On the Weight Spectrum of Good Linear Binary Codes

Ruoheng Liu, *Student Member, IEEE*,
Predrag Spasojević, *Member, IEEE*, and
Emina Soljanin, *Senior Member, IEEE*

***Abstract*—The weight spectrum of sequences of binary linear codes that achieve arbitrarily small word error probability on a class of noisy channels at a nonzero rate is studied. We refer to such sequences as good codes. The class of good codes includes turbo, low-density parity-check, and repeat-accumulate codes. We show that a sequence of codes is good when transmitted over a memoryless binary-symmetric channel (BSC) or an additive white Gaussian noise (AWGN) channel if and only if the slope of its spectrum is finite everywhere and its minimum Hamming distance goes to infinity with no requirement on its rate growth. The extension of these results to code ensembles in probabilistic terms follows in a direct manner. We also show that the sufficient condition holds for any binary-input memoryless channel.**

***Index Terms*—Binary linear code, good codes, low-density parity-check codes, maximum-likelihood (ML) decoding, turbo codes, weight spectrum.**

I. Introduction

Let $\mathcal{C} = \{\mathcal{C}(n_i)\}_{i=1}^{\infty}$ be a sequence of binary linear codes, where $\mathcal{C}(n_i), \{n_i : i = 1, 2, \ldots\}$, are (n_i, k_i) codes with a common rate $R_c = k_i/n_i$. Following MacKay [1], [2], we say that a code (sequence) $\mathcal{C}$ is *good* if it achieves arbitrarily small word error probability when transmitted over a *noisy* channel at or below a nonzero threshold rate R_c. Capacity achieving codes are good codes whose rate threshold R_c is equal to the channel capacity. We say that a code is *bad* when the corresponding code sequence can not be decoded with an arbitrarily small probability of error, or if it can be decoded with an arbitrarily small probability of error only by decreasing the information rate to zero (e.g., repetition codes).

The performance of a code ensemble is often studied when the performance corresponding to a member code is hard to analyze. We define a binary linear code ensemble $[\mathcal{C}(n)]$ as a *family* of (n, k) code sequences of a common rate. Randomly chosen code (sequence) from a *good code ensemble* is good with probability one. The class of good (ensembles of) codes includes the family of the, so called, *random-like* codes, such as parallel concatenated convolutional codes (classical turbo codes), serial concatenated convolutional codes, *l*ow-*d*ensity *p*arity-*c*heck (LDPC) codes [3], and *r*epeat–*a*ccumulate (RA) codes [4]. Since the threshold rates of these (ensembles of) codes are close to Shannon's limit, their design, analysis, and application has become a focal point of recent research efforts in coding theory.

We aim here to characterize the weight spectrum of linear binary (ensembles of) codes which are good when transmitted over a memoryless binary-symmetric channel (BSC) and an additive white Gaussian noise (AWGN) channel. We study the error probability performance of good (ensembles of) codes under *maximum-likelihood* (ML) decoding. Although, in general, prohibitively complex for long codes, ML decoding based analysis provides the ultimate performance bounds independent of the decoding algorithm.

First, we give a (necessary) condition that the code weight spectrum must satisfy to achieve arbitrarily small word error probability on a BSC and an AWGN channel. We also show that this condition is, in fact, sufficient for any binary-input memoryless channel. Next, we study good code ensembles. We summarize the results from Jin and McEliece in [5] where the authors derive a sufficient condition for goodness of code ensembles based on the average weight spectrum (see also [6] for a stronger sufficient condition). Here, we illustrate why the condition from [5] is only sufficient but not necessary and describe the necessary and sufficient condition for ensemble goodness for a BSC and an AWGN channel using probabilistic language.

The goodness condition is a function of the minimum distance and the behavior of the spectrum in the low codeword weight region. A *good* minimum distance has traditionally been adopted as a criterion for asymptotic code goodness. In [3, Sec. 2.1], Gallager noted that the minimum distance of random binary codes (which are capacity achieving) meets the Gilbert-Varshamov bound, and, thus, increases linearly with the codeword length (see also [7]). However, good minimum distance is a sufficient condition, but *not* necessary for good codes. For example, concatenated Hamming codes have *bad* minimum distance (i.e., the minimum distance normalized by the codeword length converges to zero [2]), whereas, regular LDPC codes have good minimum distance [3]; however, both are good [2]. We show that, to guarantee goodness, it is not important how quickly the minimum distance goes to infinity with the codeword length as long as the slope of the low codeword weight spectrum is finite.

We note that the previous work on capacity achieving (or ϵ-capacity achieving) LDPC code ensembles [8]–[11] has focused on vanishing bit error probability. The requirement to achieve an arbitrarily small word error probability is stricter, since vanishing word error probability implies vanishing bit error probability, but not vice versa. However, for an ensemble of regular LDPC codes (with the degree of variable nodes larger than 2) transmitted over a BSC with the crossover probability less than a threshold [3], Gallager has shown that the bit-error probability under iterative decoding decreases as $\exp(-\phi n^b)$ (when n is large enough), where ϕ and b are positive constants, and n is the codeword length. Since the word error probability is at most n times the bit error probability, it follows that the word error probability will approach zero as n increases. Moreover, for irregular graph-based LDPC code ensembles [10], [12], one can achieve a probability of bit error that approaches zero exponentially fast in terms of n with arbitrarily small loss in rate by concatenating with an appropriate outer code [10], or expanding the graph [12]. In addition, for the capacity achieving LDPC code ensembles and a binary erasure channel, the author in [8] states that one can show that the message-passing algorithm can successfully decode (asymptotically in the codeword length) once the graph is expanded [12], which implies that the word error probability converges to zero.

The remainder of this correspondence is organized as follows: We state the main results in Section II, prove Theorem 1 in Section III, the rest of the proofs are deferred to the Appendix.

II. The Main Result Statement

In this section, we study the weight spectrum of good binary codes and code ensembles. Hereafter, *low weight sequence* refers to a sequence of integers $\{F_n\}$ such that

$$1 \le F_n \le n \quad \text{and} \quad F_n/n \to 0, \qquad \text{as } n \to \infty \tag{1}$$

very low weight sequence refers to a sequence of integers $\{L_n\}$ such that

$$1 \le L_n \le n \quad \text{and} \quad L_n \to 0, \qquad \text{as } n \to \infty$$

Manuscript received November 26, 2004; revised June 2, 2005. This work was supported by the National Science Foundation under Grant CCR-0205362 and SPN-0338805.

R. Liu and P. Spasojević are with WINLAB, Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ 08854 USA (e-mail: liurh@winlab.rutgers.edu; spasojev@winlab.rutgers.edu).

E. Soljanin is with the Mathematical Sciences Research Center, Bell Labs, Lucent Technologies, Murray Hill, NJ 07974 USA (e-mail: emina@lucent.com).

Communicated by R. J. McEliece, Associate Editor for Coding Theory.
Digital Object Identifier 10.1109/TIT.2005.859294

and $\mathcal{F}$ denotes the *set* of all possible low weight sequences $\{F_n\}$. The *slope* of the (normalized) weight spectrum for a codebook $\mathcal{C}(n)$ is defined as

$$S^{\mathcal{C}(n)}(w) \triangleq \frac{\ln A^{\mathcal{C}(n)}(w)}{w}$$

where w denotes the Hamming weight, and the weight enumerator $A^{\mathcal{C}(n)}(w)$ enumerates the codewords of weight w for codebook $\mathcal{C}(n)$.

A. Good Codes

Here, we state the necessary and sufficient condition for goodness of a binary code $\mathcal{C}$ transmitted over a memoryless BSC and AWGN channel. The condition is expressed, in terms of the minimum Hamming weight and the *slope* of weight spectrum of the corresponding sequence of binary linear codes $\{\mathcal{C}(n_i)\}_{i=1}^{\infty}$, in the following theorem.

Theorem 1: A binary linear code $\mathcal{C}$ transmitted over a BSC or an AWGN channel is good if and only if the code satisfies the following condition

T1.a The minimum Hamming distance of a good code $\mathcal{C}$

$$d_{\min}^{\mathcal{C}(n)} \to \infty, \qquad \text{as } n \to \infty.$$

T1.b The low weight spectrum slope of a good code $\mathcal{C}$

$$\limsup_{n\to\infty} S^{\mathcal{C}(n)}(F_n) < \infty, \qquad \forall \{F_n\} \in \mathcal{F}.$$

Furthermore, Condition T1 is sufficient for any binary-input memoryless channel.

The proof is given in Section III. Intuitively, one can interpret Theorem 1 as two codebook design requirements: In order to successfully decode a codeword $\boldsymbol{x}$ at the receiver end, it is required that the distance from $\boldsymbol{x}$ to other codewords is large and the number of $\boldsymbol{x}$'s neighbors is small. Clearly, the former requirement corresponds to Condition T1.a and the latter relates to Condition T1.b.

An immediate consequence of Theorem 1 is the following corollary.

Corollary 1: The good minimum distance property requiring that the ratio $d_{\min}^{\mathcal{C}(n)}/n$ converges to a nonzero constant is a sufficient condition for code goodness.

The proof is in the Appendix.

B. Ensemble Goodness

In this subsection, we focus on a good code ensemble $[\mathcal{C}(n)]$ of rate R_c transmitted over a binary-input memoryless channel. First, we summarize the sufficient condition for goodness of code ensembles in terms of the average weight spectrum [5]. Next, we illustrate that this condition is *not* necessary using counterexamples. Finally, we state a necessary and sufficient condition for ensemble goodness for a BSC and an AWGN channel using probabilistic language.

Following [5], the following theorem states a sufficient condition for the code ensemble goodness in terms of the average weight enumerator $\bar{A}^{[\mathcal{C}(n)]}(w)$ averaged over the ensemble family of codes.

Theorem 2 [5, Theorems 5.1 and 5.3]: A binary code ensemble $[\mathcal{C}]$ is good for any binary-input memoryless channel if it satisfies the following condition

T2.a There exists a sequence of integers $\{D_n\}$ such that

$$D_n \to \infty \quad \text{and} \quad \lim_{n\to\infty} \sum_{w=1}^{D_n} \bar{A}^{[\mathcal{C}(n)]}(w) = 0.$$

T2.b The average low weight spectrum slope of a code ensemble $[\mathcal{C}]$

$$\limsup_{n\to\infty} S^{[\mathcal{C}(n)]}(F_n) < \infty, \qquad \forall \{F_n\} \in \mathcal{F}$$

where

$$S^{[\mathcal{C}(n)]}(w) \triangleq [\ln \bar{A}^{[\mathcal{C}(n)]}(w)]/w.$$

We provide the proof in the Appendix for completeness.

To illustrate the fact that Condition T2 is not necessary for ensemble goodness, we consider the following two counterexamples. Suppose that there is a code ensemble $[\mathcal{C}(n)]$ such that the proportion of good codes in $[\mathcal{C}(n)]$ is $(\sqrt{n}-1)/\sqrt{n}$, and that the proportion of bad codes in $[\mathcal{C}(n)]$ is $1/\sqrt{n}$. By definition, such ensemble is good. Let us now also assume that, for each bad code $\mathcal{C}_b(n)$ in $[\mathcal{C}(n)]$, there exists a constant D such that $\sum_{w=1}^{D} A^{\mathcal{C}_b(n)}(w) = \Theta(\sqrt{n})$, where $f(x) = \Theta(x)$ implies that positive constants c_1, c_2, and k_1 exist, such that $0 \le c_1 x \le f(x) \le c_2 x$ for all $x \ge k_1$. The average weight enumerator of the code ensemble $[\mathcal{C}(n)]$ is now $\sum_{w=1}^{D} \bar{A}^{[\mathcal{C}(n)]}(w) = \Theta(1)$ and, thus, Condition T2.a is not satisfied. An example which illustrates this development is the ensemble of quasicyclic turbo codes [13] with two parallel concatenated branches concatenated by a random interleaver. In [14], the authors note that, for this ensemble of codes, the *expected* number of turbo codewords of small weight averaged over all possible interleavers does not vanish with the code length n (it is constant) but that there is an *all-or-nothing* phenomenon: by randomly choosing a code from the ensemble of turbo codes, we either have no very-low-weight codewords, or relatively many (at least in the order of $\sqrt{n}$). This implies that the proportion of good turbo codes (with no very-low-weight codewords) is 1 in probability even though the ensemble does not satisfy Condition T2.a. Similarly, we can construct another good code ensemble which does not satisfy Condition T2.b as follows. Let $[\mathcal{C}(n)]$ be a good code ensemble such that the proportion of bad codes in $[\mathcal{C}(n)]$ is $1/\sqrt{n}$. We further assume that, for any bad code $\mathcal{C}_b(n)$ in $[\mathcal{C}(n)]$, there exists a sequence of integers $w_n = \lfloor \ln n \rfloor^{1/2}$ such that

$$A^{\mathcal{C}_b(n)}(w_n) \ge n.$$

We now have that $\{w_n\} \in \mathcal{F}$ and

$$\frac{\ln \bar{A}^{[\mathcal{C}(n)]}(w_n)}{w_n} \ge \frac{(\ln n)^{1/2}}{2} \to \infty.$$

Clearly, the above example is a good code ensemble which does not satisfy Condition T2.b.

Finally, we state the necessary and sufficient ensemble goodness condition for a BSC and an AWGN channel using probabilistic language in the following theorem.

Theorem 3: A binary code ensemble $[\mathcal{C}]$ transmitted over a BSC or an AWGN channel is good if and only if a randomly chosen code $\mathcal{C}$ from the ensemble $[\mathcal{C}]$ satisfies

T3.a

$$P\left\{\lim_{n\to\infty} d_{\min}^{\mathcal{C}(n)} \le \alpha\right\} = 0, \qquad \forall \alpha < \infty.$$

T3.b $\exists \beta < \infty$

$$P\left\{\limsup_{n\to\infty} S^{\mathcal{C}(n)}(F_n) > \beta\right\} = 0, \qquad \forall \{F_n\} \in \mathcal{F}.$$

Proof: The proof follows in a straightforward manner from the definition of code ensemble goodness and Theorem 1. □

We note that (see also [5] for the related comment)

$$\begin{aligned}
&P\left\{\lim_{n\to\infty} d_{\min}^{\mathcal{C}(n)} < D_n\right\}\\
&\quad = \frac{1}{|[\mathcal{C}(n)]|}\sum_{\mathcal{C}(n)\in[\mathcal{C}(n)]} \mathbf{1}\left(\lim_{n\to\infty} d_{\min}^{\mathcal{C}}(n) \le D_n\right)\\
&\quad \le \lim_{n\to\infty}\frac{1}{|[\mathcal{C}(n)]|}\sum_{\mathcal{C}(n)\in[\mathcal{C}(n)]}\sum_{w=1}^{D_n} A^{\mathcal{C}(n)}(w)\\
&\quad = \lim_{n\to\infty}\sum_{w=1}^{D_n}\bar{A}^{[\mathcal{C}(n)]}(w)
\end{aligned}$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. Therefore, we say that Condition 2.a implies Condition 3.a.

III. The Proof of Theorem 1

Let $\chi_0^{\mathcal{C}}$ be the union-Bhattacharyya (UB) threshold of the code $\mathcal{C}$, that is

$$\chi_0^{\mathcal{C}} \triangleq \limsup_{n\to\infty}\max_{1\le w\le n} S^{\mathcal{C}(n)}(w). \tag{2}$$

Theorem 1 is based on the following lemma.

Lemma 1: For a code $\mathcal{C}$, the UB noise threshold $\chi_0^{\mathcal{C}}$ is finite if and only if

$$\limsup_{n\to\infty} S^{\mathcal{C}(n)}(F_n) < \infty, \qquad \forall\{F_n\}\in\mathcal{F}$$

i.e., Condition T1.b is satisfied.

Lemma 1 follows from ([5], Theorem 5.3). We provide its proof in the Appendix for completeness. We derive the proof of Theorem 1 in the following.

A. Sufficiency in Theorem 1

The proof of sufficiency is based on the UB bound on the ML decoding word error probability for code $\mathcal{C}(n)$

$$P_W^{\mathcal{C}(n)}(\gamma) \le \sum_{w=d_{\min}^{\mathcal{C}(n)}}^{n} A^{\mathcal{C}(n)}(w)\gamma^w$$

where γ is the Bhattacharyya noise parameter of a binary-input memoryless channel [15]. The UB threshold definition (2) implies that

$$P_W^{\mathcal{C}(n)}(\gamma) \le_n \sum_{w=d_{\min}^{\mathcal{C}(n)}}^{n}\exp\left[-w\left(-\ln\gamma-\chi_0^{\mathcal{C}}-\theta\right)\right], \qquad \forall\theta>0 \tag{3}$$

where $\le_n$ means that the inequality holds for sufficiently large n. Condition T1.b and Lemma 1 imply that $\chi_0^{\mathcal{C}}$ is finite. Thus, for any $\epsilon,\theta>0$, there exists a $\gamma_0(\epsilon,\theta)>0$ such that $-\ln\gamma_0(\epsilon,\theta)-\chi_0^{\mathcal{C}}>\epsilon+\theta>0$. Then, for any $0<\gamma\le\gamma_0(\epsilon,\theta)$, (3) can be upper-bounded as

$$P_W^{\mathcal{C}(n)}(\gamma) \le_n \sum_{w=d_{\min}^{\mathcal{C}(n)}}^{n}\exp(-w\epsilon) \le B(\epsilon)\exp\left[-d_{\min}^{\mathcal{C}(n)}\cdot\epsilon\right]$$

where $B(\epsilon)=1/(1-e^{-\epsilon})$. Now, Condition T1.a implies that

$$\lim_{n\to\infty} P_W^{\mathcal{C}(n)}(\gamma)=0 \quad \forall\gamma\le\gamma_0(\epsilon,\theta) \tag{4}$$

where $\gamma_0(\epsilon,\theta)$ can be made arbitrarily close to $\exp(-\chi_0^{\mathcal{C}})$ by choosing sufficiently small ϵ and θ. Therefore, code $\mathcal{C}$ is good. □

B. Necessity in Theorem 1

Necessity of Condition T1.a: In order to prove the necessity of Condition T1.a, we assume that it does not hold. Then, there exists a constant d_0 such that

$$\liminf_{n\to\infty} d_{\min}^{\mathcal{C}(n)} < d_0 < \infty.$$

Let $\boldsymbol{x}\in\mathcal{C}(n)\subseteq\{0,1\}^n$ be a transmitted codeword. Then, there exists a codeword $\boldsymbol{x}'\in\mathcal{C}(n)$ (where $\boldsymbol{x}'\neq\boldsymbol{x}$) such that the Hamming distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$ can be bounded as follows:

$$0 < d_H(\boldsymbol{x},\boldsymbol{x}') \le_n d_0.$$

Hence, for an AWGN channel with a received SNR $\Gamma<\infty$, the probability that the maximum likelihood (ML) detector selects $\boldsymbol{x}'$ as the more likely one is

$$P_e(\boldsymbol{x},\boldsymbol{x}') \ge_n Q(\sqrt{d_0\cdot\Gamma}) > 0.$$

For a BSC with the cross error probability $p>0$, the pairwise error probability is

$$P_e(\boldsymbol{x},\boldsymbol{x}') \ge_n p^{d_0} > 0.$$

Note that

$$P_W^{\mathcal{C}(n)}(\mu) = P_{W|\boldsymbol{x}}^{\mathcal{C}(n)}(\mu) \ge P_e(\boldsymbol{x},\boldsymbol{x}') > 0$$

where μ denotes the channel parameter (i.e., Γ for an AWGN channel and p for a BSC) and $P_{W|\boldsymbol{x}}^{\mathcal{C}(n)}(\mu)$ is the ML decoding word error probability when $\boldsymbol{x}$ is transmitted over either channel. Thus, the asymptotic ML decoding word error probabilities for an AWGN channel and a BSC are also positive, i.e.,

$$\begin{aligned}
&\lim_{n\to\infty} P_W^{\mathcal{C}(n)}(\Gamma) > 0, \qquad \text{for all } \Gamma<\infty\\
&\lim_{n\to\infty} P_W^{\mathcal{C}(n)}(p) > 0, \qquad \text{for all } p>0
\end{aligned}$$

which contradicts the definition of good codes.

Necessity of Condition T1.b: Let $\{F_n\}$ be an arbitrary low weight sequence from (1), and w be an arbitrary integer such that $1\le w\le F_n$. Consider a constant-weight subset $\mathcal{C}_w^*(n)$ of the codebook $\mathcal{C}(n)$

$$\mathcal{C}_w^*(n) = \{\boldsymbol{x}\in\mathcal{C}(n) : d_H(\boldsymbol{x},\boldsymbol{x}_0)=w\} \tag{5}$$

where $\boldsymbol{x}_0\in\mathcal{C}(n)$ is the all-zero codeword. Since

$$P_W^{\mathcal{C}(n)}(\mu) = P_{W|\boldsymbol{x}}^{\mathcal{C}(n)}(\mu) \ge P_{W|\boldsymbol{x}}^{\mathcal{C}_w^*(n)}(\mu), \qquad \text{for } \boldsymbol{x}\in\mathcal{C}_w^*(n)$$

then

$$\begin{aligned}
P_W^{\mathcal{C}_w^*(n)}(\mu) &= \sum_{\boldsymbol{x}\in\mathcal{C}_w^*(n)} p^*(\boldsymbol{x})P_{W|\boldsymbol{x}}^{\mathcal{C}_w^*(n)}(\mu)\\
&\le \sum_{\boldsymbol{x}\in\mathcal{C}_w^*(n)} p^*(\boldsymbol{x})P_{W|\boldsymbol{x}}^{\mathcal{C}(n)}(\mu) \qquad (6)\\
&= P_W^{\mathcal{C}(n)}(\mu) \qquad (7)
\end{aligned}$$

where $p^*(\boldsymbol{x})$ is codeword probability distribution over $\mathcal{C}_w^*(n)$. In the following, instead of considering the whole codebook $\mathcal{C}(n)$, we only consider the codeword subset $\mathcal{C}_w^*(n)$. Let X^n be the channel-input sequence distributed over $\mathcal{C}_w^*(n)$ and Y^n be the channel-output sequence with the n-tuple alphabet $\mathcal{Y}^n$. Fano's inequality ([16, p. 38]) implies that

$$\log|\mathcal{C}_w^*(n)| \le 1 + P_W^{\mathcal{C}_w^*(n)}(\mu)\log|\mathcal{C}_w^*(n)| + I(X^n;Y^n).$$

Note that $|\mathcal{C}_w^*(n)| = A^{\mathcal{C}(n)}(w)$ and, thus,

$$\log A^{\mathcal{C}(n)}(w) \le \frac{1+I(X^n;Y^n)}{1-P_W^{\mathcal{C}_w^*(n)}(\mu)}. \tag{8}$$

Since the considered channels are memoryless and output-symmetric, the channel transition probability is given by $p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} p(y|x)$ and the conditional entropy satisfies

$$H(Y|X=0) = H(Y|X=1) = H(Y|X).$$

Hence,

$$H(Y^n|X^n) = \sum_{i=1}^{n} H(Y_i|X_i) = nH(Y|X).$$

The mutual information $I(X^n;Y^n)$ can now be expressed as

$$\begin{aligned} I(X^n;Y^n) &= H(Y^n) - H(Y^n|X^n) \\ &\le \sum_{i=1}^{n} H(Y_i) - nH(Y|X) \le n[H_{\bar{p}}(Y) - H(Y|X)] \end{aligned} \tag{9}$$

where

$$H_{\bar{p}}(Y) = -\sum_{y\in\mathcal{Y}} \bar{p}(y)\log\bar{p}(y)$$

and

$$\bar{p}(y) = \frac{1}{n}\sum_{i=1}^{n} p(Y_i = y).$$

The first inequality in (9) corresponds to the subadditivity property of the entropy, and the second inequality follows from the concavity of the entropy $H_p(Y)$ as a function of the density function $p(y)$ (see, e.g., [16]). Combining (7), (8), and (9), we have

$$\frac{\log A^{\mathcal{C}(n)}(w)}{w} \le \frac{1}{1 - P_W^{\mathcal{C}(n)}(\mu)}\left[\frac{1}{w} + \frac{H_{\bar{p}}(Y) - H(Y|X)}{\delta_n}\right] \tag{10}$$

where $\delta_n = w/n$. Note that (10) holds for differential entropy as well.

First, we study (10) for a memoryless BSC with the crossover error probability $0 < p < 1/2$. Since the input sequence is distributed $p^*(\boldsymbol{x})$ over $\mathcal{C}_w^*(n)$

$$\begin{aligned} p(X_i = 0) &= \sum_{\boldsymbol{x}\in\mathcal{C}_w^*(n)} p^*(\boldsymbol{x})p(X_i = 0|\boldsymbol{x}) \\ &= \sum_{\boldsymbol{x}\in\mathcal{C}_w^*(n)} p^*(\boldsymbol{x})\mathbf{1}\{x_i = 0\} \\ p(X_i = 1) &= \sum_{\boldsymbol{x}\in\mathcal{C}_w^*(n)} p^*(\boldsymbol{x})p(X_i = 1|\boldsymbol{x}) \\ &= \sum_{\boldsymbol{x}\in\mathcal{C}_w^*(n)} p^*(\boldsymbol{x})\mathbf{1}\{x_i = 1\} \end{aligned}$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function and x_i denotes the i-th element of codeword $\boldsymbol{x}$. Thus, $\bar{p}(Y=0)$ is given by

$$\begin{aligned} \bar{p}(Y=0) &= \frac{1}{n}\sum_{i=1}^{n} p(Y_i = 0) \\ &= \frac{1}{n}\sum_{i=1}^{n} [p(X_i = 0)p(Y_i = 0|X_i = 0) \\ &\quad + p(X_i = 1)p(Y_i = 0|X_i = 1)] \\ &= (1-p)\times\frac{1}{n}\sum_{i=1}^{n}\sum_{\boldsymbol{x}\in\mathcal{C}_w^*(n)} p^*(\boldsymbol{x})\mathbf{1}\{x_i = 0\} \\ &\quad + p\times\frac{1}{n}\sum_{i=1}^{n}\sum_{\boldsymbol{x}\in\mathcal{C}_w^*(n)} p^*(\boldsymbol{x})\mathbf{1}\{x_i = 1\} \\ &= (1-p)\times\sum_{\boldsymbol{x}\in\mathcal{C}_w^*(n)} p^*(\boldsymbol{x})\frac{\sum_{i=1}^{n}\mathbf{1}\{x_i = 0\}}{n} \\ &\quad + p\times\sum_{\boldsymbol{x}\in\mathcal{C}_w^*(n)} p^*(\boldsymbol{x})\frac{\sum_{i=1}^{n}\mathbf{1}\{x_i = 1\}}{n} \\ &= (1-p)\times\frac{n-w}{n} + p\times\frac{w}{n} \\ &= (1-p) - \delta_n(1-2p) \end{aligned}$$

and, similarly, $\bar{p}(Y=1) = p + \delta_n(1-2p)$. Hence,

$$H_{\bar{p}}(Y) - H(Y|X) = H(p + \delta_n(1-2p)) - H(p)$$

where $H(x) \triangleq -x\log x - (1-x)\log(1-x)$. Note that

$$\delta_n \le \frac{F_n}{n} \to 0 \quad \text{as } n\to\infty$$

and, thus, by employing the L'Hopital's Rule

$$\lim_{\delta_n\to 0}\frac{H_{\bar{p}}(Y) - H(Y|X)}{\delta_n} = (1-2p)\log\frac{1-p}{p} < \infty. \tag{11}$$

Next, we focus on a memoryless AWGN channel

$$Y_i = X_i + Z_i, \qquad \text{for } i = 1,\ldots,n$$

where X_i is a sequence of signals from the alphabet $\mathcal{X} = \{-\sqrt{E_s}, \sqrt{E_s}\}$ ($-\sqrt{E_s}$ and $\sqrt{E_s}$ correspond to the coded symbols "0" and "1," respectively), E_s is the energy per symbol, and $Z_i \sim \mathcal{N}(0,\sigma_z^2)$ is the i.i.d. additive Gaussian noise sequence. The received SNR $\Gamma = E_s/\sigma_z^2$. Since the input codeword $\boldsymbol{x}\in\mathcal{C}_w^*(n)$ and $\boldsymbol{x}_0$ is the all-zero codeword

$$\begin{aligned} E_{\bar{p}}[Y] &= \frac{1}{n}\sum_{i=1}^{n} E[X_i] \\ &= -\frac{n-w}{n}\sqrt{E_s} + \frac{w}{n}\sqrt{E_s} \\ &= -(1-2\delta_n)\sqrt{E_s} \end{aligned}$$

and

$$\begin{aligned} Var_{\bar{p}}(Y) &= E_{\bar{p}}[Y^2] - E_{\bar{p}}[Y]^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} E\left[X_i^2\right] + \sigma_z^2 - (1-2\delta_n)^2 E_s \\ &= \sigma_z^2 + 4\delta_n(1-\delta_n)E_s. \end{aligned}$$

Also,

$$\begin{aligned} h_{\bar{p}}(Y) - h(Y|X) &= h_{\bar{p}}(Y + (1-2\delta_n)\sqrt{E_s}) - \frac{1}{2}\log 2\pi e\sigma_z^2 \\ &\le \frac{1}{2}\log[2\pi e\cdot Var_{\bar{p}}(Y)] - \frac{1}{2}\log 2\pi e\sigma_z^2 \\ &= \frac{1}{2}\log[1 + 4\delta_n(1-\delta_n)\Gamma]. \end{aligned}$$

Finally,

$$\lim_{\delta_n\to 0}\frac{h_{\bar{p}}(Y) - h(Y|X)}{\delta_n} \le 2\Gamma\cdot\log e < \infty. \tag{12}$$

Note that, by definition, a good codes requires

$$\lim_{n\to\infty} P_W^{\mathcal{C}(n)}(\mu) = 0. \tag{13}$$

Thus, for both a BSC and an AWGN channel, (10)–(13) imply that

$$\chi_\flat^{\mathcal{C}} = \limsup_{n\to\infty}\max_{1\le w\le F_n}\frac{\ln A^{\mathcal{C}(n)}(w)}{w} < \infty. \qquad \square$$

Appendix

A. Proof of Corollary 1

Proof: The good minimum distance property implies

$$d_{\min}^{\mathcal{C}(n)} \to \infty \quad \text{and} \quad \limsup_{n\to\infty} S^{\mathcal{C}(n)}(F_n) < 0, \forall \{F_n\} \in \mathcal{F}$$

i.e., Condition T1 is satisfied. Now, Theorem 1 implies that the good minimum distance property is sufficient for code goodness. □

B. Proof of the Lemma 2

Proof ⇒: This direction is trivial. Note that $1 \le F_n \le n$, we have

$$S^{\mathcal{C}(n)}(F_n) \le \max_{1\le w\le n} S^{\mathcal{C}(n)}(w) < \infty, \forall \{F_n\} \in \mathcal{F}.$$

⇐: This direction will be proved using contradiction. Let us assume that $\chi_0^{\mathcal{C}} = \infty$. Thus, following definitions (1) and (2), there exists a convergent subsequence $0 \le \delta_n \le 1$ such that $\delta_n \to \delta_0$ and

$$\lim_{n\to\infty} \frac{\ln A^{\mathcal{C}(n)}(\lfloor n\delta_n \rfloor)}{\lfloor n\delta_n \rfloor} = \infty. \tag{14}$$

Since the weight enumerator can be bounded as (the second inequality follows from [16, p. 284])

$$A^{\mathcal{C}(n)}(w) \le \binom{n}{w} \le e^{nH(w/n)}.$$

Then for any $\delta_0 > 0$ we have

$$\lim_{n\to\infty} \frac{\ln A^{\mathcal{C}(n)}(\lfloor n\delta_n \rfloor)}{\lfloor n\delta_n \rfloor} \le \frac{H(\delta_0)}{\delta_0} < \infty.$$

Thus, (14) can only happen if $\delta_0 = 0$. Let now $F_n = \lfloor n\delta_n \rfloor$. We have $\{F_n\} \in \mathcal{F}$, and

$$S^{\mathcal{C}(n)}(F_n) = \frac{\ln A^{\mathcal{C}(n)}(\lfloor n\delta_n \rfloor)}{\lfloor n\delta_n \rfloor} \to \infty, \qquad \text{as } n \to \infty$$

which contradict the condition

$$\limsup_{n\to\infty} S^{\mathcal{C}(n)}(F_n) < \infty, \forall \{F_n\} \in \mathcal{F}. \quad \square$$

C. Proof of Theorem 2

Proof: Let code ensemble $[\mathcal{C}]$ satisfy Condition T2 and let $\chi_0^{[\mathcal{C}]}$ be the UB threshold of $[\mathcal{C}]$, where

$$\begin{aligned}\chi_0^{[\mathcal{C}]} &\triangleq \limsup_{n\to\infty} \max_{1\le w\le n} S^{[\mathcal{C}(n)]}(w) \\ &= \limsup_{n\to\infty} \max_{1\le w\le n} \frac{\ln \bar{A}^{[\mathcal{C}(n)]}(w)}{w}.\end{aligned} \tag{15}$$

Note that the proof of Lemma 1 (see the Appendix) holds if we consider code ensembles and replace $\chi_0^{\mathcal{C}}$, $S^{\mathcal{C}(n)}(w)$, and $A^{\mathcal{C}(n)}(w)$ by $\chi_0^{[\mathcal{C}]}$, $S^{[\mathcal{C}(n)]}(w)$, and $\bar{A}^{[\mathcal{C}(n)]}(w)$, respectively. Therefore, following Condition T2.b, we have $\chi_0^{[\mathcal{C}]} < \infty$. The definition of (15) implies that

$$\bar{A}^{[\mathcal{C}(n)]}(w) \le_n \exp[w(\chi_0^{[\mathcal{C}]} + \theta)], \qquad \forall \theta > 0.$$

We now can bound the average ML decoding word error probability of $[\mathcal{C}]$ as follows:

$$\begin{aligned}\bar{P}_W^{[\mathcal{C}(n)]}(\gamma) &\le \sum_{w=1}^{n} \bar{A}^{[\mathcal{C}(n)]}(w)\gamma^w \\ &\le \sum_{w=1}^{D_n} \bar{A}^{[\mathcal{C}(n)]}(w) \\ &\quad + \sum_{w=D_n+1}^{n} e^{-w(-\ln\gamma - \chi_0^{[\mathcal{C}]} - \theta)}\end{aligned} \tag{16}$$

where D_n is defined in Condition T2.a. Note that $\chi_0^{[\mathcal{C}]} < \infty$. Hence, for any $\epsilon, \theta > 0$, there exists a $\gamma_0(\epsilon,\theta) > 0$ such that $-\ln\gamma_0(\epsilon,\theta) - \chi_0^{[\mathcal{C}]} > \epsilon + \theta > 0$. Then, for any $0 < \gamma \le \gamma_0(\epsilon,\theta)$, (16) can be upper-bounded as

$$\begin{aligned}\bar{P}_W^{[\mathcal{C}(n)]}(\gamma) &\le_n \sum_{w=1}^{D_n} \bar{A}^{[\mathcal{C}(n)]}(w) + \sum_{w=D_n+1}^{n} \exp(-w\epsilon) \\ &\le \sum_{w=1}^{D_n} \bar{A}^{[\mathcal{C}(n)]}(w) + B'(\epsilon)\exp[-D_n\epsilon]\end{aligned}$$

where $B'(\epsilon) = e^{-\epsilon}/(1-e^{-\epsilon})$. Now, Condition T2.a implies that

$$\lim_{n\to\infty} \bar{P}_W^{[\mathcal{C}(n)]}(\gamma) = 0, \quad \forall \gamma \le \gamma_0(\epsilon,\theta).$$

Therefore, the code ensemble $[\mathcal{C}]$ is good.

References

[1] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 399–431, Mar. 1999.

[2] ——, *Information Theory, Inference and Learning Algorithms*. New York: Cambridge Univ. Press, 2003.

[3] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA: MIT Press, 1963.

[4] D. Divsalar, H. Jin, and R. J. McEliece, "Coding theorems for 'turbo-like' codes," in *Proc. 1998 Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sep. 1998, pp. 201–210.

[5] H. Jin and R. J. McEliece, "Coding theorems for turbo code ensembles," *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 1451–1461, Jun. 2002.

[6] D. Divsalar, S. Dolinar, H. Jin, and R. J. McEliece, "AWGN coding theorems from ensemble weight enumerators," in *Proc. IEEE Int. Symp. Information Theory*, Sorrento, Italy, Jun. 2000, p. 459.

[7] A. Barg and G. D. Forney, Jr., "Random codes: Minimum distances and error exponents," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2568–2573, Sep. 2002.

[8] A. Shokrollahi, *Codes, Systems and Graphical Models*, B. Marcus and J. Rosenthal, Eds. New York: Springer-Verlag, 2000. Capacity-achieving sequences, ser. IMA Volumes in Mathematics and Its Applications.

[9] P. Oswald and A. Shokrollahi, "Capacity-achieving sequences for the erasure channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3017–3028, Dec. 2002.

[10] T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[11] I. Sason and R. Urbanke, "Parity-check density versus performance of binary linear block codes over memoryless symmetric channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1611–1635, Jul. 2003.

[12] M. G. Luby, M. Mitzenmacher, A. Shokrollahi, and D. A. Spielman, "Improved low-density parity-check codes using irregular graphs," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 585–598, Feb. 2001.

[13] R. M. Tanner, "Toward an algebraic theory for turbo codes," in *Proc. 2nd Int. Symp. Turbo Codes*, Brest, France, Sep. 2000, pp. 17–25.

[14] J. Boutros and G. Zemor, "Interleavers for turbo codes that yield a minimum distance growing with blocklength," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 55.

[15] R. J. McEliece, *The Theory of Information and Coding*, 2nd ed. New York: Cambridge Univ. Press, 2001.

[16] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.