear block codes over BSCs with cross-over probability p. A corresponding Hamming weight-based IS algorithm is then presented.

For a given linear block code with blocklength n, the complexity of searching for a proper n-dimensional IS distribution is high. We define $W \triangleq \mathrm{wt}(\mathbf{Z})$ as a random variable, named as the error weight. It follows the binomial distribution $\mathrm{Bin}(w; p)$ with p.m.f. $P_w = \binom{n}{w} p^w (1-p)^{n-w}$ and its sample space is the one-dimensional Hamming weight space $\{0, 1, 2 \cdots, n\}$. We search for the optimal one-dimensional IS distribution in the Hamming weight space that minimized the variance of the IS estimator. The complexity of the searching process can be significantly reduced while the most sensitive dimension of the randomness to the cross-over probability remains.

A. OPTIMAL IS DISTRIBUTION

Define the error probability conditioned on the weight-w vectors as

$$\theta_e(w) \triangleq \Pr\left(I_e(\mathbf{Z}) = 1 | \mathrm{wt}(\mathbf{Z}) = w\right), \tag{24}$$

which is named as the error ratio. Since $\theta_e(w)$ is independent from p, assigning a larger sample size to the error weight that contributes the most to the accuracy of the estimation helps accelerate the simulation process. Assume P_w^* for $w = 0, 1, \cdots, n$ is the applied IS distribution in the Hamming weight space. The corresponding IS distribution over BSCs with cross-over probability p can be derived as

$$f^*(\mathbf{z}) = \frac{f(\mathbf{z})}{P_{\mathrm{wt}(\mathbf{z})}} \cdot P^*_{\mathrm{wt}(\mathbf{z})} = \frac{P^*_{\mathrm{wt}(\mathbf{z})}}{\binom{n}{\mathrm{wt}(\mathbf{z})}}. \tag{25}$$

The physical meaning of the IS distribution (25) can be illustrated as that we need to adjust the occurrence probability for each weight from a binomial distribution to an optimized $\{P_w^*\}_{w=0}^n$ while keeping the distribution conditioned on each weight unchanged. Therefore, the IS estimator for WER in (13) specializes to

$$\hat{P}_e^{\mathrm{IS}} = \frac{1}{N}\sum_{i=1}^{N} I_e(\mathbf{z}_i)\frac{f(\mathbf{z}_i)}{f^*(\mathbf{z}_i)} = \frac{1}{N}\sum_{i=1}^{N} I_e(\mathbf{z}_i)\frac{P_{\mathrm{wt}(\mathbf{z}_i)}}{P^*_{\mathrm{wt}(\mathbf{z}_i)}}, \tag{26}$$

where $\mathbf{z}_i$ are generated from $f^*(\mathbf{z})$.

By substituting the IS distribution (25) into (14), we can derive the expression of the variance of the IS estimator as

$$\begin{aligned}
\mathrm{Var}\left[\hat{P}_e^{\mathrm{IS}}\right] &= \frac{1}{N}\left(\sum_{\mathbf{z}\in\mathcal{X}^n} I_e(\mathbf{z})\frac{f^2(\mathbf{z})}{f^*(\mathbf{z})} - P_e^2\right) \\
&= \frac{1}{N}\left(\sum_{w=0}^{n}\binom{n}{w}\sum_{\mathbf{z}\in\mathcal{Y}_w} I_e(\mathbf{z})\frac{f^2(\mathbf{z})}{f^*(\mathbf{z})} - P_e^2\right) \\
&= \frac{1}{N}\left(\sum_{w=0}^{n}\theta_e(w)\frac{P_w^2}{P_w^*} - P_e^2\right),
\end{aligned} \tag{27}$$

where $\mathcal{Y}_w = \{\mathbf{z} \in \mathcal{X}^n : \mathrm{wt}(\mathbf{z}) = w\}$ is the weight-w Hamming shell. One remark is that the variance expression in Lemma 1 of benchmark paper [26] is not rigorous.

The following theorem provides the general expression for the p.m.f. $\{P_w^*\}_{w=0}^n$ that minimizes the variance of the proposed IS estimator (i.e., the sample size required to achieve a specific relative error is minimized).

Theorem 2: The optimal p.m.f. $\{P_w^*\}_{w=0}^n$ on the Hamming weight space that minimizes the variance of the proposed IS estimator (27) is given by

$$P_w^\star = \frac{\sqrt{\theta_e(w)}P_w}{\sum_{j=0}^{n}\sqrt{\theta_e(j)}P_j}, \quad \text{for } w = 0, 1, \ldots, n. \tag{28}$$

Proof: Since P_w^*, for $w = 0, 1, \cdots, n$, are probabilities and only involved in the first term of the variance (27), the minimization problem can be formulated as

$$\begin{aligned}
\underset{P_0^*,\cdots,P_n^*}{\text{minimize}} \quad & \sum_{w=0}^{n}\frac{\theta_e(w)P_w^2}{P_w^*} \\
\text{s.t.} \quad & \sum_{w=0}^{n} P_w^* = 1 \\
& 0 \le P_w^* \le 1, \ \text{for } w = 0, 1, \ldots, n.
\end{aligned} \tag{29}$$

Denote $J(P_0^*, P_2^*, \cdots, P_n^*) = \sum_{w=0}^{n}\frac{C_w}{P_w^*}$ as the objective function, where $C_w = \theta_e(w)P_w^2$. The Hessian of J is

$$\nabla^2 J = \mathrm{diag}\left(\frac{2C_0}{P_0^{*3}}, \frac{2C_1}{P_1^{*3}}, \ldots, \frac{2C_n}{P_n^{*3}}\right), \tag{30}$$

where $\mathrm{diag}(\cdot)$ represents the diagonal matrix.

The Hessian matrix is positive definite $\nabla^2 J \succeq 0, \forall P_w^* \in \mathbb{R}_+^n$, where $\mathbb{R}_+^n$ is the n-dimensional non-negative real space. The feasible set is a subset of $\mathbb{R}_+^n$ and is convex. Furthermore, the objective function is a subset of $\mathbb{R}_+^n$. Since both the objective function and the feasible set are convex on $\mathbb{R}_+^n$, the optimization problem is convex.

Relax the problem by removing all the inequality constraints (i.e., extending the feasible set). The relaxed problem is still convex on $\mathbb{R}_+^n$. Its Lagrangian can be derived as

$$L(P_0^*, \cdots, P_n^*) = \sum_{w=0}^{n}\frac{C_w}{P_w^*} + \lambda\left(\sum_{j=0}^{n} P_j^* - 1\right), \tag{31}$$

where λ denotes the Lagrange multiplier.

The solution of the relaxed problem can be derived by setting the derivative of the Lagrangian to 0

$$\begin{aligned}
&\frac{\partial L}{\partial P_w^*} = -\frac{C_w}{P_w^{*2}} + \lambda = 0 \quad \Rightarrow \sum_{w=0}^{n} P_w^* = \sum_{w=0}^{n}\sqrt{\frac{C_w}{\lambda}} = 1 \\
&\Rightarrow \lambda = \left(\sum_{w=0}^{n}\sqrt{C_w}\right)^2, \ \text{and } P_w^* = \frac{\sqrt{C_w}}{\sum_{j=0}^{n}\sqrt{C_j}}.
\end{aligned}$$

The above solution satisfies $0 \le P_w^* \le 1$, which indicates it falls inside the feasible set of the original problem.

Therefore, the optimal solution for the problem in (29) is

$$P_w^* = \frac{\sqrt{\theta_e(w)}P_w}{\sum_{j=0}^{n}\sqrt{\theta_e(j)}P_j}, \text{ for } w = 0, 1, \ldots, n. \tag{32}$$

□

By substituting (28) into (25), the optimal IS distribution w.r.t. p can be written as

$$f^\star(\mathbf{z}) = \frac{\sqrt{\theta_e(\mathrm{wt}(\mathbf{z}))}P_{\mathrm{wt}(\mathbf{z})}}{\binom{n}{\mathrm{wt}(\mathbf{z})}\sum_{j=0}^{n}\sqrt{\theta_e(j)}P_j}. \tag{33}$$

As we can notice, the optimal solution (28) contains part of the target information itself $\theta_e(w)$, with which the error probability can be straightforwardly calculated. Therefore, it is impossible to obtain the optimal solution in practice. A straightforward way that can avoid the problem is to use the estimated or approximated $\theta_e(w)$, denoted as $\hat{\theta}_e(w)$, to evaluate the p.m.f. $\{P_w^*\}_{w=0}^n$. There exists some knowledge of the codebook like the error-correcting capability t and the IDWEF conditioned on weight $t+1$ that can help to get a good approximation of $\theta_e(w)$ for the asymptotic case.

A naive IS estimator is therefore introduced. Based on either some knowledge of the codebook or a preprocessing with limited computational power, a rough estimation of some dominant terms of $\theta_e(w)$ is assumed. Then, an approximated IS distribution is derived and used for further error probability simulation. Also, since error ratios $\theta_e(w)$ for small weights are important properties for the code, it's worthwhile to do an off-line simulation and tabulate them like the examples shown in [24].

One remark is that the counterparts of θ_e for the BER case θ_b can be obtained by substituting I_b for I_e in (24). All the above results still hold for BER case if these counterparts are replaced in all equations.

B. HAMMING WEIGHT-BASED IS ESTIMATOR

We propose the Hamming weight-based importance sampling (HW-IS) algorithm that is more efficient than the state-of-the-art IS-MC basic algorithm [26] for the fast simulation purpose.

Since the optimal IS distribution requires the information of $\theta_e(w)$ which is also absent at the very beginning, an iterative algorithm is proposed. With the initial guess $\hat{\theta}_0(w)$, the iteration alternates between performing a p.m.f. $\{P_w^*\}_{w=0}^n$ update and a WER or BER estimation. The implementation of the HW-IS algorithm is shown in Algorithm IV-B. Initially, the sample size counter N_tot is set as 0 and the relative error WERre is set as 1. The p.m.f. $\{P_w^*\}_{w=0}^n$ is initialized with the inputs of the cross-over probability and $\hat{\theta}_0(w)$.

The sample generation from the IS distribution $f^*(\mathbf{z})$ consists of two phases. During the first phase, an integer w is randomly generated from the p.m.f. $\{P_w^*\}_{w=0}^n$. In the second phase, a sample $\mathbf{z}$ is generated by randomly permuting the element order of a weight-w length-n binary vector to mimic the uniform distribution of error vectors conditioned on $\mathrm{wt}(\mathbf{z}) = w$. During the iteration, the error ratio $\hat{\theta}_e(w)$ for each Hamming weight is updated by

$$\hat{\theta}_e(w) = \frac{\sum_{i=1}^{N} I_e(\mathbf{z}_i)I_w(\mathbf{z}_i)}{\sum_{i=1}^{N} I_e(\mathbf{z}_i)}, \tag{34}$$

where $I_w(\mathbf{z})$ returns 1 if $\mathrm{wt}(\mathbf{z}) = w$ and 0 otherwise. The p.m.f. $\{P_w^*\}_{w=0}^n$ for the next iteration is updated with (28). The error probability is estimated inside the while loop until the relative error calculated by (15) meets the stopping criterion or the maximum number of iteration is reached.

Algorithm 1 Hamming Weight-Based Is Algorithm for Fast Simulation Over BSCs

Input: Cross-over probability p, initial error ratio $\hat{\theta}_0$ and relative error re
Output: Error probability $\hat{P}_e$ and sample size N_tot
Initialization: N_tot:= 0, WERre:= 1, initialize $\{P_w^*\}_{w=0}^n$ with (28) **while** *WERre* > *re* **do**
  Randomly generate w from the p.m.f. $\{P_w^*\}_{w=0}^n$ Generate $\mathbf{z}$ uniformly conditioned on $\mathrm{wt}(\mathbf{z}) = w$ Pass $\mathbf{c}_0 + \mathbf{z}$ through the decoder **if** *N_tot* > *N_min* **then**
    Compute $\hat{P}_e$ according to (26) Compute WERre according to (15) Update $\hat{\theta}_e$ with (34) Update $\{P_w^*\}_{w=0}^n$ with (28)
  end
  N_tot:= N_tot + 1
end
return $\hat{P}_e$ and N_tot

In order to avoid the violation due to the insufficient number of samples during the first several iterations, we set the minimum sample size N_min needed for both $\hat{P}_e$ and $\{P_w^*\}_{w=0}^n$ to start an update. However, for some small $\theta_e(w)$ terms, it's possible that no error has been found after N_min samples are generated. This may lead to $\hat{\theta}_e(w) = 0 \Rightarrow P_w^* = 0$ after the update, which means no more samples with those weights will be generated afterwards and P_w^* will be frozen to be 0. In order to avoid this frozen distribution parameter problem, we sacrifice part of the efficiency by forcing $P_{w_0-1}^* = \beta P_{w_0}^*$, where w_0 is the first weight with non-zero $\hat{\theta}(w_0)$ by far and β is a factor set heuristically. This will make the algorithm keep searching for weight-(w_0-1) vectors that may contain errors. Apparently, for the cases with the knowledge of t, no efficiency loss happens. While for the cases without t, we need to keep generating samples with weight $w_0 - 1$ until the stopping criterion satisfies. One can regard this as the efficiency loss led by the lack of knowledge of t.

Furthermore, the algorithm can be extended to an SNR invariant version by accurately estimating $\theta_e(w)$ during one simulation. Compared with the SNR-invariant algorithm in [26] the advantage is that even for long codes, not only the error floor part but also the water-falling region can be accurately predicted. Similar to [26], the error-correcting capability t can be estimated according to the result. The algorithm is also applicable for the BER case if all the notations for WER are replaced by those for BER.

V. ASYMPTOTIC RELATIVE SAVING ON THE REQUIRED SAMPLE SIZE

In this section, the efficiency of the proposed IS estimator is analyzed. From the simulation results of the benchmark IS estimator in [26], one can observe that the number of the generated samples becomes saturated at some point in the high SNR region. A similar phenomenon can be found in the results of the proposed estimator as well, which means the IS distribution does not depend on the SNR or p anymore. Since for the asymptotic case, the weight-$(t+1)$ error vectors dominate the performance, once the reliable estimation of the error ratio $\theta_e(t+1)$ is obtained, the error probability result meets the accuracy requirement and no more samples are needed. Above all, it's natural to show the advantage of the proposed estimator by comparing the efficiencies of these two estimators in the asymptotic case.

In order to make a fair comparison, define the relative saving η as the percentage that the proposed estimator can save compared to the benchmark estimator in terms of the sample size under the reliability (i.e., the relative error values are equal).

$$\eta = 1 - \frac{N_{\text{prop}}}{N_{\text{bench}}}, \tag{35}$$

where N_{prop} and N_{bench} are denoted as the sample sizes for the proposed estimator and the benchmark estimator, respectively.

For the asymptotic case where the assumption $np \ll 1$ is usually made, [31] suggests that the following approximation for (27) holds

$$\text{Var}\left[\hat{P}_e^{\text{IS}}\right] \approx \frac{1}{N}\left(\theta_e(t+1)\frac{P_{t+1}^2}{P_{t+1}^*} - P_e^2\right). \tag{36}$$

which indicates that the weight-$(t+1)$ error vectors not only dominates the error probability but also the variance of the estimator. It can be foreseen that the saturated value of the number of samples is only related to the term $\theta_e(t+1)$ in the error ratios of the code. For the proposed estimator, this will make the parameter of the optimal distribution P_{t+1}^* approaches 1 to generated samples on the target weight as many as possible. An asymptotic approximation of the relative saving on the required sample size can be derived.

Theorem 3: Assume that $np \ll 1$. The relative saving η_{MC} (in terms of the sample size) of the proposed IS estimator w.r.t. the MC estimator is

$$\eta_{MC} \approx 1 - \binom{n}{t+1} p^{t+1}(1-p)^{n-t-1}. \tag{37}$$

The relative saving η of the proposed IS estimator w.r.t. the state-of-the-art counterpart in [26] is

$$\eta \approx 1 - \binom{n}{t+1}\left(\frac{t+1}{n}\right)^{t+1}\left(1 - \frac{t+1}{n}\right)^{n-t-1}. \tag{38}$$

Proof: According to the approximation of the variance in (36) and the definition of the relative error in (15), we can derive the total number of samples required for the MC estimator, the benchmark IS estimator [26] and the proposed IS estimator with relative error κ as

$$\begin{aligned}
N_{\text{MC}} &= \frac{1}{\kappa^2 P_e^2}\left(\theta_e(t+1)P_{t+1} - P_e^2\right),\\
N_{\text{bench}} &= \frac{1}{\kappa^2 P_e^2}\left(\frac{\theta_e(t+1)P_{t+1}^2}{\binom{n}{t+1}q^{t+1}(1-q)^{n-t-1}} - P_e^2\right),\\
N_{\text{prop}} &= \frac{1}{\kappa^2 P_e^2}\left(\frac{\theta_e(t+1)P_{t+1}^2}{P_{t+1}^*} - P_e^2\right).
\end{aligned}$$

Under the assumption $np \ll 1$, the approximation $P_e \approx \theta_e(t+1)P_{t+1}$ holds. Therefore, the number of samples needed for the proposed estimator can be written as

$$\begin{aligned}
N_{\text{prop}} &= \frac{P_{t+1}^2\theta_e(t+1)}{\kappa^2 P_e^2 P_{t+1}^*} - \frac{1}{\kappa^2}\\
&\approx \frac{P_{t+1}^2}{(\theta_e(t+1)P_{t+1})^2}\cdot\frac{\theta_e(t+1)}{\kappa^2 P_{t+1}^*} - \frac{1}{\kappa^2}\\
&\approx \frac{1}{\kappa^2 P_{t+1}^*\theta_e(t+1)}.
\end{aligned}$$

where the last approximation is made due to that most of the cases, the error ratio satisfies $\theta_e(t+1) \ll 1$. Similarly, for the MC and the benchmark estimator, the sample sizes can be derived as

$$\begin{aligned}
N_{\text{MC}} &\approx \frac{1}{\kappa^2 P_{t+1}\theta_e(t+1)},\\
N_{\text{bench}} &\approx \frac{1}{\kappa^2\binom{n}{t+1}q^{t+1}(1-q)^{n-t-1}\theta_e(t+1)}.
\end{aligned}$$

As the probability P_{t+1} dominates the tail part $w \geq t+1$ of the p.m.f. P_w under the assumption $np \ll 1$, the approximation $P_{t+1}^* \approx 1$ can be achieved according to (28). Furthermore, from [26], we know that the optimal parameter for the benchmark estimator in the asymptotic case is $q^\star = \frac{t+1}{n}$ for the minimum-variance purpose within the parametric family of Bernoulli distribution.

Hence, the efficiencies that the proposed IS estimator can achieve compared to the MC and the benchmark are

$$\eta_{\text{MC}} \approx 1 - \frac{P_{t+1}}{P_{t+1}^*} \approx 1 - \binom{n}{t+1}p^{t+1}(1-p)^{n-t-1},$$

and

$$\begin{aligned}
\eta &\approx 1 - \frac{\binom{n}{t+1}q^{\star t+1}(1-q^\star)^{n-t-1}}{P_{t+1}^*}\\
&\approx 1 - \binom{n}{t+1}\left(\frac{t+1}{n}\right)^{t+1}\left(1 - \frac{t+1}{n}\right)^{n-t-1},
\end{aligned}$$

respectively. □

As p tends to 0, the saving η_{MC} compared to the MC estimator in (37) approaches 1. We are more interested in η compared to the benchmark estimator [26]. Since most of the block codes satisfy the condition $n \gg t+1$, the influence of the blocklength on η is negligible when n becomes large. This also can be verified through Fig. 1, where the curves become

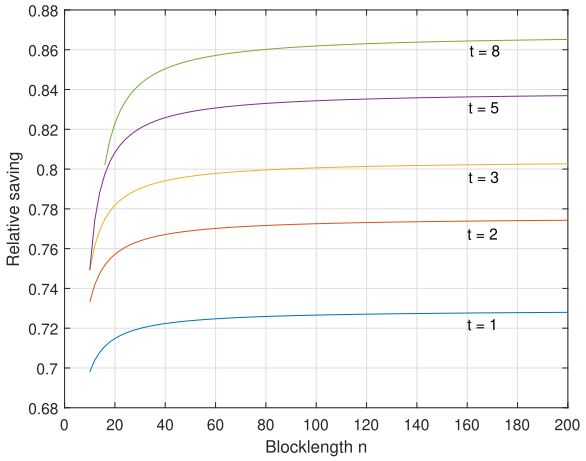


FIGURE 1. Asymptotic relative saving (38) on the required sample size that the proposed IS estimator can achieve compared to the benchmark estimator [26] in terms of the blocklength n with different error-correcting capabilities t.

steady as n increases. Therefore, η can be further simplified by approximation as stated in the following corollary.

Corollary 1: Assume that $np \ll 1$ *and the code satisfies* $n \gg t+1$. *The relative saving* η *of the proposed IS estimator given in Theorem 3 can be further simplified as*

$$\eta \approx \frac{1}{2} + Q\left(\frac{1}{\sqrt{t+1}}\right), \tag{39}$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{x^2}{2}} dx$.

Proof: When n is large, we can use the normal distribution $\mathcal{N}(np, np(1-p))$ to approximate the binomial distribution $\mathcal{B}(n,p)$. Here define a random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ with mean $\mu = t+1$ and variance $\sigma^2 = (t+1)\left(1 - \frac{t+1}{n}\right) \approx t+1$ under the assumption $n \gg t+1$. The following approximation holds

$$\begin{aligned}
&\binom{n}{t+1} q^{\star t+1} \left(1-q^\star\right)^{n-t-1} \\
&\quad \approx \Pr\left(t+1 \le X < t+2\right) \\
&\quad \approx \Pr\left(0 \le \frac{X-\mu}{\sqrt{t+1}} < \frac{1}{\sqrt{t+1}}\right) \\
&\quad = \frac{1}{2} - Q\left(\frac{1}{\sqrt{t+1}}\right).
\end{aligned}$$

□

From the above corollary, we know that the asymptotic relative saving η can be represented as a function of t only and is indepedent of the blocklength n. The larger t the code has, the higher asymptotic η one can achieve. And we will show examples in Section VI that although several approximations are made during the analysis, the asymptotic η predicts the efficiency very well in practice.

Finally, since (39) is irrelevant to the error ratios of the code, both the BER and WER simulations have the same the relative saving on the sample size. There also exist ways to further improve η if more knowledge of the codebook is provided. Since the samples are uniformly drawn within each weight, the sample size needed for a reliable estimation of $\theta_e(w)$ won't reduce no matter what kind of biased distributions is applied. Combining the derived IS distribution with a smarter biased distribution conditioned on each weight instead of the uniform one will further reduce the required sample size. But that may be code specific since the codebook knowledge, as well as the decoding algorithms' mechanism, are required.

TABLE 2. The weight-$(t+1)$ IDWEF coefficients (see (17)) of the (15,7) primitive BCH code with MLD and $t = 2$.

h	0	1	2	3	4	5	6	7
$S_{h,3}$	87	145	130	63	30	0	0	0

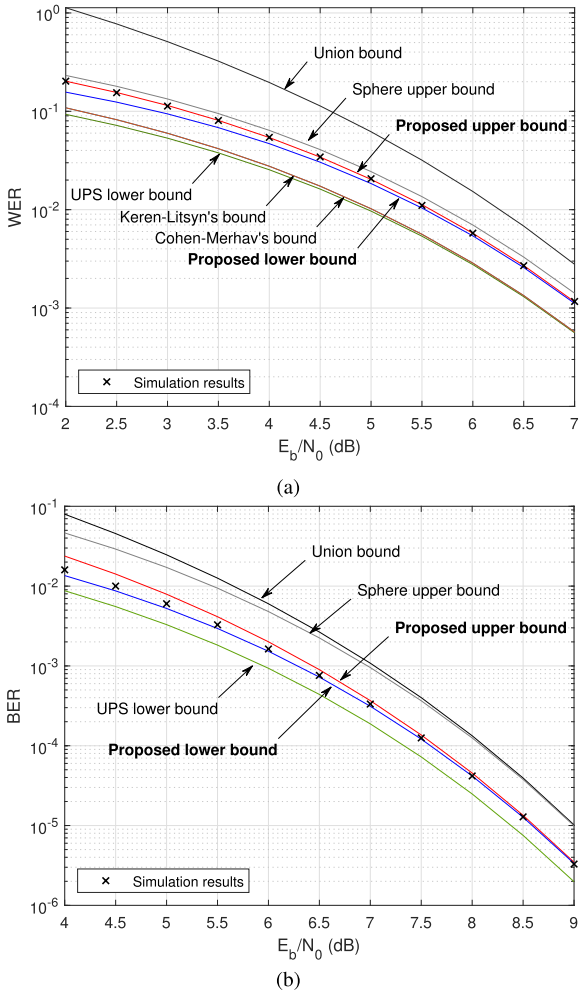
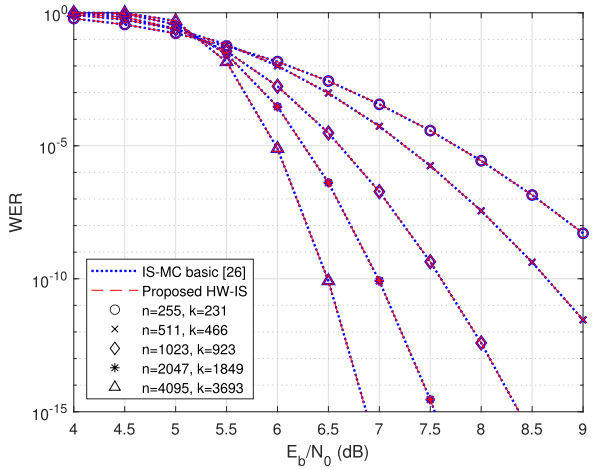


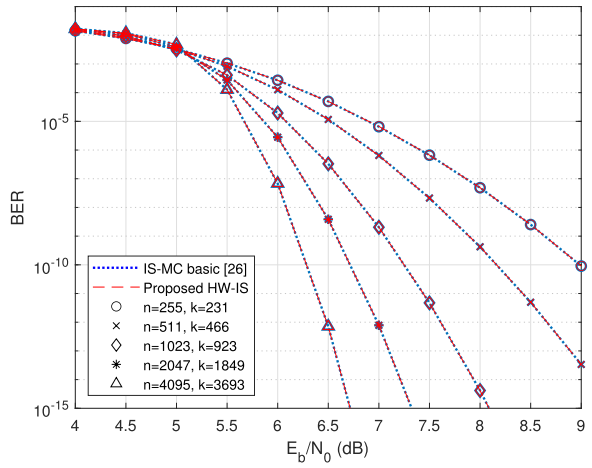
FIGURE 2. A comparison of the proposed upper and lower bounds with the union bound, Poltyrev's sphere bound [10], the UPS bound [29], Keren-Litsyn's bound [32], Cohen-Merhav's bound [33] and the MLD simulation results using the (15,7) primitive BCH code.

VI. NUMERICAL RESULTS AND DISCUSSIONS

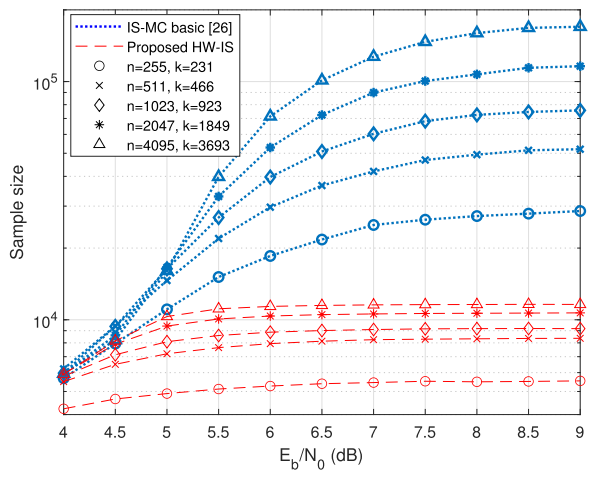
In this section, various bounds mentioned in Section II are firstly compared, and then the simulation results about the



(a)

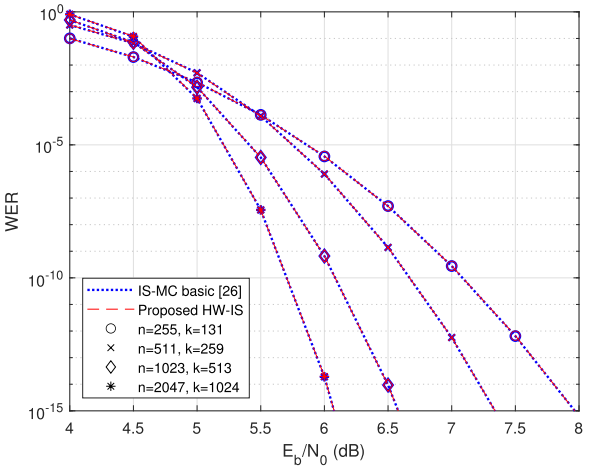


(b)

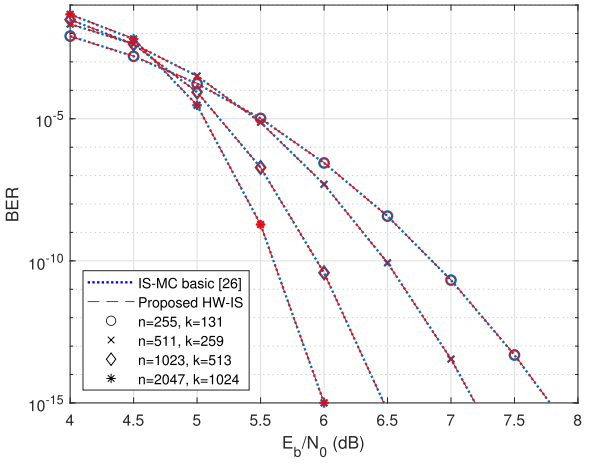


(c)

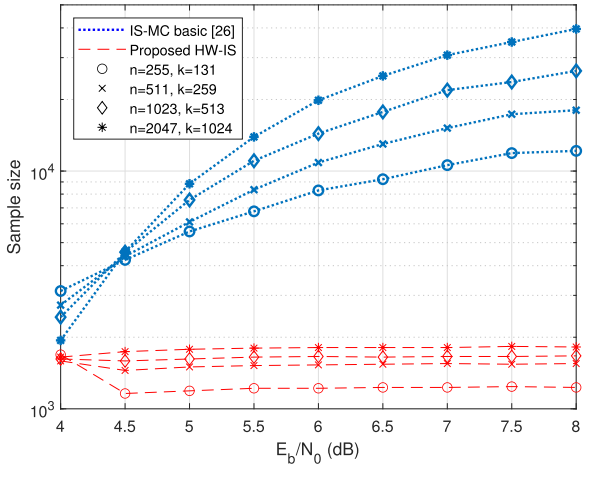
FIGURE 3. Simulation results of BCH codes with rate $R \approx 0.9$ using the proposed HW-IS algorithm and the IS-MC basic algorithm [26].



(a)



(b)



(c)

FIGURE 4. Simulation results of BCH codes with rate $R \approx 0.5$ using the proposed HW-IS algorithm and the IS-MC basic algorithm [26].

sample size needed for reliable estimation of the proposed HW-IS algorithm are shown. The comparisons between the results of the proposed algorithm with those of the state-of-the-art "IS-MC basic" algorithm in [26] are presented. Furthermore, the relative savings on the sample size of the example codes in the high SNR region are listed to show the accuracy of the derived asymptotic η. All the simulation results can be reproduced by our open-source tool [28].

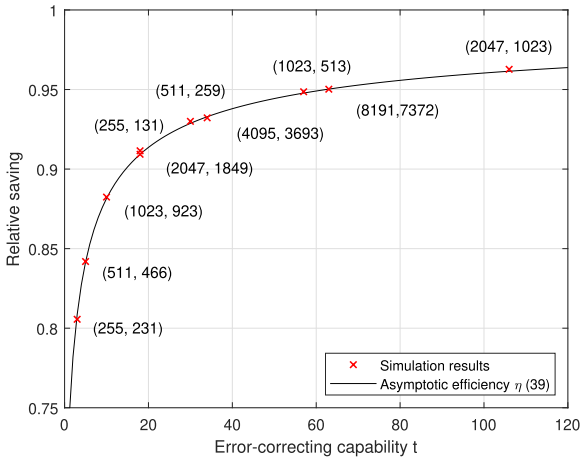


FIGURE 5. A comparison of the asymptotic relative saving in (39) and the corresponding simulation results of the proposed HW-IS algorithm using the BCH codes considered in in Fig. 3 and Fig. 4.

A. COMPARISONS OF BOUNDS

We consider the (15,7) primitive BCH code with MLD and $t = 2$. The weight-$(t+1)$ IDWEF coefficients of the code are obtained by an exhaustive search, and the results are listed in Table 2.

In Fig. 2, we compare the proposed upper and lower bounds in Proposition 2 and 3 with several known bounds in the literature. For reference, the union bound (4), Poltyrev's sphere bound [10], the UPS bound [29] and the MLD simulation results are plotted for both WER and BER in Fig. 2(a) and (b), respectively. In addition, Keren-Litsyn's bound [32, Theorem 1] and Cohen-Merhav's bound [33, Proposition 4.2] on WER are plotted in Fig. 2(a). These two bounds have almost the same performance for the considered code and and are marginally tighter than the UPS bound. It can be seen that all three lower bounds have the same asymptotic performance and are not asymptotically tight. Since there are no BER counterparts derived in [32] and [33], only the UPS lower bound on BER is plotted in Fig. 2(b) for comparison with the proposed lower bound. One can see that the proposed bounds are asymptotically tight in the high SNR region, while the others have non-vanishing gaps from the MLD simulation results. For the union bound and the sphere bound, the gaps are caused by over-counting the number of wrongly decoded weight-$(t+1)$ vectors as we stated in Theorem 1.

Although the bounds are powerful tools for performance evaluation, it requires parameters of the applied encoder-decoder pair (such as the IOWEF or the IDWEF). Determining such parameters may be computationally costly or intractable. Also, MLD is not always applicable for the practical case. For example, people usually decode the BCH codes by the Berlekamp-Massey (BM) algorithm [34], [35] in practice, which is suboptimal. Simulation is a commonly used tool to estimate the error performance for the coded systems. Next, we will show how the proposed IS method can significantly improve the efficiency in terms of the sample size.

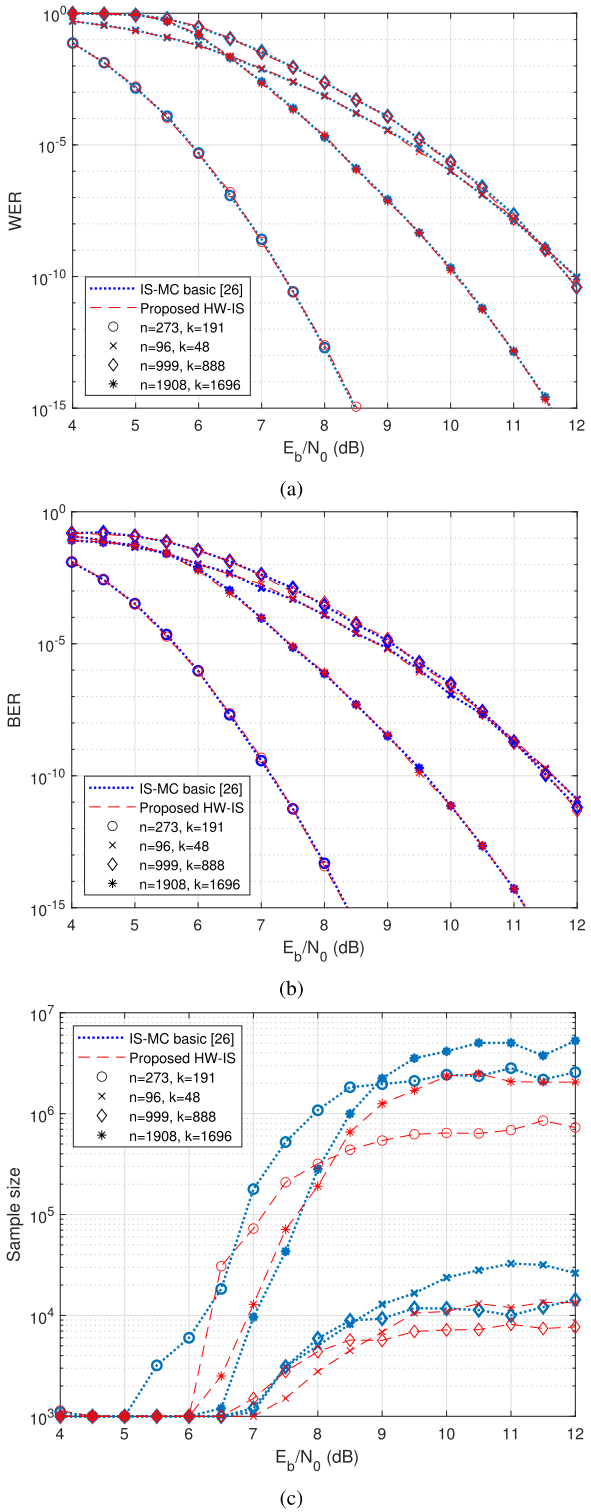


FIGURE 6. Simulation results of some representative LDPC codes using the proposed HW-IS algorithm and the IS-MC basic algorithm [26]. The (273, 191) DSC LDPC code is taken from [36] and the others are MacKay's LDPC codes from [37].

B. COMPARISONS OF SIMULATION RESULTS

According to the HW-IS algorithm in Section IV, it is natural to think that a good initial guess of $\hat{\theta}_0(w)$ will accelerate the

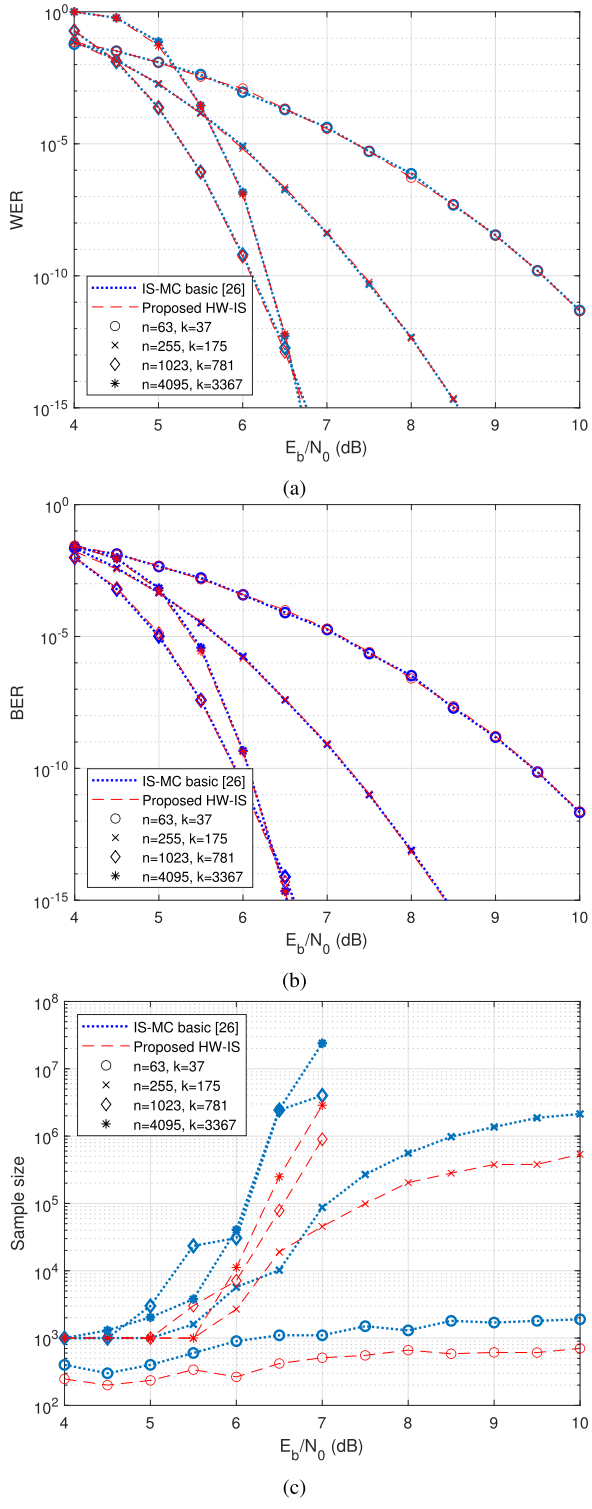


FIGURE 7. Simulation results of EG-LDPC codes [38] using the proposed HW-IS algorithm and the IS-MC basic algorithm [26].

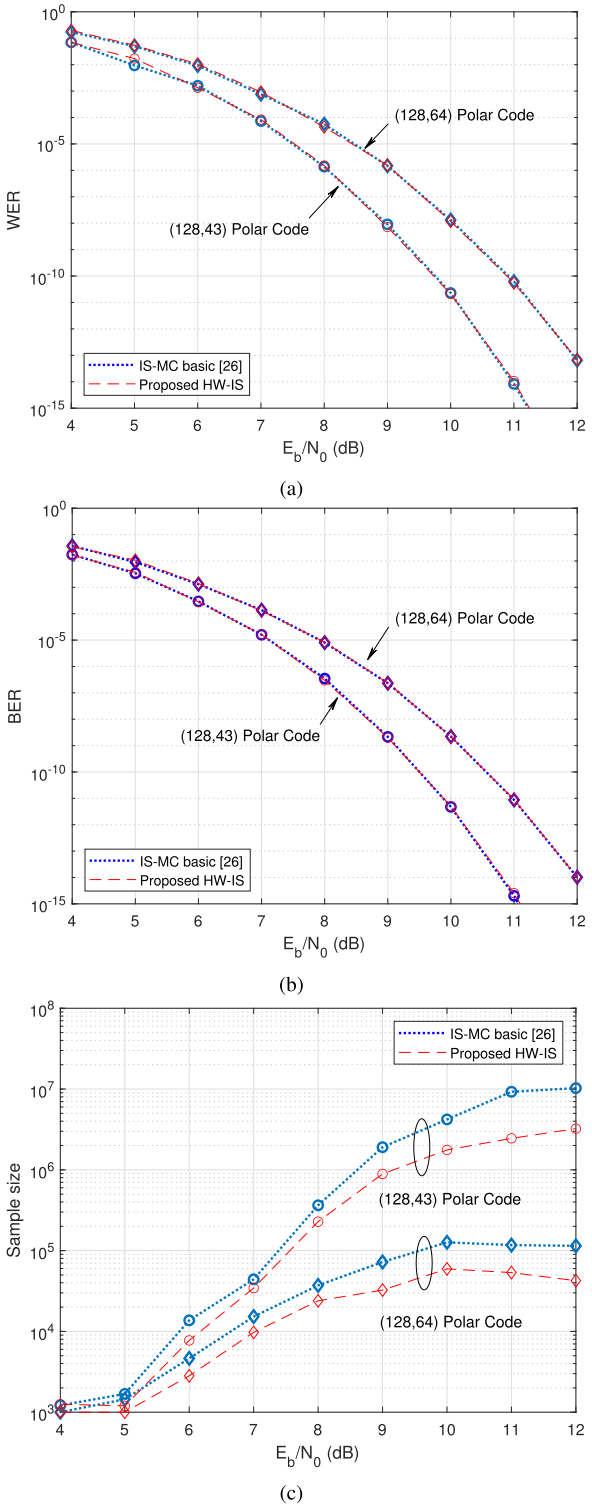


FIGURE 8. Simulation results of Polar codes decoded by the SC decoder [39] using the proposed HW-IS algorithm and the IS-MC basic algorithm [26].

convergence of the optimal IS distribution. The choice of $\hat{\theta}_0(w)$ depends on what kind of knowledge is assumed. In this part, we consider two cases - the one with the knowledge about the error-correcting capability t and the other without any side information. For the former one, we can set $\hat{\theta}_0(w)$ to be that of the worst-case bounded distance decoder (i.e., $\hat{\theta}_0(w_i) = 0$ and $\hat{\theta}_0(w_j) = 1$, for $w_i = 0, 1, \cdots, t$ and $w_j = t+1, 2, \cdots, n$). While the latter one can be regarded

TABLE 3. A comparison of the asymptotic relative saving ("Asymptotic η") in (40) and the simulated relative saving ("Simulated η") of the proposed IS estimator (HW-IS algorithm) w.r.t. the benchmark IS estimator (IS-MC basic algorithm [26]) at WER $\approx 10^{-13}$ using various LDPC and Polar codes. N_{prop} and N_{bench} denote their required sample sizes, respectively.

	(n, k)	WER	N_{bench} [26]	N_{prop}	Simulated η	Asymptotic η
MacKay's LDPC [37]	(96, 48)	1.18×10^{-13}	3.15×10^{4}	1.46×10^{4}	0.5365	0.5637
	(999, 888)	1.13×10^{-13}	1.40×10^{4}	7393	0.4742	0.4795
	(1908, 1696)	1.53×10^{-13}	5.03×10^{6}	2.08×10^{6}	0.5865	0.5637
DSC LDPC [36]	(273, 191)	2.44×10^{-13}	1.08×10^{6}	3.19×10^{5}	0.7046	0.7389
EG-LDPC [38]	(63, 37)	1.13×10^{-13}	1748	583	0.6665	0.6547
	(255, 175)	1.35×10^{-13}	6.69×10^{5}	2.18×10^{5}	0.6741	0.7389
	(1023, 781)	2.64×10^{-14}	4.56×10^{6}	1.07×10^{6}	0.7654	0.8084
	(4095, 3367)	6.16×10^{-13}	2.49×10^{6}	2.46×10^{5}	0.9012	0.8618
Polar codes [39]	(128, 43)	1.07×10^{-13}	7.89×10^{6}	2.23×10^{6}	0.7174	0.7237
	(128, 64)	9.91×10^{-14}	1.15×10^{5}	4.26×10^{4}	0.6296	0.6171

as an uncoded system (i.e., $t = 0$). The stopping criterion of the simulations in this section is set as $\kappa = 0.1$.

We choose some representative BCH codes [31] with rate $R = k/n \approx 0.9$ as examples. The BER results as well as the generated sample size w.r.t. E_b/N_0 are shown in Fig. 3(a) and (b), respectively. All the codes are decoded by the BM algorithm. A sample size $\mathsf{N\_min} = 10^3$ is assured before we start to update the IS distribution.

It can be noticed that all the curves in Fig. 3 (b) increase exponentially and become saturated after reaching some specific E_b/N_0's. The reason is that as SNR increases, the sample size needed keeps growing until the error ratio of the dominant weight-$(t + 1)$ Hamming shell is reliably estimated. The E_b/N_0's for the curves become saturated vary for different codes as they depend on the codebook knowledge and the decoding algorithm. Further increasing the sample size wastes the computational power. We may claim the HW-IS algorithm is SNR-invariant for high SNR.

The BER and sample size vs E_b/N_0 of BCH codes with rate $R \approx 0.5$ using the proposed HW-IS algorithm and the benchmark IS-MC algorithm are shown in Fig. 4(a) and (b), respectively. Similar observations as those with rate $R \approx 0.9$ also hold.

The relative saving η vs the error-correcting capability t for the BCH codes considered in Fig. 3 and Fig. 4 are shown in Fig. 5, where the solid line represents the approximated asymptotic result derived in Corollary 1. All the simulation results are calculated in the high SNR region (i.e., the part that the number of samples becomes saturated). As we can see, the curve for the asymptotic η is achievable and can predict the efficiencies of the HW-IS algorithm pretty well. Also, the trend that a larger saving on the sample size can be obtained for the codes with bigger t is verified according to these simulation results.

Next, we consider the cases with no side information provided and apply the HW-IS algorithm on the LDPC and Polar codes. In order to make a fair comparison, let us start with the same examples used for the IS-MC basic algorithm in [26]. The LDPC codes taken from [36], [37] are implemented. In Fig. 6 (a) and (b), the simulation results of the WER and the sample size are shown, respectively. Since the knowledge of t is absent, the factor $\beta = 1$ is set to keep the estimator unbiased. Subsequently, the asymptotic η will be degraded to

$$\eta \approx 1 - \frac{1+\beta}{2} + (1+\beta)Q\left(\frac{1}{\sqrt{t+1}}\right). \tag{40}$$

All the settings of the algorithm are the same as the cases for the BCH code. The LDPC codes are decoded by the bit-flipping decoder presented in [24] with a maximum iteration number of 20.

Similar to the BCH codes, we can observe that from Fig. 7 (b), the curves of the sample size reach a saturated value after some E_b/N_0 points. This flat region indicates that the weight-$t + 1$ errors dominate the performance and the assumptions for the asymptotic analysis for the error probability hold now. Hence, the estimated $\hat{\theta}_e(w)$ for this region can be used to predict the t of the LDPC codes. Since those saturated values are determined by the value of $\theta_e(t + 1)$, for the code with a thinner weight spectrum, the smaller $\theta_e(t+1)$ it owns, the more samples needed for a reliable estimation in the asymptotic case.

We further implement the HW-IS algorithm on the EG-LDPC codes [38] and the Polar codes. The EG-LDPC codes are decoded by the same decoder as stated in the previous example. Due to the values of t for these codes have been tabulated, we can verify the derived asymptotic efficiencies η with the simulated ones. The Polar codes are decoded with the successive-cancellation algorithm proposed in [39].

One can see by Fig. 7 and 8 that for all of these examples, the proposed HW-IS algorithm can beat the IS-MC basic algorithm in the aspect of the efficiency while keeping the same accuracy of the WER estimation.

Usually, people are more interested in the performance of the LDPC codes over the water-falling region instead of the asymptotic cases. Since in practical applications, there is no need to get the specific position of the error floor as long as it's below the error probability level of interest. Hence, for the purpose of studying the performance of the HW-IS algorithm in the non-asymptotic cases, we apply it on two more practical EG-LDPC codes [38] with parameters (1023, 781) and (4095, 3367). The level of WER $P_e \approx 10^{-13}$ (approximated from the level of required BER $P_b = 10^{-15}$)

is focused on, which is commonly required for the optical communication applications [12].

The sample sizes and the asymptotic efficiencies of all the used LDPC codes and Polar codes compared to the IS-MC basic algorithm are summarized in Table 3.

Due to the efficiency loss caused by the lack of knowledge of t, the asymptotic η's of these codes are smaller than those of the BCH codes under the same t. Though, compared to the large sample size required for the low error probability estimation, the saving is still considerable. By combining the results shown in Table 3 and Fig. 6-8, we can see that the asymptotic η predicts the performance quite well even for $(1023, 781)$ and $(4095, 3367)$ codes, whose curves for the sample size have not reached the saturated region at $P_e = 10^{-13}$. These provide evidences that our method can outperform the IS-MC basic algorithm for the non-asymptotic cases as well.

VII. CONCLUSION

In this paper, the problem of efficiently evaluating the BER and WER of linear block codes over BSCs was studied. Firstly, we showed that any GFBT-based bounds are not asymptotically tight for all possible choices of the Gallager region. By proposing the IDWEF of a coding scheme, the asymptotically tight MLD upper and lower bounds were proposed. Secondly, aiming at accelerating the simulation process for the low BER and WER region, a Hamming weight-based IS estimator was proposed. Its relative saving on the sample size required for a reliable estimation compared with the state-of-the-art IS-MC basic algorithm [26] was investigated. The derived asymptotic η can predict the efficiency of the proposed IS estimator accurately. Our simulation results showed that the proposed IS estimator is more efficient than the benchmark [26] in all cases under consideration. The saving on the sample size ranges from 47% to 97% and increases with the error-correcting capability of the code. As a future work, it is interesting to extend the presented results for BSCs to the counterparts for continuous channels, including but not limited to the Gaussian channel.

REFERENCES

[1] W. H. Tranter, T. S. Rappaport, K. L. Kosbar, and K. S. Shanmugan, *Principles of Communication Systems Simulation With Wireless Applications*. vol. 1, Upper Saddle River, NJ, USA: Prentice-Hall, 2004.

[2] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems: Modeling, Methodology and Techniques*. Cham, Switzerland: Springer, 2006.

[3] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[4] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, Apr. 2010.

[5] I. Sason and S. Shamai (Shitz), "Improved upper bounds on the ML decoding error probability of parallel and serial concatenated turbo codes via their ensemble distance spectrum," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 24–47, Jan. 2000.

[6] I. Sason and S. Shamai (Shitz), "Performance analysis of linear codes under maximum-likelihood decoding: A tutorial," *Found. Trends Commun. Inf. Theory*, vol. 3, pp. 1–222, Jul. 2006.

[7] E. Agrell, "Voronoi regions for binary linear block codes," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 310–316, Jan. 1996.

[8] D. Divsalar, "A simple tight bound on error probability of block codes with application to turbo codes," *Telecommun. Mission Oper. Prog. Rep.*, vol. 19, pp. 42–139, Jul. 1999.

[9] E. R. Berlekamp, "The technology of error-correcting codes," *Proc. IEEE*, vol. 68, no. 5, pp. 564–593, May 1980.

[10] H. Herzberg and G. Poltyrev, "Techniques of bounding the probability of decoding error for block coded modulation structures," *IEEE Trans. Inf. Theory*, vol. 40, no. 3, pp. 903–911, May 1994.

[11] G. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectra," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1284–1292, Jul. 1994.

[12] G. Tzimpragos, C. Kachris, I. B. Djordjevic, M. Cvijetic, D. Soudris, and I. Tomkos, "A survey on FEC codes for 100 G and beyond optical networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 209–221, 1st Quart., 2016.

[13] S. Ghosh and P. D. Lincoln, "Dynamic LDPC codes for nanoscale memory with varying fault arrival rates," in *Proc. 6th Int. Conf. Design Technol. Integr. Syst. Nanosc. Era (DTIS)*, Apr. 2011, pp. 1–4.

[14] J. Wang, K. Vakilinia, T.-Yi Chen, T. Courtade, G. Dong T. Zhang, H. Shankar, and R. Wesel, "Enhanced precision through multiple reads for LDPC decoding in flash memories," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 880–891, May 2014.

[15] P. J. Smith, M. Shafi, and H. Gao, "Quick simulation: A review of importance sampling techniques in communications systems," *IEEE J. Sel. Areas Commun.*, vol. 15, no. 4, pp. 597–613, May 1997.

[16] B. Xia and W. E. Ryan, "On importance sampling for linear block codes," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 4, May 2003, pp. 2904–2908.

[17] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Generalized multiple importance sampling," *Stat. Sci.*, vol. 34, no. 1, pp. 129–155, Feb. 2019.

[18] V. Elvira and I. Santamaria, "Multiple importance sampling for symbol error rate estimation of maximum-likelihood detectors in MIMO channels," *IEEE Trans. Signal Process.*, vol. 69, pp. 1200–1212, 2021.

[19] N. Kurtz and J. Song, "Cross-entropy-based adaptive importance sampling using Gaussian mixture," *Structural Saf.*, vol. 42, pp. 35–44, May 2013.

[20] M. F. Bugallo, V. Elvira, L. Martino, D. Luengo, J. Miguez, and P. M. Djuric, "Adaptive importance sampling: The past, the present, and the future," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 60–79, Jul. 2017.

[21] F. Han, S. Zhao, H. Jiang, H. Chen, and C. Zhang, "Low-overhead evaluation of multiuser detection performance for physical-layer multiple access systems," *IEEE Access*, vol. 8, pp. 20537–20545, 2020.

[22] J. Font-Segura, A. Martinez, and A. Guillén i Fàbregas, "Importance sampling for coded-modulation error probability estimation," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 289–300, Jan. 2020.

[23] R. Holzlohner, A. Mahadevan, C. R. Menyuk, J. M. Morris, and J. Zweck, "Evaluation of the very low BER of FEC codes using dual adaptive importance sampling," *IEEE Commun. Lett.*, vol. 9, no. 2, pp. 163–165, Feb. 2005.

[24] A. Mahadevan and J. M. Morris, "SNR-invariant importance sampling for hard-decision decoding performance of linear block codes," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 100–111, Jan. 2007.

[25] A. Minja and V. Senk, "Quasi-analytical simulation method for estimating the error probability of star domain decoders," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3101–3113, May 2019.

[26] G. Romano and D. Ciuonzo, "Minimum-variance importance-sampling Bernoulli estimator for fast simulation of linear block codes over binary symmetric channels," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 486–496, Jan. 2014.

[27] J. Pan and W. H. Mow, "A new importance sampling algorithm for fast simulation of linear block codes over BSCs," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Aug. 2019, pp. 1–5.

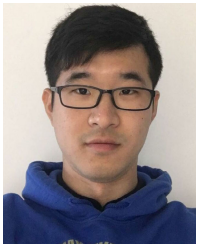
[28] J. Pan. (Aug. 2021). *Minimum-Variance Importance Sampling Estimator of Linear Block Codes Over BsCs*. [Online]. Available: https://github.com/pjzklcb/Importance-sampling-for-BSCs

[29] Q. F. Zhou, W. H. Mow, S. Zhang, and D. Toumpakaris, "Two-way decode-and-forward for low-complexity wireless relaying: Selective forwarding versus one-bit soft forwarding," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1866–1880, Mar. 2016.

[30] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo Method*, vol. 10. Hoboken, NJ, USA: Wiley, 2016.

[31] S. Benedetto and E. Biglieri, *Principles of Digital Transmission: With Wireless Applications*. Cham, Switzerland: Springer, 1999.

[32] O. Keren and S. Litsyn, "A lower bound on the probability of decoding error over a BSC channel," in *Proc. 21st IEEE Conv. Electr. Electron. Eng. Israel*, Apr. 2000, pp. 271–273.
[33] A. Cohen and N. Merhav, "Lower bounds on the error probability of block codes based on improvements on de Caen's inequality," *IEEE Trans. Inf. Theory*, vol. 50, no. 2, pp. 290–310, Feb. 2004.
[34] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*, vol. 1. Upper Saddle River, NJ, USA: Prentice-Hall, 1995.
[35] E. Berlekamp, *Algebraic Coding Theory*. Singapore: World Scientific, 2015.
[36] R. H. Morelos-Zaragoza, *The Art Error Correcting Coding*. Hoboken, NJ, USA: Wiley, 2006.
[37] D. MacKay. *Encyclopedia of Sparse Graph Codes*. Accessed: Nov. 2018. [Online]. Available: http://www.inference.org.U.K./mackay/codes/data.html
[38] Y. Kou, S. Lin, and M. P. Fossorier, "Low-density parity-check codes based on finite geometries: A rediscovery and new results," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2711–2736, Nov. 2001.
[39] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2019.



JINZHE PAN received the B.Eng. degree in information and communication engineering from Zhejiang University (ZJU), Hangzhou, China, in 2015. He is currently pursuing the Ph.D. degree in electronic and computer engineering with The Hong Kong University of Science and Technology (HKUST), Hong Kong, China. His research interests include communication theory, coding, and information theory.

WAI HO MOW (Senior Member, IEEE) received the Ph.D. degree in information engineering from the Chinese University of Hong Kong, in 1993. From 1997 to 1999, he was with the Nanyang Technological University, Singapore. He has been with the Hong Kong University of Science and Technology (HKUST) since 2000 and is currently a Professor. His research areas include coding and information theory, wireless communications, optical camera communications, and thermographic signal processing. He pioneered the lattice approach to signal detection problems, including sphere decoding and complex lattice reduction-aided detection. He unified all known constructions of perfect roots-of-unity (aka CAZAC) sequences, which have been widely used as communication preambles and radar signals. He published two books and 220+ journal/conference publications and is the inventor of 38 patents. His joint work won the top prizes of 10+ project/paper competitions, including the 2014 HK U-21 IoT Gold Award for Revolutional Concept, the Best Paper Award of 2013 and 2016 Asia-Pacific Communications Conference, and the Best Mobile App Award at ACM MobiCom'2013. His co-invented picture barcode PiCode was highlighted as one of the four local innovations in the 2015 International IT Fest organized by the Office of the Government Chief Information Officer, Hong Kong. He is a past chair of the Hong Kong Chapter of the IEEE Information Theory Society, and was the general/program chair of six conferences, incl. SETA'2018 held in HK. He served on the editorial boards of six journals, including the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He is a past member of the Radio Spectrum Advisory Committee, office of the Telecommunications Authority of the Hong Kong S.A.R. Government.