Distance-Spectrum Formulas on the Largest Minimum Distance of Block Codes

Po-Ning Chen, *Member, IEEE*, Tzong-Yow Lee, and Yunghsiang S. Han, *Member, IEEE*

***Abstract*—A general formula for the asymptotic largest minimum distance (in block length) of deterministic block codes under *generalized* distance functions (not necessarily additive, symmetric, and bounded) is presented. As revealed in the formula, the largest minimum distance can be fully determined by the ultimate statistical characteristics of the normalized distance function evaluated under a properly chosen random-code generating distribution. Interestingly, the new formula has an analogous form to the general information-spectrum expressions of the channel capacity and the optimistic channel capacity, respectively derived by Verdú–Han [29] and Chen–Alajaji [7], [8]. As a result, a minor class of distance functions for which the largest minimum distance can be derived is characterized. A general Varshamov–Gilbert lower bound is next addressed. Some discussions on the tightness of the general Varshamov–Gilbert bound are also provided. Finally, lower bounds on the largest minimum distances for several specific block coding schemes are rederived in terms of the new formulas, followed by comparisons with the known results devoted to the same codes.**

***Index Terms*—Block codes, information spectrum, Varshamov–Gilbert bound.**

I. Introduction

THE ultimate capabilities and limitations of error-correcting codes are quite important, especially for code designers who want to estimate the relative efficacy of the designed code. In fairly general situations, this information is closely related to the largest minimum distance of the codes [24]. One of the examples is that for a binary block code employing the Hamming distance, the error correcting capability of the code is half of the minimum distance among codewords. Hence, the knowledge of the largest minimum distance can be considered as a reference of the optimal error correcting capability of codes.

The problem on the largest minimum distance can be described as follows. Over a given code alphabet, and a given measurable function on the "distance" between two code symbols, determine the asymptotic ratio, the largest minimum distance attainable among M selected codewords divided by the code block length n, as n tends to infinity, subject to a fixed rate $R \triangleq \log(M)/n$.

Research on this problem has been done for years. Up to the present, only bounds on this ratio are established. The best known bound on this problem is the Varshamov–Gilbert lower bound, which is usually derived in terms of a combinatorial approximation under the assumption that the code alphabet is finite and the measure on the "distance" between code letters is symmetric [20]. If the size of the code alphabet q is an even power of a prime, satisfying $q \geq 49$, and the distance measure is the Hamming distance, a better lower bound can be obtained through the construction of the Algebraic-Geometric code [14], [28], the idea of which was first proposed by Goppa. Later, Zinoviev and Litsyn proved that a better lower bound than the Varshamov–Gilbert bound is actually possible for any $q \geq 46$ [32]. Other improvements of the bounds can be found in [12], [21], and [30].

In addition to the combinatorial techniques, some researchers also apply the probabilistic and analytical methodologies to this problem. For example, by means of the random coding argument with expurgation, the Varshamov–Gilbert bound in its most general form can be established by simply using the Chebyshev inequality ([3] or cf. Appendix A), and restrictions on the code alphabet (such as finite, countable, ...) and the distance measure (such as additive, symmetric, bounded, ...) are no longer necessary for the validity of its proof.

Recently, channels without statistical assumptions such as memoryless, information stability, stationarity, causality, and ergodicity, ..., etc., have been successfully handled by employing the notions of *liminf in probability* and *limsup in probability*[1] of the information spectrum. As a consequence, the channel capacity C is shown to equal the supremum, over all input processes, of the input–output *inf-information rate* defined as the liminf in probability of the normalized information density [29]. More specifically, given a channel $\boldsymbol{W} = \{W^n = P_{Y^n|X^n}\}_{n=1}^{\infty}$,

$$C = \sup_{\boldsymbol{X}} \sup \left\{ a \in \Re : \limsup_{n\to\infty} \Pr\left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n) < a\right] = 0 \right\}$$

[1]If $\{A_n\}_{n\geq 1}$ is a sequence of random variables, then its *liminf in probability* is the largest extended real number $\underline{U}$ such that $\lim_{n\to\infty} \Pr[A_n < \underline{U}] = 0$. Similarly, its *limsup in probability* is the smallest extended real number $\bar{U}$ such that $\lim_{n\to\infty} \Pr[A_n > \bar{U}] = 0$ [17].

Manuscript received January 17, 1999; revised August 11, 1999. The work of P.-N. Chen is supported by the National Science Council under Grant NSC 87-2213-E-009-139. The work of T.-Y. Lee was supported by the University of Maryland, College Park. The material in this paper was presented in part at the International Symposium on Communications, Hsin Chu, Taiwan, R.O.C., Septepmber 1993.

P.-N. Chen is with the Department of Communications Engineering, National Chiao Tung University, Hsin Chu, Taiwan 30050, R.O.C. (e-mail: poning@cc.nctu.edu.tw).

T.-Y. Lee is with the Department of Mathematics, University of Maryland, College Park, MD 20742 USA (e-mail: tyl@math.umd.edu).

Y. S. Han is with the Department of Computer Science and Information Engineering, National Chi Nan University, Nan Tou, Taiwan, R.O.C. (e-mail: yshan@csie.ncnu.edu.tw).

Communicated by S. Shamai, Assocaite Editor for Shannon Theory.

Publisher Item Identifier S 0018-9448(00)03095-9.

where X^n and Y^n are, respectively, the n-fold input process drawn from

$$\boldsymbol{X} = \left\{X^n = \left(X_1^{(n)}, \cdots, X_n^{(n)}\right)\right\}_{n=1}^{\infty}$$

and the corresponding output process induced by X^n via the channel $W^n = P_{Y^n|X^n}$, and

$$\frac{1}{n} i_{X^n W^n}(x^n; y^n) \triangleq \frac{1}{n} \log \frac{P_{Y^n|X^n}(y^n \mid x^n)}{P_{Y^n}(y^n)}$$

is the normalized information density. If the conventional definition of channel capacity, which requires the existence of reliable block codes for *all sufficiently large block lengths*, is replaced by that reliable codes exist for *infinitely many block lengths*, a new *optimistic* definition of capacity $\bar{C}$ is obtained [29]. Its information-spectrum expression is then given by [7], [8]

$$\bar{C} = \sup_{\boldsymbol{X}} \sup \left\{ a \in \Re : \liminf_{n\to\infty} \Pr\left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n) < a\right] = 0 \right\}.$$

Inspired by such probabilistic methodology, together with the random-coding scheme with expurgation, a spectrum formula on the largest minimum distance of deterministic block codes for *generalized* distance functions[2] (not necessarily additive, symmetric, and bounded) is established in this work. As revealed in the formula, the largest minimum distance is completely determined by the ultimate statistical characteristics of the normalized distance function evaluated under a properly chosen random-code generating distribution. Interestingly, the new formula has an analogous form to the general information-spectrum expressions of the channel capacity and the optimistic channel capacity. This somehow confirms the connection between the problem of designing a reliable code for a given channel and that of finding a code with sufficiently large distance among codewords, if the distance function is properly defined in terms of the channel statistics.

With the help of the new formula, we characterize a minor class of distance metrics for which the ultimate largest minimum distance among codewords can be derived. Although these distance functions may be of secondary interest, it sheds some light on the determination of the largest minimum distance for a more general class of distance functions. Discussions on the general properties of the new formula will follow.

We next derive a general Varshamov–Gilbert lower bound directly from the new distance-spectrum formula. Some remarks on its properties are given. A sufficient condition under which the general Varshamov–Gilbert bound is tight, as well as examples to demonstrate its strict inferiority to the distance-spectrum formula, are also provided.

Finally, we demonstrate that the new formula can be used to derive the known lower bounds for a few specific block coding schemes of general interests, such as constant-weight codes and the codes that corrects arbitrary noise [10]–[12], [15], [16], [18], [23], [27], [31]. Transformation of the asymptotic distance determination problem into an alternative problem setting over a graph for a possible improvement of these known bounds is also addressed.

The rest of the paper is organized as follows. The distance-spectrum formula is derived in Section II. The determination of the asymptotic largest minimum distances among codewords for a class of distance functions is covered in Section III. Section IV presents the general properties of the distance-spectrum formula, followed by examples and remarks on these properties. Section V establishes the general Varshamov–Gilbert lower bound directly from the distance-spectrum formula. Also covered in the same section is a sufficient condition under which the general Varshamov–Gilbert bound is tight, as well as examples to demonstrate the strict superiority of the new formula to the general Varshamov–Gilbert bound. Section VI shows that the new formula can be used to derive the known bounds for specific coding schemes of general interests. Final comments appear in Section VII.

Throughout this paper, the natural logarithm is employed unless otherwise stated.

II. Distance-Spectrum Formula on the Largest Minimum Distance of Block Codes

We first introduce some notations. The n-tuple code alphabet is denoted by $\mathcal{X}^n$. For any two elements $\hat{x}^n$ and x^n in $\mathcal{X}^n$, we use $\mu_n(\hat{x}^n, x^n)$ to denote the n-fold measure on the "distance" of these two elements. A codebook with block length n and size M is represented by

$$\mathfrak{C}_{n,M} \triangleq \left\{\boldsymbol{c}_0^{(n)}, \boldsymbol{c}_1^{(n)}, \boldsymbol{c}_2^{(n)}, \cdots, \boldsymbol{c}_{M-1}^{(n)}\right\}$$

where $\boldsymbol{c}_m^{(n)} \triangleq (c_{m1}, c_{m2}, \cdots, c_{mn})$, and each c_{mk} belongs to $\mathcal{X}$. We define the minimum distance

$$d_m(\mathfrak{C}_{n,M}) \triangleq \min_{\substack{0\le \hat{m}\le M-1 \\ \hat{m}\ne m}} \mu_n\left(\boldsymbol{c}_{\hat{m}}^{(n)}, \boldsymbol{c}_m^{(n)}\right)$$

and the largest minimum distance

$$d_{n,M} \triangleq \max_{\mathfrak{C}_{n,M}} \min_{0\le m\le M-1} d_m(\mathfrak{C}_{n,M}).$$

Note that there is no assumption on the code alphabet $\mathcal{X}$ and the sequence of the functions $\{\mu_n(\cdot,\cdot)\}_{n\ge 1}$.

Based on the above definitions, the problem considered in this paper becomes to find the limit, as $n \to \infty$, of $d_{n,M}/n$ under a fixed rate $R = \log(M)/n$. Since the quantity is investigated as n goes to infinity, it is justified to take $M = e^{nR}$ as integers.

The concept of our method is similar to that of the random coding technique employed in the channel reliability function [2]. Each codeword is assumed to be selected independently of all others from $\mathcal{X}^n$ through a generic distribution P_{X^n}. Then the distance between codewords $\boldsymbol{c}_m^{(n)}$ and $\boldsymbol{c}_{\hat{m}}^{(n)}$ becomes a random variable, and so does $d_m(\mathfrak{C}_{n,M})$. For clarity, we will use D_m to denote the random variable corresponding to $d_m(\mathfrak{C}_{n,M})$. Also note that $\{D_m\}_{m=0}^{M-1}$ are identically distributed. We therefore have the following lemma.

[2]Conventionally, a *distance* or *metric* [19], [26, pp. 139] should satisfy the properties of i) nonnegativity; ii) being zero iff two points coincide; iii) symmetry; and iv) triangle inequality. The derivation in this paper, however, is applicable to any measurable function defined over the code alphabets. Since none of the above four properties are assumed, the measurable function on the "distance" between two code letters is therefore termed *generalized distance* function. One can, for example, apply our formula to situation where the code alphabet is a distribution space, and the "distance" measure is the Kullback–Leibler divergence. For simplicity, we will abbreviate the *generalized distance* function simply as the *distance* function in the remaining part of the paper.

Lemma 1: Fix a triangular-array random process

$$\boldsymbol{X} = \{X^n = (X_1^{(n)}, \cdots, X_n^{(n)})\}_{n=1}^{\infty}.$$

Let

$$D_m = D_m(X^n), \qquad 0 \le m \le (M-1)$$

be defined for the random codebook of block length n and size M, where each codeword is drawn independently according to the distribution P_{X^n}. Then

1) for any $\gamma > 0$, there exists a universal constant $\alpha = \alpha(\gamma) \in (0,1)$ (independent of block length n) and a codebook sequence $\{\mathfrak{C}_{n,\alpha M}\}_{n\ge 1}$ such that

$$\frac{1}{n}\min_{0\le m\le \alpha M-1} d_m(\mathfrak{C}_{n,\alpha M}) > \inf\left\{a \in \Re : \limsup_{n\to\infty} \Pr\left[\frac{1}{n}D_m > a\right] = 0\right\} - \gamma \tag{2.1}$$

for infinitely many n;

2) for any $\gamma > 0$, there exists a universal constant $\alpha = \alpha(\gamma) \in (0,1)$ (independent of block length n) and a codebook sequence $\{\mathfrak{C}_{n,\alpha M}\}_{n\ge 1}$ such that

$$\frac{1}{n}\min_{0\le m\le \alpha M-1} d_m(\mathfrak{C}_{n,\alpha M}) > \inf\left\{a \in \Re : \liminf_{n\to\infty} \Pr\left[\frac{1}{n}D_m > a\right] = 0\right\} - \gamma \tag{2.2}$$

for sufficiently large n.

Proof: We will only prove (2.1). (2.2) can be proved by simply following the same procedure.

Define

$$\bar{L}_{\boldsymbol{X}}(R) \triangleq \inf\left\{a : \limsup_{n\to\infty} \Pr\left[\frac{1}{n}D_m > a\right] = 0\right\}.$$

Let $\mathbf{1}(\mathcal{A})$ be the indicator function of a set $\mathcal{A}$, and let

$$\phi_m \triangleq \mathbf{1}\left(\frac{1}{n}D_m > \bar{L}_{\boldsymbol{X}}(R) - \gamma\right).$$

By definition of $\bar{L}_{\boldsymbol{X}}(R)$

$$\limsup_{n\to\infty} \Pr\left[\frac{1}{n}D_m > \bar{L}_{\boldsymbol{X}}(R) - \gamma\right] > 0.$$

Let

$$2\alpha \triangleq \limsup_{n\to\infty} \Pr[(1/n)D_m > \bar{L}_{\boldsymbol{X}}(R) - \gamma].$$

Then for infinitely many n

$$\Pr\left[\frac{1}{n}D_m > \bar{L}_{\boldsymbol{X}}(R) - \gamma\right] > \alpha.$$

For those n that satisfy the above inequality

$$E\left[\sum_{m=0}^{M-1}\phi_m\right] = \sum_{m=0}^{M-1} E[\phi_m] > \alpha M$$

which implies that among all possible selections, there exist (for infinite many n) a codebook $\mathfrak{C}_{n,M}$ in which αM codewords satisfy $\phi_m = 1$, i.e.,

$$\frac{1}{n}d_m(\mathfrak{C}_{n,M}) > \bar{L}_{\boldsymbol{X}}(R) - \gamma \quad \text{for at least } \alpha M \text{ codewords in the codebook } \mathfrak{C}_{n,M}.$$

The collection of these αM codewords is a desired codebook for the validity of (2.1). □

Our second lemma concerns the spectrum of $(1/n)D_m$.

Lemma 2: Let each codeword be independently selected through the distribution P_{X^n}. Suppose that $\hat{X}^n$ is independent of, and has the same distribution as, X^n. Then

$$\Pr\left[\frac{1}{n}D_m > a\right] \ge \left(\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n, X^n) > a\right\}\right)^M.$$

Proof: Let $\boldsymbol{C}_m^{(n)}$ denote the mth randomly selected codeword. From the definition of D_m, we have

$$\begin{aligned}
&\Pr\left\{\frac{1}{n}D_m > a \,\middle|\, \boldsymbol{C}_m^{(n)}\right\} \\
&= \Pr\left\{\min_{\substack{0\le \hat{m}<M-1\\ \hat{m}\ne m}} \frac{1}{n}\mu_n\left(\boldsymbol{C}_{\hat{m}}^{(n)}, \boldsymbol{C}_m^{(n)}\right) > a \,\middle|\, \boldsymbol{C}_m^{(n)}\right\} \\
&= \prod_{\substack{0\le \hat{m}<M-1\\ \hat{m}\ne m}} \Pr\left\{\frac{1}{n}\mu_n\left(\boldsymbol{C}_{\hat{m}}^{(n)}, \boldsymbol{C}_m^{(n)}\right) > a \,\middle|\, \boldsymbol{C}_m^{(n)}\right\} \\
&= \left(\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n, X^n) > a \,\middle|\, X^n\right\}\right)^{M-1}
\end{aligned} \tag{2.3}$$

where (2.3) holds because

$$\{(1/n)\mu_n(\boldsymbol{C}_{\hat{m}}^{(n)}, \boldsymbol{C}_m^{(n)})\}_{0\le \hat{m}\le M-1, \hat{m}\ne m}$$

is conditionally independent given $\boldsymbol{C}_m^{(n)}$. Hence

$$\begin{aligned}
&\Pr\left\{\frac{1}{n}D_m > a\right\} \\
&= \int_{\mathcal{X}^n} \left(\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n, x^n) > a\right\}\right)^{M-1} dP_{X^n}(x^n) \\
&\ge \int_{\mathcal{X}^n} \left(\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n, x^n) > a\right\}\right)^{M} dP_{X^n}(x^n) \\
&= E_{X^n}\left[\left(\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n, X^n) > a \,\middle|\, X^n\right\}\right)^M\right] \\
&\ge E_{X^n}^M\left[\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n, X^n) > a \,\middle|\, X^n\right\}\right] \\
&= \left(\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n, X^n) > a\right\}\right)^M
\end{aligned} \tag{2.4}$$

where (2.4) follows from Lyapounov's inequality [1, p. 76], i.e., $E^{1/M}[U^M] \ge E[U]$ for a nonnegative random variable U. □

We are now ready to prove the main theorem of the paper. For simplicity, throughout the article, $\hat{X}^n$ and X^n are used specifically to denote two independent random variables having common distribution P_{X^n}.

Theorem 1 (Distance-Spectrum Formula):

$$\sup_{\boldsymbol{X}} \bar{\Lambda}_{\boldsymbol{X}}(R) \geq \limsup_{n\to\infty} \frac{d_{n,M}}{n} \geq \sup_{\boldsymbol{X}} \bar{\Lambda}_{\boldsymbol{X}}(R+\delta) \tag{2.5}$$

and

$$\sup_{\boldsymbol{X}} \underline{\Lambda}_{\boldsymbol{X}}(R) \geq \liminf_{n\to\infty} \frac{d_{n,M}}{n} \geq \sup_{\boldsymbol{X}} \underline{\Lambda}_{\boldsymbol{X}}(R+\delta) \tag{2.6}$$

for every $\delta > 0$, where

$$\bar{\Lambda}_{\boldsymbol{X}}(R) \triangleq \inf\left\{a \in \Re : \limsup_{n\to\infty}\left(\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n, X^n) > a\right\}\right)^M = 0\right\}$$

and

$$\underline{\Lambda}_{\boldsymbol{X}}(R) \triangleq \inf\left\{a \in \Re : \liminf_{n\to\infty}\left(\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n, X^n) > a\right\}\right)^M = 0\right\}.$$

Proof:

1) *Lower bound*. Observe that in Lemma 1, the rate only decreases by the amount $-\log(\alpha)/n$ when employing a code $\mathfrak{C}_{n,\alpha M}$. Also note that for any $\delta > 0, -\log(\alpha)/n < \delta$ for sufficiently large n. These observations, together with Lemma 2, imply the validity of the lower bound.
2) *Upper bound*. Again, we will only prove (2.5), since (2.6) can be proved by simply following the same procedure.

   To show that the upper bound of (2.5) holds, it suffices to prove the existence of $\boldsymbol{X}$ such that

$$\bar{\Lambda}_{\boldsymbol{X}}(R) \geq \limsup_{n\to\infty} \frac{d_{n,M}}{n}.$$

   Let X^n be uniformly distributed over one of the optimal codes $\mathfrak{C}^*_{n,M}$. (By "optimal" we mean that the code has the largest minimum distance among all codes of the same size.) Define

$$\lambda_n \triangleq \frac{1}{n}\min_{0\leq m\leq M-1} d_m(\mathfrak{C}^*_{n,M}) \quad \text{and} \quad \bar{\lambda} \triangleq \limsup_{n\to\infty} \lambda_n.$$

   Then for any $\delta > 0$

$$\lambda_n > \bar{\lambda} - \delta \quad \text{for infinitely many } n. \tag{2.7}$$

   For those n satisfying (2.7)

$$\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n, X^n) > \bar{\lambda} - \delta\right\} \geq \Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n, X^n) \geq \lambda_n\right\} \geq \Pr\{\hat{X}^n \neq X^n\} = 1 - \frac{1}{M}$$

   which implies

$$\limsup_{n\to\infty}\left(\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n, X^n) > \bar{\lambda} - \delta\right\}\right)^M \geq \limsup_{n\to\infty}\left(1 - \frac{1}{M}\right)^M = \limsup_{n\to\infty}\left(1 - \frac{1}{e^{nR}}\right)^{e^{nR}} = e^{-1} > 0.$$

   Consequently, $\bar{\Lambda}_{\boldsymbol{X}}(R) \geq \bar{\lambda} - \delta$. Since δ can be made arbitrarily small, the upper bound holds. □

Observe that $\sup_{\boldsymbol{X}} \bar{\Lambda}_{\boldsymbol{X}}(R)$ is nonincreasing in R, and hence, the number of discontinuities is countable. This fact implies that

$$\sup_{\boldsymbol{X}} \bar{\Lambda}_{\boldsymbol{X}}(R) = \lim_{\delta\downarrow 0} \sup_{\boldsymbol{X}} -(R+\delta)$$

except for countably many R. Similar argument applies to $\sup_{\boldsymbol{X}} \underline{\Lambda}_{\boldsymbol{X}}(R)$. We can then rephrase the above theorem as appeared in the next corollary.

Corollary 1:

$$\limsup_{n\to\infty} \frac{d_{n,M}}{n} = \sup_{\boldsymbol{X}} \bar{\Lambda}_{\boldsymbol{X}}(R) \quad \left(\text{resp., } \liminf_{n\to\infty} \frac{d_{n,M}}{n} = \sup_{\boldsymbol{X}} \underline{\Lambda}_{\boldsymbol{X}}(R)\right)$$

except possibly at the points of discontinuities of $\sup_{\boldsymbol{X}} \bar{\Lambda}_{\boldsymbol{X}}(R)$ (resp., $\sup_{\boldsymbol{X}} \underline{\Lambda}_{\boldsymbol{X}}(R)$), which are countable.

From the above theorem (or corollary), we can characterize the largest minimum distance of deterministic block codes in terms of the distance spectrum. We thus name it *distance-spectrum formula*. For convenience, $\bar{\Lambda}_{\boldsymbol{X}}(\cdot)$ and $\underline{\Lambda}_{\boldsymbol{X}}(\cdot)$ will be, respectively, called the *sup-distance-spectrum function* and the *inf-distance-spectrum function* in the remainder of the paper.

We conclude this section by remarking that the distance-spectrum formula obtained above indeed has an analogous form to the information-spectrum formulas of the channel capacity and the optimistic channel capacity. Furthermore, by taking the distance metric to be the n-fold Bhattacharyya distance [2, Definition 5.8.3], an upper bound on channel reliability [2, Theorem 10.6.1] can be obtained, i.e.,

$$\limsup_{n\to\infty} -\frac{1}{n}\log P_e(n, M = e^{nR}) \leq \sup_{\boldsymbol{X}} \inf\left\{a : \limsup_{n\to\infty}\left(\Pr\left[-\frac{1}{n} \times \log \sum_{y^n \in \mathcal{Y}^n} P^{1/2}_{Y^n|X^n}(y^n \mid \hat{X}^n) P^{1/2}_{Y^n|X^n}(y^n \mid X^n) > a\right]\right)^M = 0\right\}$$

and

$$\liminf_{n\to\infty} -\frac{1}{n}\log P_e(n, M = e^{nR}) \leq \sup_{\boldsymbol{X}} \inf\left\{a : \liminf_{n\to\infty}\left(\Pr\left[-\frac{1}{n} \times \log \sum_{y^n \in \mathcal{Y}^n} P^{1/2}_{Y^n|X^n}(y^n \mid \hat{X}^n) P^{1/2}_{Y^n|X^n}(y^n \mid X^n) > a\right]\right)^M = 0\right\}$$

where $P_{Y^n|X^n}$ is the n-dimensional channel transition distribution from code alphabet $\mathcal{X}^n$ to channel output alphabet $\mathcal{Y}^n$, and $P_e(n, M)$ is the average probability of error for optimal channel code of block length n and size M. (The proof of this sphere-packing-type bound is given in Appendix C for the sake

of completeness.) Note that the formula of the above channel reliability bound is quite different from those formulated in terms of the exponents of the information spectrum (cf. [5, Sec. V] and [25, eq. (14)]).

III. Determination of the Largest Minimum Distance for a Class of Distance Functions

In this section, we will present a minor class of distance functions for which the optimization input $\boldsymbol{X}$ for the distance-spectrum function can be characterized, and thereby, the ultimate largest minimum distance among codewords can be established in terms of the distance-spectrum formula.

A simple example for which the largest minimum distance can be derived in terms of the new formula is the probability-of-error distortion measure, which is defined as

$$\mu_n(\hat{x}^n, x^n) = \begin{cases} 0, & \text{if } \hat{x}^n = x^n \\ n, & \text{if } \hat{x}^n \neq x^n. \end{cases}$$

It can be easily shown that

$$\begin{aligned} &\sup_{\boldsymbol{X}} \bar{\Lambda}_{\boldsymbol{X}}(R) \\ &\quad \leq \inf\left\{ a \in \Re : \limsup_{n\to\infty} \sup_{X^n} \left(\Pr\left\{ \frac{1}{n}\mu_n(\hat{X}^n, X^n) > a \right\} \right)^M = 0 \right\} \\ &\quad = \begin{cases} 1, & \text{for } 0 \leq R < \log|\mathcal{X}| \\ 0, & \text{for } R \geq \log|\mathcal{X}| \end{cases} \end{aligned}$$

and the upper bound can be achieved by letting $\boldsymbol{X}$ be uniformly distributed over the code alphabet. Similarly

$$\sup_{\boldsymbol{X}} \underline{\Lambda}_{\boldsymbol{X}}(R) = \begin{cases} 1, & \text{for } 0 \leq R < \log|\mathcal{X}| \\ 0, & \text{for } R \geq \log|\mathcal{X}|. \end{cases}$$

Another example for which the optimizer $\boldsymbol{X}$ of the distance-spectrum function can be characterized is the *separable distance function* defined below.

Definition 1 (Separable Distance Functions):

$$\mu_n(\hat{x}^n, x^n) \triangleq f_n(|g_n(\hat{x}^n) - g_n(x^n)|)$$

where $f_n(\cdot)$ and $g_n(\cdot)$ are real-valued functions.

Next, we derive the basis for finding one of the optimization distributions for

$$\sup_{X^n} \Pr\left\{ \frac{1}{n}\mu_n(\hat{X}^n, X^n) > a \right\}$$

under separable distance functionals.

Lemma 3: Define

$$G(\alpha) \triangleq \inf_{U} \Pr\{|\hat{U} - U| \leq \alpha\}$$

where the infinum is taken over all $(\hat{U}, U)$ pair having independent and identical distribution on $[0, 1]$. Then for $j = 2, 3, 4, \cdots$ and $1/j \leq \alpha < 1/(j-1)$

$$G(\alpha) = \frac{1}{j}.$$

In addition, $G(\alpha)$ is achieved by uniform distribution over

$$\left\{0, \frac{1}{j-1}, \frac{2}{j-1}, \cdots, \frac{j-2}{j-1}, 1\right\}.$$

Proof:

$$\begin{aligned} \Pr\{|\hat{U} - U| \leq \alpha\} &\geq \Pr\left\{|\hat{U} - U| \leq \frac{1}{j}\right\} \\ &\geq \sum_{i=0}^{j-2} \Pr\left\{(\hat{U}, U) \in \left[\frac{i}{j}, \frac{i+1}{j}\right)^2\right\} \\ &\quad + \Pr\left\{(\hat{U}, U) \in \left[\frac{j-1}{j}, 1\right]^2\right\} \\ &= \sum_{i=0}^{j-2} \left(\Pr\left\{U \in \left[\frac{i}{j}, \frac{i+1}{j}\right)\right\}\right)^2 \\ &\quad + \left(\Pr\left\{U \in \left[\frac{j-1}{j}, 1\right]\right\}\right)^2 \\ &\geq \frac{1}{j}. \end{aligned}$$

Achievability of $G(\alpha)$ by uniform distribution over

$$\left\{0, \frac{1}{j-1}, \frac{2}{j-1}, \cdots, \frac{j-2}{j-1}, 1\right\}$$

can be easily verified, and hence, we omit here. □

Lemma 4: For $1/j \leq \alpha < 1/(j-1)$

$$\frac{1}{j} \leq \inf_{U_n} \Pr\{|\hat{U}_n - U_n| \leq \alpha\} \leq \frac{1}{j} + \frac{1}{n+0.5}$$

where the infinum is taken over all $(\hat{U}_n, U_n)$ pair having independent and identical distribution on

$$\left\{0, \frac{1}{n}, \frac{2}{n}, \cdots, \frac{n-1}{n}, 1\right\}$$

and $j = 2, 3, 4, \cdots$ etc.

Proof: The lower bound follows immediately from Lemma 3.

To prove the upper bound, let U_n^* be uniformly distributed over

$$\left\{0, \frac{\ell}{n}, \frac{2\ell}{n}, \cdots, \frac{k\ell}{n}, 1\right\}$$

where $\ell = \lfloor n\alpha \rfloor + 1$ and k is an integer satisfying

$$\frac{n}{n/(j-1)+1} - 1 \leq k < \frac{n}{n/(j-1)+1}.$$

(Note that

$$\frac{n}{j-1} + 1 \geq \left\lfloor \frac{n}{j-1} \right\rfloor + 1 \geq \ell.)$$

Then

$$\begin{aligned} \inf_{U_n} \Pr\{|\hat{U}_n - U_n| \leq \alpha\} &\leq \Pr\{|\hat{U}_n^* - U_n^*| \leq \alpha\} \\ &= \frac{1}{k+2} \leq \frac{1}{\frac{1}{1/(j-1)+1/n} + 1} \\ &= \frac{1}{j} + \frac{(j-1)^2}{j^2} \frac{1}{n + (j-1)/j} \\ &\leq \frac{1}{j} + \frac{1}{n+0.5}. \end{aligned}$$

□

Based on the above lemmas, we can then proceed to compute the asymptotic largest minimum distance among codewords of the following examples. It needs to be pointed out that in these examples, our objective is not to attempt to solve any related problems of practical interests, but simply to demonstrate the computation of the distance spectrum function for general readers.

Example 1: Assume that the n-tuple code alphabet is $\{0,1\}^n$. Let the n-fold distance function be defined as

$$\mu_n(\hat{x}^n, x^n) \triangleq |\#1(\hat{x}^n) - \#1(x^n)|$$

where $\#1(x^n)$ represents the number of 1's in x^n. Then

$$\begin{aligned}
&\sup_{\boldsymbol{X}} \bar{\Lambda}_{\boldsymbol{X}}(R) \\
&\quad \le \inf\left\{a \in \Re : \limsup_{n\to\infty} \sup_{X^n}\left(\Pr\left\{\frac{1}{n}|\#1(\hat{X}^n) - \#1(X^n)| > a\right\}\right)^M = 0\right\} \\
&\quad = \inf\left\{a \in \Re : \limsup_{n\to\infty}\left(1 - \inf_{X^n}\Pr\left\{\frac{1}{n}|\#1(\hat{X}^n) - \#1(X^n)| \le a\right\}\right)^M = 0\right\} \\
&\quad \le \inf\left\{a \in \Re : \limsup_{n\to\infty}\left(1 - \inf_{U}\Pr\{|\hat{U} - U| \le a\}\right)^{\exp\{nR\}} = 0\right\} \\
&\quad = \inf\left\{a \in \Re : \limsup_{n\to\infty}\left(1 - \frac{1}{\lceil 1/a\rceil}\right)^{\exp\{nR\}} = 0\right\} \\
&\quad = 0.
\end{aligned}$$

Hence, the asymptotic largest minimum distance among block codewords is zero. This conclusion is not surprising because the code with nonzero minimal distance should contain the codewords of different Hamming weights and the whole number of such words is $n+1$.

Example 2: Assume that the code alphabet is binary. Define

$$\mu_n(\hat{x}^n, x^n) \triangleq \log_2(|g_n(\hat{x}^n) - g_n(x^n)| + 1)$$

where

$$g_n(x^n) = x_{n-1}\cdot 2^{n-1} + x_{n-2}\cdot 2^{n-2} + \cdots + x_1 \cdot 2 + x_0.$$

Then

$$\begin{aligned}
&\sup_{\boldsymbol{X}} \bar{\Lambda}_{\boldsymbol{X}}(R) \\
&\quad \le \inf\left\{a \in \Re : \limsup_{n\to\infty} \sup_{X^n}\left(\Pr\left\{\frac{1}{n}\log_2(|g_n(\hat{X}^n) - g_n(X^n)| + 1) > a\right\}\right)^M = 0\right\} \\
&\quad \le \inf\left\{a \in \Re : \limsup_{n\to\infty}\left(1 - \inf_{U}\Pr\left\{|\hat{U} - U| \le \frac{2^{na}-1}{2^n - 1}\right\}\right)^{\exp\{nR\}} = 0\right\} \\
&\quad = \inf\left\{a \in \Re : \limsup_{n\to\infty}\left(1 - \frac{1}{\left\lceil \frac{2^n-1}{2^{na}-1}\right\rceil}\right)^{\exp\{nR\}} = 0\right\} \\
&\quad = 1 - \frac{R}{\log 2}. \qquad (3.8)
\end{aligned}$$

By taking $\boldsymbol{X}^*$ to be the one under which

$$U_K \triangleq g_n(X^n)/(2^n - 1)$$

has the distribution as used in the proof of the upper bound of Lemma 4, where $K \triangleq 2^n - 1$, we obtain

$$\begin{aligned}
\bar{\Lambda}_{\boldsymbol{X}^*}(R) &\ge \inf\left\{a \in \Re : \limsup_{n\to\infty}\left(1 - \frac{1}{\left\lceil \frac{2^n-1}{2^{na}-1}\right\rceil} - \frac{1}{K + 0.5}\right)^{\exp\{nR\}} = 0\right\} \\
&= \inf\left\{a \in \Re : \limsup_{n\to\infty}\left(1 - \frac{1}{\left\lceil \frac{2^n-1}{2^{na}-1}\right\rceil} - \frac{1}{2^n - 0.5}\right)^{\exp\{nR\}} = 0\right\} \\
&= 1 - \frac{R}{\log 2}.
\end{aligned}$$

This proved the achievability of (3.8).

IV. General Properties of Distance-Spectrum Function

We next address some general functional properties of $\bar{\Lambda}_{\boldsymbol{X}}(R)$ and $\underline{\Lambda}_{\boldsymbol{X}}(R)$.

Lemma 5 (General Properties of $\bar{\Lambda}_{\boldsymbol{X}}(R)$ and $\underline{\Lambda}_{\boldsymbol{X}}(R)$):

1) $\bar{\Lambda}_{\boldsymbol{X}}(R)$ and $\underline{\Lambda}_{\boldsymbol{X}}(R)$ are nonincreasing and right-continuous functions of R.
2)

$$\bar{\Lambda}_{\boldsymbol{X}}(R) \ge \bar{D}_0(\boldsymbol{X}) \triangleq \limsup_{n\to\infty} \operatorname{ess\,inf} \frac{1}{n}\mu_n(\hat{X}^n, X^n) \qquad (4.9)$$

$$\underline{\Lambda}_{\boldsymbol{X}}(R) \ge \underline{D}_0(\boldsymbol{X}) \triangleq \liminf_{n\to\infty} \operatorname{ess\,inf} \frac{1}{n}\mu_n(\hat{X}^n, X^n) \qquad (4.10)$$

where *ess inf* represents *essential infinum*.[3] In addition, equality holds for (4.9) and (4.10), respectively, when

$$R > \bar{R}_0(\boldsymbol{X}) \triangleq \limsup_{n\to\infty} -\frac{1}{n}\log\Pr\{\hat{X}^n = X^n\} \qquad (4.11)$$

and

$$R > \underline{R}_0(\boldsymbol{X}) \triangleq \liminf_{n\to\infty} -\frac{1}{n}\log\Pr\{\hat{X}^n = X^n\} \qquad (4.12)$$

provided that

$$(\forall x^n \in \mathcal{X}^n) \min_{\hat{x}^n \in \mathcal{X}^n} \mu_n(\hat{x}^n, x^n) = \mu_n(x^n, x^n) = \text{const.} \qquad (4.13)$$

[3]For a given random variable Z, its *essential infinum* is defined as

$$\operatorname{ess\,inf} Z \triangleq \sup\{z : \Pr[Z \ge z] = 1\}.$$

3)

$$\bar{\Lambda}_{\boldsymbol{X}}(R) \le \bar{D}_p(\boldsymbol{X}) \triangleq \limsup_{n\to\infty} \frac{1}{n} E[\mu_n(\hat{X}^n, X^n) \,|\, \mu_n(\hat{X}^n, X^n) < \infty], \quad \text{for } R > \bar{R}_p(\boldsymbol{X}) \quad (4.14)$$

$$\underline{\Lambda}_{\boldsymbol{X}}(R) \le \underline{D}_p(\boldsymbol{X}) \triangleq \liminf_{n\to\infty} \frac{1}{n} E[\mu_n(\hat{X}^n, X^n) \,|\, \mu_n(\hat{X}^n, X^n) < \infty], \quad \text{for } R > \underline{R}_p(\boldsymbol{X}) \quad (4.15)$$

where

$$\bar{R}_p(\boldsymbol{X}) \triangleq \limsup_{n\to\infty} -\frac{1}{n} \log \Pr\{\mu_n(\hat{X}^n, X^n) < \infty\}$$

and

$$\underline{R}_p(\boldsymbol{X}) \triangleq \liminf_{n\to\infty} -\frac{1}{n} \log \Pr\{\mu_n(\hat{X}^n, X^n) < \infty\}.$$

In addition, equality holds for (4.14) and (4.15), respectively, when $R = \bar{R}_p(\boldsymbol{X})$ and $R = \underline{R}_p(\boldsymbol{X})$, provided that $[\mu_n(\hat{X}^n, X^n) \,|\, \mu_n(\hat{X}^n, X^n) < \infty]$ has the large deviation type of behavior, i.e., for all $\delta > 0$

$$\liminf_{n\to\infty} -\frac{1}{n} \log \Pr\left\{ \frac{1}{n}(Y_n - E[Y_n \,|\, Y_n < \infty]) \le -\delta \,|\, Y_n < \infty \right\} > 0 \quad (4.16)$$

where $Y_n \triangleq \mu_n(\hat{X}^n, X^n)$.

4) For $0 < R < \bar{R}_p(\boldsymbol{X})$, $\bar{\Lambda}_{\boldsymbol{X}}(R) = \infty$. Similarly, for $0 < R < \underline{R}_p(\boldsymbol{X})$, $\underline{\Lambda}_{\boldsymbol{X}}(R) = \infty$.

Proof: Again, only the proof regarding $\bar{\Lambda}_{\boldsymbol{X}}(R)$ will be provided. The properties of $\underline{\Lambda}_{\boldsymbol{X}}(R)$ can be proved similarly.

1) Property 1 follows by definition.
2) Formula (4.9) can be proved as follows. Let

$$e_n(\boldsymbol{X}) \triangleq \operatorname{ess\,inf} \frac{1}{n} \mu_n(\hat{X}^n, X^n)$$

and, hence, $\bar{D}_0(\boldsymbol{X}) = \limsup_{n\to\infty} e_n(\boldsymbol{X})$. Observe that for any $\delta > 0$ and for infinitely many n

$$\left(\Pr\left[\frac{1}{n} \mu_n(\hat{X}^n, X^n) > \bar{D}_0(\boldsymbol{X}) - 2\delta \right] \right)^M \ge \left(\Pr\left[\frac{1}{n} \mu_n(\hat{X}^n, X^n) > e_n(\boldsymbol{X}) - \delta \right] \right)^M = 1.$$

Therefore, $\bar{\Lambda}_{\boldsymbol{X}}(R) \ge \bar{D}_0(\boldsymbol{X}) - 2\delta$ for arbitrarily small $\delta > 0$. This completes the proof of (4.9).

To prove the equality condition for (4.9), it suffices to show that for any $\delta > 0$

$$\bar{\Lambda}_{\boldsymbol{X}}(R) \le \bar{D}_0(\boldsymbol{X}) + \delta \quad \text{for } R > \limsup_{n\to\infty} -\tfrac{1}{n} \log \Pr\{\hat{X}^n \ne X^n\}. \quad (4.17)$$

By the assumption on the range of R, there exists $\gamma > 0$ such that

$$R > -\frac{1}{n} \log \Pr\{\hat{X}^n \ne X^n\} + \gamma \quad \text{for sufficiently large } n. \quad (4.18)$$

Then for those n satisfying $e_n(X^n) \le \bar{D}_0(\boldsymbol{X}) + \delta/2$ and (4.18) (of which there are sufficiently many)

$$\begin{aligned} &\left(\Pr\left\{ \frac{1}{n} \mu_n(\hat{X}^n, X^n) > \bar{D}_0(\boldsymbol{X}) + \delta \right\} \right)^M \\ &\le \left(\Pr\left\{ \frac{1}{n} \mu_n(\hat{X}^n, X^n) > e_n(X^n) + \frac{\delta}{2} \right\} \right)^M \\ &\le (\Pr\{\hat{X}^n \ne X^n\})^M \\ &= (1 - \Pr\{\hat{X}^n = X^n\})^M \end{aligned} \quad (4.19)$$

where (4.19) holds because (4.13). Consequently,

$$\limsup_{n\to\infty} \left(\Pr\left\{ \frac{1}{n} \mu_n(\hat{X}^n, X^n) > \bar{D}_0(\boldsymbol{X}) + \delta \right\} \right)^M = 0$$

which immediately implies (4.17).

3) Formula (4.14) holds trivially if $\bar{D}_p(\boldsymbol{X}) = \infty$. Thus without loss of generality, we assume that $\bar{D}_p(\boldsymbol{X}) < \infty$. Formula (4.14) can then be proved by observing that for any $\delta > 0$ and sufficiently large n

$$\begin{aligned} &\left(\Pr\left\{ \frac{1}{n} Y_n > (1+\delta)^2 \cdot \bar{D}_p(\boldsymbol{X}) \right\} \right)^M \\ &\le \left(\Pr\left\{ \frac{1}{n} Y_n > (1+\delta) \cdot \frac{1}{n} E[Y_n \,|\, Y_n < \infty] \right\} \right)^M \\ &= (\Pr\{Y_n > (1+\delta) \cdot E[Y_n \,|\, Y_n < \infty]\})^M \\ &= (\Pr\{Y_n = \infty\} + \Pr\{Y_n < \infty\} \\ &\quad \times \Pr\{Y_n > (1+\delta) \cdot E[Y_n \,|\, Y_n < \infty] \,|\, Y_n < \infty\})^M \\ &\le \left(\Pr\{Y_n = \infty\} + \Pr\{Y_n < \infty\} \frac{1}{1+\delta} \right)^M \\ &= \left(1 - \frac{\delta}{1+\delta} \Pr\{Y_n < \infty\} \right)^M \end{aligned} \quad (4.20)$$

where (4.20) follows from Markov's inequality. Consequently, for $R > \bar{R}_p(\boldsymbol{X})$

$$\limsup_{n\to\infty} \left(\Pr\left\{ \frac{1}{n} \mu_n(\hat{X}^n, X^n) > (1+\delta)^2 \cdot \bar{D}_p(\boldsymbol{X}) \right\} \right)^M = 0.$$

To prove the equality holds for (4.14) at $R = \bar{R}_p(\boldsymbol{X})$, it suffices to show the achievability of $\bar{\Lambda}_{\boldsymbol{X}}(R)$ to $\bar{D}_p(\boldsymbol{X})$ by $R \downarrow \bar{R}_p(\boldsymbol{X})$, since $\bar{\Lambda}_{\boldsymbol{X}}(R)$ is right-continuous. This can be shown as follows. For any $\delta > 0$, we note from (4.16) that there exists $\gamma = \gamma(\delta)$ such that for sufficiently large n

$$\Pr\left\{ \frac{1}{n} Y_n - \frac{1}{n} E[Y_n \,|\, Y_n < \infty] \le -\delta \,\middle|\, Y_n < \infty \right\} \le e^{-n\gamma}.$$

Therefore, for infinitely many n

$$\begin{aligned} &\left(\Pr\left\{ \frac{1}{n} Y_n > \bar{D}_p(\boldsymbol{X}) - 2\delta \right\} \right)^M \\ &\ge \left(\Pr\left\{ \frac{1}{n} Y_n > \frac{1}{n} E[Y_n \,|\, Y_n < \infty] - \delta \right\} \right)^M \\ &= (\Pr\{Y_n = \infty\} + \Pr\{Y_n < \infty\} \\ &\quad \times \Pr\left\{ \frac{1}{n} Y_n > \frac{1}{n} E[Y_n \,|\, Y_n < \infty] - \delta \,|\, Y_n < \infty \right\})^M \end{aligned}$$

$$= \left(1 - \Pr\left\{\frac{1}{n}Y_n - \frac{1}{n}E[Y_n \,|\, Y_n < \infty] \le -\delta \,|\, Y_n < \infty\right\} \cdot \Pr\{Y_n < \infty\}\right)^M$$
$$\ge (1 - e^{-n\gamma} \cdot \Pr\{Y_n < \infty\})^M.$$

Accordingly, for $\bar{R}_p(\boldsymbol{X}) < R < \bar{R}_p(\boldsymbol{X}) + \gamma$

$$\limsup_{n\to\infty}\left(\Pr\left\{\frac{1}{n}Y_n > \bar{D}_p(\boldsymbol{X}) - 2\delta\right\}\right)^M > 0$$

which, in turn, implies that

$$\bar{\Lambda}_{\boldsymbol{X}}(R) \ge \bar{D}_p(\boldsymbol{X}) - 2\delta.$$

This completes the proof of achievability of (4.14) by $R \downarrow \bar{R}_p(\boldsymbol{X})$.

4) This is an immediate consequence of

$$(\forall L > 0) \quad \left(\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n, X^n) > L\right\}\right)^M \ge (1 - \Pr\{\mu_n(\hat{X}^n, X^n) < \infty\})^M. \quad \square$$

Remarks:

- A weaker condition for (4.11) and (4.12) is

$$R > \limsup_{n\to\infty}\frac{1}{n}\log|\mathcal{S}_n| \quad \text{and} \quad R > \liminf_{n\to\infty}\frac{1}{n}\log|\mathcal{S}_n|$$

where

$$\mathcal{S}_n \triangleq \{x^n \in \mathcal{X}^n : P_{X^n}(x^n) > 0\}.$$

This indicates an expected result that when the rate is larger than $\log|\mathcal{X}|$, the asymptotic largest minimum distance among codewords remains at its smallest value $\sup_{\boldsymbol{X}} \bar{D}_0(\boldsymbol{X})$ (resp., $\sup_{\boldsymbol{X}} \underline{D}_0(\boldsymbol{X})$), which is usually zero.

- Based on Lemma 5, the general relation between $\bar{\Lambda}_{\boldsymbol{X}}(R)$ and the spectrum of

$$\frac{1}{n}\mu_n(\hat{X}^n, X^n)$$

can be illustrated as in Fig. 1, which shows that $\bar{\Lambda}_{\boldsymbol{X}}(R)$ lies asymptotically within $\operatorname{ess\,inf}(1/n)\mu(\hat{X}^n, X^n)$ and $(1/n)E[\mu_n(\hat{X}^n, X^n) \,|\, \mu_n(\hat{X}^n, X^n)]$for $\bar{R}_p(\boldsymbol{X}) < R < \bar{R}_0(\boldsymbol{X})$. Similar remarks can be made on $\underline{\Lambda}_{\boldsymbol{X}}(R)$.

On the other hand, the general curve of $\bar{\Lambda}_{\boldsymbol{X}}(R)$ (similarly for $\underline{\Lambda}_{\boldsymbol{X}}(R)$) can be plotted as shown in Fig. 2. To summarize, we remark that under fairly general situations -4pt-4pt

$$\bar{\Lambda}_{\boldsymbol{X}}(R)\begin{cases} = \infty, & \text{for } 0 < R < \bar{R}_p(\boldsymbol{X}) \\ = \bar{D}_p(\boldsymbol{X}), & \text{at } R = R_p(\boldsymbol{X}) \\ \in (\bar{D}_0(\boldsymbol{X}), \bar{D}_p(\boldsymbol{X})], & \text{for } R_p(\boldsymbol{X}) < R < R_0(\boldsymbol{X}) \\ = \bar{D}_0(\boldsymbol{X}), & \text{for } R \ge R_0(\boldsymbol{X}). \end{cases}$$

- A simple universal upper bound on the largest minimum distance among block codewords is the Plotkin bound. Its usual expression is given by [13] for which a straightforward generalization (cf. Appendix B) is

$$\sup_{\boldsymbol{X}} \limsup_{n\to\infty}\frac{1}{n}E[\mu_n(\hat{X}^n, X^n)].$$

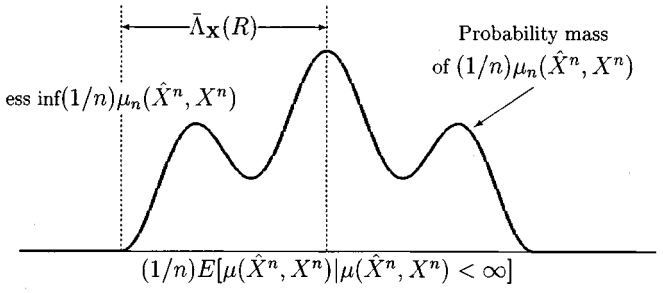

Fig. 1. $\bar{\Lambda}_{\boldsymbol{X}}(R)$ asymptotically lies between $\operatorname{ess\,inf}(1/n)\mu(\hat{X}^n, X^n)$ and $(1/n)E[\mu_n(\hat{X}^n, X^n) \,|\, \mu_n(\hat{X}^n, X^n)]$ for $\bar{R}_p(\boldsymbol{X}) < R < \bar{R}_0(\boldsymbol{X})$.

Property 3 of Lemma 5, however, provides a slightly better form for the general Plotkin bound.

We now, based on Lemma 5, calculate the distance-spectrum function of the following examples. The first example deals with the case of infinite code alphabet, and the second example derives the distance-spectrum function under unbounded generalized distance measure.

Example 3 (Continuous Code Alphabet): Let $\mathcal{X} = [0, 1)$, and let the marginal distance metric be

$$\mu(x_1, x_2) = \min\{1 - |x_1 - x_2|, |x_1 - x_2|\}.$$

Note that the metric is nothing but treating $[0, 1)$ as a circle (0 and 1 are glued together), and then to measure the shorter distance between two positions. Also, the additivity property is assumed for the n-fold distance function, i.e.,

$$\mu_n(\hat{x}^n, x^n) \triangleq \sum_{i=1}^n \mu(\hat{x}_i, x_i).$$

Using the product $\boldsymbol{X}$ of uniform distributions over $\mathcal{X}$, the sup-distance-spectrum function becomes

$$\bar{\Lambda}_{\boldsymbol{X}}(R) = \inf\left\{a \in \Re : \limsup_{n\to\infty}\left(1 - \Pr\left\{\frac{1}{n}\sum_{i=1}^n \mu(\hat{X}_i, X_i) \le a\right\}\right)^M = 0\right\}.$$

By Cramér Theorem [4]

$$\lim_{n\to\infty} -\frac{1}{n}\Pr\left\{\frac{1}{n}\sum_{i=1}^n \mu(\hat{X}_i, X_i) \le a\right\} = I_X(a)$$

where

$$I_X(a) \triangleq \sup_{t>0}\left\{-ta - \log E\left[e^{-t\cdot\mu(\hat{X}, X)}\right]\right\}$$

is the large deviation rate function.[4] Since $I_X(a)$ is convex in a, there exists a supporting line to it satisfying

$$I_X(a) = -t^* a - \log E\left[e^{-t^*\cdot\mu(\hat{X}, X)}\right]$$

[4]We take the range of supremum to be $[t > 0]$ (instead of $[t \in \Re]$ as conventional large deviation rate function does) since what concerns us here is the exponent of the *cumulative* probability mass.

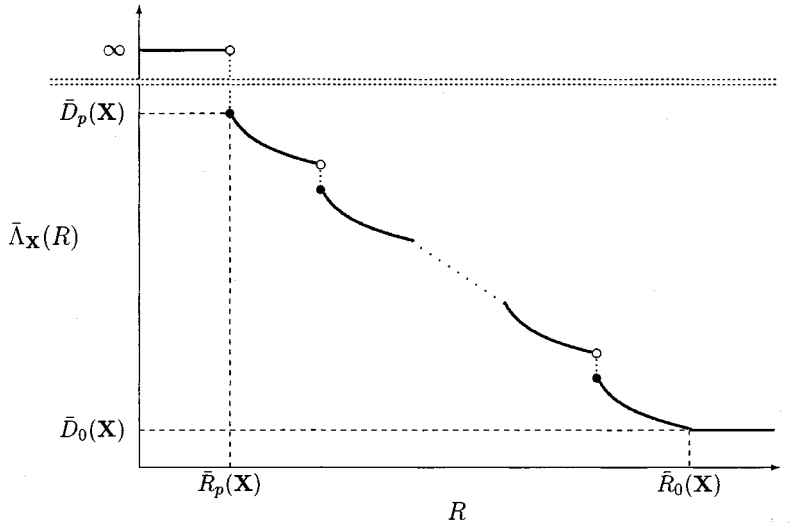

Fig. 2. General curve of $\bar{\Lambda}_{\boldsymbol{X}}(R)$.

which implies

$$a = -s^* I_X(a) - s^* \cdot \log E\left[e^{-\mu(\hat{X},X)/s^*}\right]$$

for $s^* \stackrel{\triangle}{=} 1/t^* > 0$; or, equivalently, the inverse function of $I_X(\cdot)$ is given as

$$\begin{aligned} I_X^{-1}(R) &= -s^* R - s^* \cdot \log E\left[e^{-\mu(\hat{X},X)/s^*}\right] \\ &= \sup_{s>0}\left\{-sR - s \cdot \log E\left[e^{-\mu(\hat{X},X)/s}\right]\right\} \end{aligned} \quad (4.21)$$

where[5] the last step follows from the observation that (4.21) is also a supporting line to the convex $I_X^{-1}(R)$. Consequently,

$$\begin{aligned} \bar{\Lambda}_{\boldsymbol{X}}(R) &= \inf\{a \in \Re : I_X(a) < R\} \\ &= \sup_{s>0}\{-sR - s \cdot \log[2s \cdot (1 - e^{-1/2s})]\} \end{aligned} \quad (4.22)$$

which is plotted in Fig. 3.

Also from Lemma 5, we can easily compute the marginal points of the distance-spectrum function as follows.

$$\bar{R}_0(\boldsymbol{X}) = -\log \Pr\{\hat{X} = X\} = \infty$$
$$\text{and } \bar{D}_0(\boldsymbol{X}) = \operatorname{ess\,inf} \mu(\hat{X}, X) = 0$$

$$\bar{R}_p(\boldsymbol{X}) = -\log \Pr\{\mu(\hat{X}, X) < \infty\} = 0$$
$$\text{and } \bar{D}_p(\boldsymbol{X}) = E[\mu(\hat{X}, X) \,|\, \mu(\hat{X}, X) < \infty] = \frac{1}{4}.$$

[5]One may notice the analog between the expression of the large deviation rate function $I_X(a)$ and that of the *error exponent function* [2, Theorem 4.6.4)] (or the *channel reliability exponent function* [2, Theorem 10.1.5]). Here, we demonstrate in Example 3 the basic procedure of obtaining

$$\begin{aligned} &\inf\{a \in \Re : \sup_{t>0}(-ta - \log E[e^{-tZ}]) < R\} \\ &\qquad = \sup_{s>0}\{-sR - s \cdot \log E[e^{-Z/s}]\} \end{aligned}$$

for a random variable Z so that readers do not have to refer to literatures regarding to error exponent function or channel reliability exponent function for the validity of the above equality. This equality will be used later in Examples 4 and 5, and also (5.26) and (5.27).

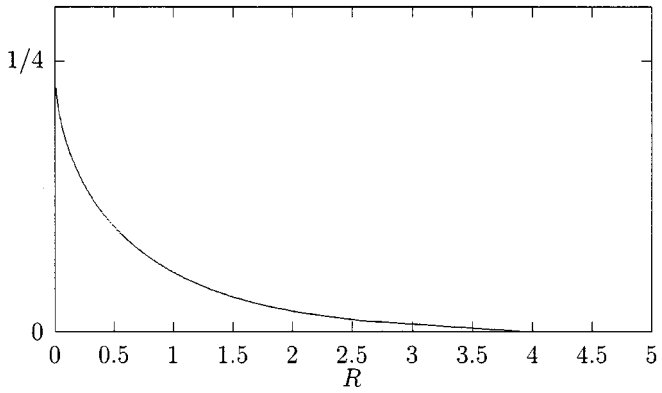

Fig. 3. Function of $\sup_{s>0}\{-sR - s\log[2s(1 - e^{-1/2s})]\}$.

Example 4: Under the case that $\mathcal{X} = \{0, 1\}$ and $\mu_n(\cdot, \cdot)$ is additive with marginal distance metric $\mu(0,0) = \mu(1,1) = 0$, $\mu(0,1) = 1$, and $\mu(1,0) = \infty$, the sup-distance-spectrum function is obtained using the product of uniform (on $\mathcal{X}$) distributions as

$$\begin{aligned} \bar{\Lambda}_{\boldsymbol{X}}(R) &= \inf\{a \in \Re : I_X(a) < R\} \\ &= \sup_{s>0}\left\{-sR - s \cdot \log \frac{2 + e^{-1/s}}{4}\right\} \end{aligned} \quad (4.23)$$

where

$$\begin{aligned} I_X(a) &\stackrel{\triangle}{=} \sup_{t>0}\left\{-ta - \log E\left[e^{-t \cdot \mu(\hat{X},X)}\right]\right\} \\ &= \sup_{t>0}\left\{-ta - \log \frac{2 + e^{-t}}{4}\right\}. \end{aligned}$$

This curve is plotted in Fig. 4. It is worth noting that there exists a region that the sup-distance-spectrum function is infinity. This is justified by deriving

$$\bar{R}_0(\boldsymbol{X}) = -\log \Pr\{\hat{X} = X\} = \log 2$$
$$\text{and } \bar{D}_0(\boldsymbol{X}) = \operatorname{ess\,inf} \mu(\hat{X}, X) = 0$$

$$\bar{R}_p(\boldsymbol{X}) = -\log \Pr\{\mu(\hat{X}, X) < \infty\} = \log \frac{4}{3}$$
$$\text{and } \bar{D}_p(\boldsymbol{X}) = E[\mu(\hat{X}, X) \,|\, \mu(\hat{X}, X) < \infty] = \frac{1}{3}.$$

One can draw the same conclusion by simply taking the derivative of (4.23) with respect to s, and obtaining that the derivative

$$-R-\log(2+e^{-1/s})-\frac{e^{-1/s}}{s(2+e^{-1/s})}+\log(4)$$

is always positive when $R < \log(4/3)$. Therefore, when $0 < R < \log(4/3)$, the distance-spectrum function is infinity.

From the above two examples, it is natural to question whether the formula of the largest minimum distance can be simplified to the *quantile function*[6] of the large deviation rate function (cf. (4.22) and (4.23)), especially when the distance functional is symmetric and additive. Note that the quantile function of the large deviation rate function is exactly the well-known Varshamov–Gilbert lower bound (cf. the next section). This inquiry then becomes to find the answer of an open question: *under what conditions is the Varshamov–Gilbert lower bound tight?* Some insight on this inquiry will be discussed in the next section.

V. General Varshamov–Gilbert Lower Bound

In this section, a general Varshamov–Gilbert lower bound will be derived directly from the distance-spectrum formulas. Conditions under which this lower bound is tight will then be explored.

Lemma 6 (Large Deviation Formulas for $\bar{\Lambda}_{\boldsymbol{X}}(R)$ and $\underline{\Lambda}_{\boldsymbol{X}}(R)$):

$$\bar{\Lambda}_{\boldsymbol{X}}(R)=\inf\{a\in\Re:\bar{\ell}_{\boldsymbol{X}}(a)<R\}$$
$$\text{and } \underline{\Lambda}_{\boldsymbol{X}}(R)=\inf\{a\in\Re:\underline{\ell}_{\boldsymbol{X}}(a)<R\}$$

where $\bar{\ell}_{\boldsymbol{X}}(a)$ and $\underline{\ell}_{\boldsymbol{X}}(a)$ are, respectively, the sup- and the inf-large deviation spectra of $(1/n)\mu_n(\hat{X}^n,X^n)$, defined as

$$\bar{\ell}_{\boldsymbol{X}}(a)\triangleq\limsup_{n\to\infty}-\frac{1}{n}\log\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n,X^n)\le a\right\}$$

and

$$\underline{\ell}_{\boldsymbol{X}}(a)\triangleq\liminf_{n\to\infty}-\frac{1}{n}\log\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n,X^n)\le a\right\}.$$

Proof: We will only provide the proof regarding $\bar{\Lambda}_{\boldsymbol{X}}(R)$. All the properties of $\underline{\Lambda}_{\boldsymbol{X}}(R)$ can be proved by following similar arguments.

Define

$$\bar{\lambda}\triangleq\inf\{a\in\Re:\bar{\ell}_{\boldsymbol{X}}(a)<R\}.$$

Then for any $\gamma>0$

$$\bar{\ell}_{\boldsymbol{X}}(\bar{\lambda}+\gamma)<R \tag{5.24}$$

and

$$\bar{\ell}_{\boldsymbol{X}}(\bar{\lambda}-\gamma)\ge R. \tag{5.25}$$

[6]Note that the usual definition [1, p. 190] of the quantile function of a nondecreasing function $F(\cdot)$ is defined as $\sup\{\theta:F(\theta)<\delta\}$. Here we adopt its dual definition for a nonincreasing function $I(\cdot)$ as $\inf\{a:I(a)<R\}$. Remark that if $F(\cdot)$ is strictly increasing (resp., $I(\cdot)$ is strictly decreasing), then the quantile is nothing but the inverse of $F(\cdot)$ (resp., $I(\cdot)$).

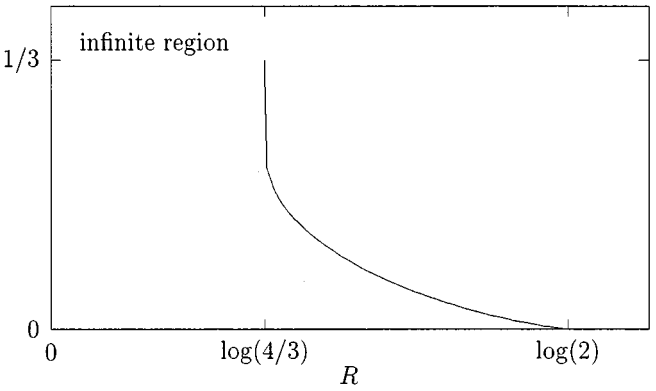

Fig. 4. Function of $\sup_{s>0}\{-sR-s\log[(2+e^{-1/s})/4]\}$.

Inequality (5.24) ensures the existence of $\delta=\delta(\gamma)>0$ such that for sufficiently large n

$$\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n,X^n)\le\bar{\lambda}+\gamma\right\}\ge e^{-n(R-\delta)}$$

which, in turn, implies

$$\limsup_{n\to\infty}\left(\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n,X^n)>\bar{\lambda}+\gamma\right\}\right)^M \le\limsup_{n\to\infty}\left(1-e^{-n(R-\delta)}\right)^{e^{nR}}=0.$$

Hence, $\bar{\Lambda}_{\boldsymbol{X}}(R)\le\bar{\lambda}+\gamma$. On the other hand, (5.25) implies the existence of subsequence $\{n_j\}_{j=1}^{\infty}$ satisfying

$$\lim_{j\to\infty}-\frac{1}{n_j}\log\Pr\left\{\frac{1}{n_j}\mu_{n_j}(\hat{X}^{n_j},X^{n_j})\le\bar{\lambda}-\gamma\right\}\ge R$$

which, in turn, implies

$$\limsup_{n\to\infty}\left(\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n,X^n)>\bar{\lambda}-\gamma\right\}\right)^M \ge\limsup_{j\to\infty}\left(1-e^{-n_jR}\right)^{e^{n_jR}}=e^{-1}.$$

Accordingly, $\Lambda_{\boldsymbol{X}}(R)\ge\bar{\lambda}-\gamma$. Since γ is arbitrary, the lemma therefore holds. □

The above lemma confirms that the distance spectrum function $\bar{\Lambda}_{\boldsymbol{X}}(\cdot)$ (resp., $\underline{\Lambda}_{\boldsymbol{X}}(\cdot)$) is exactly the quantile of the sup-(resp., inf-)large deviation spectrum of

$$\{(1/n)\mu_n(\hat{X}^n,X^n)\}_{n>1}.$$

Thus if the large deviation spectrum is known, so is the distance spectrum function.

By the generalized Gärtner–Ellis upper bound derived in [6, Theorem 2.1], we obtain

$$\bar{\ell}_{\boldsymbol{X}}(a)\ge\inf_{[x\le a]}\underline{I}_{\boldsymbol{X}}(x)=\underline{I}_{\boldsymbol{X}}(a)$$
$$\text{and}\quad \underline{\ell}_{\boldsymbol{X}}(a)\ge\inf_{[x\le a]}\bar{I}_{\boldsymbol{X}}(x)=\bar{I}_{\boldsymbol{X}}(a)$$

where the equalities follow from the convexity (and hence, continuity and strict decreasing) of

$$\underline{I}_{\boldsymbol{X}}(x)\triangleq\sup_{[\theta<0]}[\theta x-\underline{\varphi}_{\boldsymbol{X}}(\theta)] \text{ and } \bar{I}_{\boldsymbol{X}}(x)\triangleq\sup_{[\theta<0]}[\theta x-\bar{\varphi}_{\boldsymbol{X}}(\theta)]$$

and

$$\underline{\varphi}_{\boldsymbol{X}}(\theta) \triangleq \liminf_{n\to\infty} \frac{1}{n} \log E\left[e^{\theta\cdot\mu_n(\hat{X}^n,X^n)}\right]$$

$$\text{and } \bar{\varphi}_{\boldsymbol{X}}(\theta) \triangleq \limsup_{n\to\infty} \frac{1}{n} \log E\left[e^{\theta\cdot\mu_n(\hat{X}^n,X^n)}\right].$$

Based on these observations, the relation between the distance-spectrum expression and the Varshamov–Gilbert bound can be described as follows.

Corollary 2:

$$\sup_{\boldsymbol{X}} \bar{\Lambda}_{\boldsymbol{X}}(R) \geq \sup_{\boldsymbol{X}} \bar{G}_{\boldsymbol{X}}(R) \quad \text{and} \quad \sup_{\boldsymbol{X}} \underline{\Lambda}_{\boldsymbol{X}}(R) \geq \sup_{\boldsymbol{X}} \underline{G}_{\boldsymbol{X}}(R)$$

where

$$\bar{G}_{\boldsymbol{X}}(R) \triangleq \inf\{a \in \Re : \underline{I}_{\boldsymbol{X}}(a) < R\} = \sup_{s>0}[-sR - s\cdot\underline{\varphi}_{\boldsymbol{X}}(-1/s)] \tag{5.26}$$

and

$$\underline{G}_{\boldsymbol{X}}(R) \triangleq \inf\{a \in \Re : \bar{I}_{\boldsymbol{X}}(a) < R\} = \sup_{s>0}[-sR - s\cdot\bar{\varphi}_{\boldsymbol{X}}(-1/s)]. \tag{5.27}$$

Some remarks on the Varshamov–Gilbert bound obtained above are given below.

Remarks:

- One can easily see from [2, p. 400], where the Varshamov–Gilbert bound is given under Bhattacharyya distance and finite code alphabet, that
$$\sup_{\boldsymbol{X}} \bar{G}_{\boldsymbol{X}}(R) \quad \text{and} \quad \sup_{\boldsymbol{X}} \underline{G}_{\boldsymbol{X}}(R)$$
are indeed the generalization of the conventional Varshamov–Gilbert bound.
- Since $0 \leq \exp\{-\mu_n(\hat{x}^n, x^n)/s\} \leq 1$ for $s > 0$, the function $\exp\{-\mu(\cdot,\cdot)/s\}$ is always integrable. Hence, (5.26) and (5.27) can be evaluated under any nonnegative measurable function $\mu_n(\cdot,\cdot)$. In addition, no assumption on the alphabet space $\mathcal{X}$ is needed in deriving the lower bound. Its full generality can be displayed using, again, Examples 3 and 4, which result in exactly the same curves as shown in Figs. 3 and 4.
- Observe that $\bar{G}_{\boldsymbol{X}}(R)$ and $\underline{G}_{\boldsymbol{X}}(R)$ are both the pointwise supremum of a collection of affine functions, and hence, they are both convex, which immediately implies their continuity and strict decreasing property on the interior of their domains, i.e.,
$$\{R : \bar{G}_{\boldsymbol{X}}(R) < \infty\} \quad \text{and} \quad \{R : \underline{G}_{\boldsymbol{X}}(R) < \infty\}.$$
However, as pointed out in [6], $\bar{\ell}_{\boldsymbol{X}}(\cdot)$ and $\underline{\ell}_{\boldsymbol{X}}(\cdot)$ are not necessarily convex, which, in turn, hints the possibility of yielding nonconvex $\bar{\Lambda}_{\boldsymbol{X}}(\cdot)$ and $\underline{\Lambda}_{\boldsymbol{X}}(\cdot)$. This clearly indicates that the Varshamov–Gilbert bound is not tight whenever the asymptotic largest minimum distance among codewords is nonconvex.

  An immediate improvement from [6] to the Varshamov–Gilbert bound is to employ the *twisted* large deviation rate function (instead of $\bar{I}_{\boldsymbol{X}}(\cdot)$)
$$\bar{J}_{\boldsymbol{X},h}(x) \triangleq \sup_{\{\theta\in\Re:\bar{\varphi}_{\boldsymbol{X}}(\theta;h)>-\infty\}} [\theta\cdot h(x) - \bar{\varphi}_{\boldsymbol{X}}(\theta;h)]$$
and yields a (potentially) nonconvex Varshamov–Gilbert-type bound, where $h(\cdot)$ is a continuous real-valued function, and
$$\bar{\varphi}_{\boldsymbol{X}}(\theta;h) \triangleq \limsup_{n\to\infty} \frac{1}{n} \log E\left[e^{n\cdot\theta\cdot h(\mu_n(\hat{X}^n,X^n)/n)}\right].$$
The question of how to find a proper $h(\cdot)$ for such improvement is beyond the scope of this paper and hence is deferred for further study.
- We now demonstrate that $\bar{\Lambda}_{\boldsymbol{X}}(R) > \bar{G}_{\boldsymbol{X}}(R)$ by a simple example.

Example 5: Assume binary code alphabet $\mathcal{X} = \{0,1\}$, and n-fold Hamming distance

$$\mu_n(\hat{x}^n, x^n) = \sum_{i=1}^{n} \mu(\hat{x}_i, x_i).$$

Define a measurable function as follows:

$$\hat{\mu}_n(\hat{x}^n, x^n) \triangleq \begin{cases} 0, & \text{if } 0 \leq \mu_n(\hat{x}^n, x^n) < \alpha n \\ \alpha n, & \text{if } \alpha n \leq \mu_n(\hat{x}^n, x^n) < 2\alpha n \\ \infty, & \text{if } 2\alpha n \leq \mu_n(\hat{x}^n, x^n) \end{cases}$$

where $0 < \alpha < 1/2$ is a universal constant. Let $\boldsymbol{X}$ be the product of uniform distributions over $\mathcal{X}$, and let $Y_i \triangleq \mu(\hat{X}_i, X_i)$ for $1 \leq i \leq n$.

Then

$$\begin{aligned}
\underline{\varphi}_{\boldsymbol{X}}(-1/s) &= \liminf_{n\to\infty} \frac{1}{n} \log E\left[e^{-\hat{\mu}_n(\hat{X}^n,X^n)/s}\right] \\
&= \liminf_{n\to\infty} \frac{1}{n} \log\left[\Pr\left(0 \leq \frac{\mu_n(\hat{X}^n,X^n)}{n} < \alpha\right) + \Pr\left(\alpha \leq \frac{\mu_n(\hat{X}^n,X^n)}{n} < 2\alpha\right) e^{-\alpha n/s}\right] \\
&= \liminf_{n\to\infty} \frac{1}{n} \log\left[\Pr\left(0 \leq \frac{Y_1+\cdots+Y_n}{n} < \alpha\right) + \Pr\left(\alpha \leq \frac{Y_1+\cdots+Y_n}{n} < 2\alpha\right) e^{-\alpha n/s}\right] \\
&= \max\left\{-I_Y(\alpha),\ -I_Y(2\alpha) - \frac{\alpha n}{s}\right\}.
\end{aligned}$$

where

$$I_Y(\alpha) \triangleq \sup_{s>0}\{s\alpha - \log[(1+e^s)/2]\}$$

(which is exactly the large deviation rate function of $(1/n)Y^n$). Hence

$$\begin{aligned}
&\bar{G}_{\boldsymbol{X}}(R) \\
&= \sup_{s>0}\left[-sR - s\cdot\max\left\{-I_Y(\alpha), -I_Y(2\alpha) - \frac{\alpha n}{s}\right\}\right] \\
&= \sup_{s>0} \min\{s[I_Y(\alpha)-R],\ s[I_Y(2\alpha)-R]+\alpha\} \\
&= \begin{cases} \infty, & \text{for } 0 \leq R < I_Y(2\alpha) \\ \dfrac{\alpha}{I_Y(\alpha)-I_Y(2\alpha)}(I_Y(\alpha)-R), & \text{for } I_Y(2\alpha) \leq R < I_Y(\alpha) \\ 0, & \text{for } R \geq I_Y(\alpha). \end{cases}
\end{aligned}$$

We next derive $\bar{\Lambda}_{\boldsymbol{X}}(R)$. Since

$$\Pr\left(\frac{1}{n}\hat{\mu}_n(\hat{X}^n, X^n) \le a\right) = \begin{cases} \Pr\left(0 \le \dfrac{Y_1+\cdots+Y_n}{n} < \alpha\right), & \text{for } 0 \le a < \alpha \\ \Pr\left(0 \le \dfrac{Y_1+\cdots+Y_n}{n} < 2\alpha\right), & \text{for } \alpha \le a < 2\alpha \end{cases}$$

we obtain

$$\bar{\Lambda}_{\boldsymbol{X}}(R) = \begin{cases} \infty, & \text{if } 0 \le R < I_Y(2\alpha) \\ \alpha, & \text{if } I_Y(2\alpha) \le R < I_Y(\alpha) \\ 0, & \text{if } I_Y(\alpha) \le R. \end{cases}$$

Consequently,

$$\bar{\Lambda}_{\boldsymbol{X}}(R) > \bar{G}_{\boldsymbol{X}}(R)$$

for $I_Y(2\alpha) < R < I_Y(\alpha)$. □

- One of the problems that remain open in the combinatorial coding theory is the tightness of the asymptotic Varshamov–Gilbert bound for the binary code and the Hamming distance [3, p. vii]. As mentioned in Section I, it is already known that the asymptotic Varshamov–Gilbert bound is in general not tight, e.g., for algebraic-geometric code with large code alphabet size and Hamming distance. Example 5 provides another example to confirm the untightness of the asymptotic Varshamov–Gilbert bound for simple binary code and quantized Hamming measure.

  By the generalized Gärtner–Ellis lower bound derived in [6, Theorem 2.1], we conclude that

$$\bar{\ell}_{\boldsymbol{X}}(R) = \underline{I}_{\boldsymbol{X}}(R) \quad (\text{or equivalently } \bar{\Lambda}_{\boldsymbol{X}}(R) = \bar{G}_{\boldsymbol{X}}(R))$$

  if

$$[\bar{D}_0(\boldsymbol{X}), \bar{D}_p(\boldsymbol{X})] \in \bigcup_{\theta<0} \left[\limsup_{t\downarrow 0} \frac{\underline{\varphi}_{\boldsymbol{X}}(\theta+t) - \underline{\varphi}_{\boldsymbol{X}}(\theta)}{t}, \liminf_{t\downarrow 0} \frac{\underline{\varphi}_{\boldsymbol{X}}(\theta) - \underline{\varphi}_{\boldsymbol{X}}(\theta-t)}{t}\right]. \tag{5.28}$$

  Note that although (5.28) guarantees that $\bar{\Lambda}_{\boldsymbol{X}}(R) = \bar{G}_{\boldsymbol{X}}(R)$, it dose not by any means ensure the tightness of the Varshamov–Gilbert bound. An additional assumption needs to be made, which is summarized in the next observation.

Observation 1: If there exists an $\tilde{\boldsymbol{X}}$ such that

$$\sup_{\boldsymbol{X}} \bar{\Lambda}_{\boldsymbol{X}}(R) = \bar{\Lambda}_{\tilde{\boldsymbol{X}}}(R)$$

and (5.28) holds for $\tilde{\boldsymbol{X}}$, then the asymptotic Varshamov–Gilbert lower bound is tight.

The problem in applying the above observation is the difficulty in finding the optimizing process $\boldsymbol{X}$. However, it does provide an alternative to prove the tightness of the asymptotic Varshamov–Gilbert bound. Instead of finding an upper bound to meet the lower bound, one could show that (5.28) holds for a fairly general class of random-code generating processes, which the optimization process surely lies in.

VI. Bounds for Specific Block Coding Schemes

A long-standing problem in coding theory is to find codes that have as many codewords as possible while their code length and minimum distance are given [3], [20], [22]. Many previous works were devoted to find bounds of the largest size that any code can achieve [12], [15], [16], [18], [23], [31]. Here, we are interested in the behavior of codes when their code lengths tend to infinity.

Enforced by the new formula established in Section II, some known lower bounds on the largest minimum distance of a few specific codes can be rederived. Comparisons with the known results devoted to the same problem are made next.

A. Average Number of Codewords in a Sphere

We first relate our formula to the average number of codewords in a sphere, which is a quantity frequently used in combinatorial coding techniques.

Let $\mathcal{S}_n \subseteq \mathcal{X}^n$, and $\mu_n(\cdot,\cdot)$ be a nonnegative integer-valued distance function defined over $\mathcal{S}_n \times \mathcal{S}_n$. We want to find a code $\mathfrak{C}_n$, a subset of $\mathcal{S}_n$, with the largest minimum distance at least na. Define

$$V_r(\hat{x}^n) \triangleq |\{x^n \in \mathcal{S}_n : \mu_n(\hat{x}^n, x^n) \le r\}|$$

and the average volume

$$V_r^{\text{avg}} \triangleq \frac{1}{|\mathcal{S}_n|} \sum_{\hat{x}^n \in \mathcal{S}_n} V_r(\hat{x}^n).$$

Let the generic distribution P_{X^n} be uniformly distributed over $\mathcal{S}_n$, i.e.,

$$P_{X^n}(x^n) = \frac{1}{|\mathcal{S}_n|}, \qquad \text{for all } x^n \in \mathcal{S}_n.$$

Then

$$\begin{aligned} &\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n, X^n) > a\right\} \\ &\quad = \sum_{\hat{x}^n \in \mathcal{S}_n} P_{X^n}(\hat{x}^n) P_{X^n}\{x^n \in \mathcal{S}_n : \mu_n(\hat{x}^n, x^n) > na\} \\ &\quad = \sum_{\hat{x}^n \in \mathcal{S}_n} P_{X^n}(\hat{x}^n)\left(1 - \frac{V_{na}(\hat{x}^n)}{|\mathcal{S}_n|}\right) \\ &\quad = 1 - \sum_{\hat{x}^n \in \mathcal{S}_n} \frac{V_{na}(\hat{x}^n)}{|\mathcal{S}_n||^2} \\ &\quad = 1 - \frac{1}{|\mathcal{S}_n|/V_{na}^{\text{avg}}}. \end{aligned}$$

Thus for $\delta > 0$ arbitrarily small

$$\begin{aligned} &\limsup_{n\to\infty} \frac{d_{n,M}}{n} \\ &\quad \ge \sup_{\boldsymbol{X}} \bar{\Lambda}_{\boldsymbol{X}}(R+\delta) \\ &\quad \ge \inf\left\{a \in \Re : \limsup_{n\to\infty} \cdot \left(\Pr\left\{\frac{1}{n}\mu_n(\hat{X}^n, X^n) > a\right\}\right)^{Me^{n\delta}} = 0\right\} \end{aligned}$$

$$= \inf\left\{a \in \Re : \limsup_{n\to\infty}\left(1 - \frac{1}{|\mathcal{S}_n|/V_{na}^{\text{avg}}}\right)^{e^{n(R+\delta)}} = 0\right\}$$
$$\geq \inf\left\{a \in \Re : \limsup_{n\to\infty} \frac{|\mathcal{S}_n|}{V_{na}^{\text{avg}} e^{n(R+\delta)}} \leq 1\right\}.$$

The above derivation indeed indicates that a lower bound to the asymptotic largest minimum distance can be established by simply calculating $|\mathcal{S}_n|/V_{na}^{\text{avg}}$. In the sequel, we will then proceed to derive the corresponding lower bounds for a few specific coding schemes.

B. Lower Bounds on the Largest Minimum Distance for Some Specific Codes

For the sake of simplicity, the Hamming distance is assumed in this subsection, except when otherwise stated.

1) Simple Binary Code (SBC): When $\mathcal{S}_n = \{0,1\}^n$, the average volume becomes[7]

$$V_{na}^{\text{avg}} = \sum_{i=0}^{na} \binom{n}{i}.$$

A known fact to this quantity is that the exponent of V_{na}^{avg} is $h_b(a)$, i.e.,

$$\lim_{n\to\infty} \frac{1}{n} \log V_{na}^{\text{avg}} = h_b(a)$$

where

$$h_b(a) \triangleq -a\log a - (1-a)\log(1-a), \qquad 0 \leq a \leq 1$$

is the binary entropy function. Therefore,

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{d_{n,M}}{n} \\
&\quad \geq \sup_{\delta>0} \inf\left\{a \in \Re : \limsup_{n\to\infty} \frac{|\mathcal{S}_n|}{V_{na}^{\text{avg}} e^{n(R+\delta)}} \leq 1\right\} \\
&\quad = \sup_{\delta>0} \inf\left\{a \in \Re : \limsup_{n\to\infty} \frac{e^{n\log(2)}}{e^{nh_b(a)} e^{n(R+\delta)}} \leq 1\right\} \\
&\quad = \sup_{\delta>0} \inf\{a \in \Re : h_b(a) > \log(2) - R - \delta\} \\
&\quad = d_{\text{SBC}} \qquad (6.29)
\end{aligned}$$

where d_{SBC} satisfies $R = \log(2) - h_b(d_{\text{SBC}})$, which is exactly the Varshamov–Gilbert bound for simple binary codes.

2) Constant-Weight Codes (CWC): Let us now turn to the constant-weight codes.

For $x^n \in \{0,1\}^n$, let weight (x^n) be the number of 1's in x^n. Define

$$\mathcal{S}_n \triangleq \{x^n \in \{0,1\}^n : \text{weight}(x^n) = w\}.$$

For any two codewords x_1^n and x_2^n in $\mathcal{S}_n$, for which their weights (1's) coincide with each other in exactly d positions, the Hamming distance between x_1^n and x_2^n is equal to $2(w-d)$. Observe that the total number of codewords, whose Hamming distance with respect to x_1^n equals $2(w-d)$, is

$$\binom{w}{d}\binom{n-w}{w-d}.$$

Therefore,

$$V_{na}^{\text{avg}} = \sum_{i=0}^{na/2} \binom{w}{w-i}\binom{n-w}{i} = \sum_{i=0}^{na/2} \binom{w}{i}\binom{n-w}{i}.$$

[7]Without loss of generality, na can be treated as an integer.

Let $\lim_{n\to\infty}(w/n) = \upsilon$. Then by using typical asymptotic approximations for binomial coefficients, we obtain

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{|\mathcal{S}_n|}{V_{na}^{\text{avg}}} = h_b(\upsilon) - \upsilon \cdot h_b\left(\frac{a}{2\upsilon}\right) - (1-\upsilon)\cdot h_b\left(\frac{a}{2(1-\upsilon)}\right).$$

Following the same procedure as (6.29), we get

$$\limsup_{n\to\infty} \frac{d_{n,M}}{n} \geq d_{\text{CWC}}$$

where d_{CWC} satisfies

$$R = h_b(\upsilon) - \upsilon \cdot h_b\left(\frac{d_{\text{CWC}}}{2\upsilon}\right) - (1-\upsilon)\cdot h_b\left(\frac{d_{\text{CWC}}}{2(1-\upsilon)}\right)$$

which is equivalent to the result given in [12].

3) Codes of Correcting Arbitrary Additive Noises (CAN): The problem of constructing the maximal codes for correction of an arbitrary set of additive errors can be summarized as follows [10], [11].

Let $\mathcal{S}_n = \{0,1\}^n$ be a vector space of dimension n over GF (2). Denote a noise set by $\mathcal{K}_n \in \{0,1\}^n$, and assume that the all-zero elements $\mathbf{0}$ always belong to $\mathcal{K}_n$. Then a code $\mathfrak{C}_n \in \{0,1\}^n$ is said to have the capability of correcting the noise $\mathcal{K}_n$ if, and only if, for every $x^n \neq \hat{x}^n$ in $\mathfrak{C}_n$, $x^n \oplus k^n \neq \hat{x}^n \oplus \hat{k}^n$ for all k^n and $\hat{k}^n$ in $\mathcal{K}_n$, where "$\oplus$" represents the n-fold addition operator over $\mathcal{S}_n$. This condition is indeed equivalent to $(\mathfrak{C}_n \ominus \mathfrak{C}_n) \cap (\mathcal{K}_n \ominus \mathcal{K}_n) = \{\mathbf{0}\}$, where

$$\mathfrak{C}_n \ominus \mathfrak{C}_n \triangleq \{x^n \ominus \hat{x}^n : x^n, \hat{x}^n \in \mathfrak{C}_n\}$$

and "$\ominus$" is the reverse operator to "$\oplus$." Then the problem of finding the maximal codes which correct arbitrary, but fixed, types of noise, is exactly the one to locate the largest $\mathfrak{C}_n = \mathfrak{C}_n^*$ that corrects the noise $\mathcal{K}_n$. An interesting query that follows the previous problem will be the determination of the asymptotic behavior of $(1/n)\log|\mathfrak{C}_n^*|$ for a sequence of given noise sets $\{\mathcal{K}_n\}_{n\geq 1}$.

This query can be transformed to the problem of finding the largest minimum distance among codewords by defining

$$\mu_n(\hat{x}^n, x^n) \triangleq \begin{cases} n, & \text{if } (\hat{x}^n \oplus \mathcal{K}_n) \cap (x^n \oplus \mathcal{K}_n) = \text{empty set} \\ 0, & \text{otherwise} \end{cases}$$

where

$$x^n \oplus \mathcal{K}_n \triangleq \{x^n \oplus k^n : k^n \in \mathcal{K}_n\}.$$

With this definition, the query then reduces to finding the largest R such that

$$\liminf_{n\to\infty} \frac{d_{n,M=e^{nR}}}{n} = 1 > 0.$$

A lower bound for this largest R can then be obtained through the following procedure.

For $a \in [0,1)$

$$\begin{aligned}
V_{na}(\hat{x}^n) &= |\{x^n \in \mathcal{S}_n : \mu_n(\hat{x}^n, x^n) \leq na\}| \\
&= |\{x^n \in \mathcal{S}_n : x^n \oplus k^n = \hat{x}^n \oplus \hat{k}^n \\
&\qquad \text{for some } k^n, \hat{k}^n \in \mathcal{K}_n\}| \\
&= |\{x^n \in \mathcal{S}_n : x^n = \hat{x}^n \oplus \hat{k}^n \ominus k^n \\
&\qquad \text{for some } k^n, \hat{k}^n \in \mathcal{K}_n\}| \\
&= |\{\hat{x}^n \oplus (\hat{k}^n \ominus k^n) \text{:for some } k^n, \hat{k}^n \in \mathcal{K}_n\}|
\end{aligned}$$

$$\le |\{\hat{k}^n \ominus k^n\text{: for } k^n, \hat{k}^n \in \mathcal{K}_n\}|$$
$$= |\mathcal{K}_n \ominus \mathcal{K}_n|$$

which implies

$$V_{na}^{\text{avg}} \le |\mathcal{K}_n \ominus \mathcal{K}_n|.$$

Also note that for $a \ge 1, V_{na}(\hat{x}^n) = 2^n$. Thus

$$\frac{|\mathcal{S}_n|}{V_{na}^{\text{avg}}} \begin{cases} \ge \dfrac{2^n}{|\mathcal{K}_n \ominus \mathcal{K}_n|}, & \text{for } a \in [0,1) \\ = 1, & \text{for } a \ge 1. \end{cases}$$

Since

$$\liminf_{n\to\infty} \frac{d_{n,M}}{n}$$
$$\ge \sup_{\delta>0} \inf \left\{ a \in \Re : \liminf_{n\to\infty} \frac{|\mathcal{S}_n|}{V_{na}^{\text{avg}} e^{n(R+\delta)}} \le 1 \right\}$$
$$\ge \sup_{\delta>0} \inf \left\{ a \in [0,1) : \liminf_{n\to\infty} \frac{e^{n\log(2)}}{|\mathcal{K}_n \ominus \mathcal{K}_n| e^{n(R+\delta)}} \le 1 \right\}$$

a sufficient condition for $\liminf_{n\to\infty} d_{n,M}/n > 0$ is

$$(\forall \delta > 0) \quad \liminf_{n\to\infty} \frac{e^{n\log(2)}}{|\mathcal{K}_n \ominus \mathcal{K}_n| e^{n(R+\delta)}} > 1. \tag{6.30}$$

It can be verified that (6.30) is valid if

$$R < \log(2) - \limsup_{n\to\infty} \frac{1}{n} \log |\mathcal{K}_n \ominus \mathcal{K}_n|. \tag{6.31}$$

The right-hand-size of (6.31) then provides a lower bound to

$$\limsup_{n\to\infty} (1/n) \log |\mathfrak{C}_n^*|.$$

4) Runlength-Limited Codes (RLC): The asymptotic value of $|\mathcal{S}_n|/V_{na}^{\text{avg}}$ for a runlenth-limited code has been investigated in [18], [23], in which an upper bound on V_{na}^{avg} is characterized by the generating function of pairwise distances in $\mathcal{S}_n$. With their results, a lower bound for the asymptotic largest minimum distance among codewords can be established through the same procedure as used in the previous three codes. Details are therefore omitted here.

C. Alternative Transformation for Problem of Finding the Largest Minimum Distance Among Codewords

The results obtained in terms of a uniform generic distribution over $\mathcal{S}_n$ in the previous subsections basically show no improvements over the known ones. A natural query following this observation is whether or not a better bound can be achieved in terms of a more general class of generic distributions.

Recall that the determination of the asymptotic largest minimum distance among codewords actually involves an optimization of the code selecting process. Specifically

$$\limsup_{n\to\infty} \frac{d_{n,M=e^{nR}}}{n}$$
$$= \sup_{\boldsymbol{X}} \bar{\Lambda}_{\boldsymbol{X}}(R)$$
$$= \sup_{\boldsymbol{X}} \inf \left\{ a \in \Re : \limsup_{n\to\infty} \left(\Pr\left\{ \frac{1}{n} \mu_n(\hat{X}^n, X^n) > a \right\} \right)^M = 0 \right\}$$

except for countably many R. A straightforward approach to determine the quantity of $\limsup_{n\to\infty} d_{n,M=e^{nR}}/n$ is to directly optimize the code selecting process, as we did in Section III. We, however, find that it may be advantageous by converting the determination of $\limsup_{n\to\infty} d_{n,M=e^{nR}}/n$ into a graph problem described below.

Let $\mathcal{G}$ be a graph whose vertices are the element of $\mathcal{S}_n$. Two vertices $\hat{x}^n$ and x^n are connected by an edge, denoted by $e(\hat{x}^n, x^n)$, if, and only if, $\mu_n(\hat{x}^n, x^n) \le na$, where $\mu_n(\cdot,\cdot)$ is the Hamming distance. Denote the set of all edges by $\mathcal{E}$. Then the number of edges in $\mathcal{E}$ (including loops on vertices) is

$$\frac{1}{2} \sum_{\hat{x}^n \in \mathcal{S}_n} [V_{na}(\hat{x}^n) + 1] = \frac{1}{2} \left(|\mathcal{S}_n| V_{na}^{\text{avg}} + |\mathcal{S}_n| \right).$$

By assigning vertex x^n a probability weight $P_{X^n}(x^n)$, we obtain

$$\Pr\left\{ \frac{1}{n} \mu_n(\hat{X}^n, X^n) \le a \right\}$$
$$= 2 \sum_{e(\hat{x}^n, x^n) \in \mathcal{E}} P_{X^n}(\hat{x}^n) P_{X^n}(x^n) - \sum_{x^n \in \mathcal{G}} P_{X^n}^2(x^n).$$

One can then analyze the original probabilistic optimization problem through an alternative problem setting over a graph.

VII. Concluding Remarks and Future Work

In this paper, we use the distance-spectrum methodology to derive the formulas of the largest minimum distance of deterministic block codes. The major advantage of this method is that no assumptions on code alphabet and "distance" measure between codewords are needed. Besides, the concept behind this method is quite simple.

We also address a general Varshamov–Gilbert lower bound, and remark on the sufficient condition under which it is tight. As for the open question on its tightness under binary codes and Hamming distance, we conjecture that it might be profitable to consider the distance spectrum, and carry out estimations either directly through the probabilistic optimization (as in Section III) or by means of a graph developing. It would be interesting to have a theory along these lines.

Appendix A
Alternative Proof of General Varshamov–Gilbert Lower Bound

The idea employed below has actually appeared before in the literature [3, pp. 9–11]. As a result of the simple probabilistic method, the well-known Varshamov–Gilbert lower bound can be obtained for generalized distance function (no necessarily additive, symmetric, and bounded) and infinite code alphabet.

Lemma A.1: For all $t < 0$, there exists a codebook $\mathfrak{C}_{n,M/2}$ such that the minimum distance satisfies

$$\min_{0 \le m \le (M/2-1)} d_m(\mathfrak{C}_{n,M/2}) > \frac{1}{t} \log(2E[e^{tD_m}])$$

where D_m, $0 \le m \le (M-1)$, are defined for the random codebook of block length n and size M.

Proof: Let $\mathbf{1}(\mathcal{A})$ be the indicator function of a set $\mathcal{A}$, and let

$$\phi_m \triangleq \mathbf{1}(e^{tD_m} < 2E[e^{tD_m}]).$$

Then, the Chebyshev inequality yields

$$E[\phi_m] \geq \frac{1}{2}.$$

Therefore,

$$E\left[\sum_{m=0}^{M-1} \phi_m\right] = \sum_{m=0}^{M-1} E[\phi_m] > \frac{M}{2}$$

which implies that among all the possible selections, there must exist a codebook in which $M/2$ codewords satisfy $\phi_m = 1$, i.e., $e^{td_m} < 2E[e^{t\cdot D_m}]$ for at least $\dfrac{M}{2}$ codewords in this codebook. The collection of these $M/2$ codewords is a desired codebook. □

Lemma A.2: Let each codeword be independently selected through the distribution P_{X^n}. For any $s > 0$ and $0 \leq m \leq (M-1)$

$$\liminf_{n\to\infty} \left\{-s \cdot \frac{1}{n} \log(E[e^{-D_m/s}])\right\}$$

and

$$\limsup_{n\to\infty} \left\{-s \cdot \frac{1}{n} \log(E[e^{-D_m/s}])\right\}$$

where

$$\bar{\varphi}_{\boldsymbol{X}}(\theta) \triangleq \limsup_{n\to\infty} \frac{1}{n} \log E\left[e^{\theta\cdot\mu_n(\hat{X}^n, X^n)}\right]$$

$$\underline{\varphi}_{\boldsymbol{X}}(\theta) \triangleq \liminf_{n\to\infty} \frac{1}{n} \log E\left[e^{\theta\cdot\mu_n(\hat{X}^n, X^n)}\right],$$

and

$$\boldsymbol{X} = \{X^n = (X_1^{(n)}, \cdots, X_n^{(n)})\}_{n\geq 1}$$

is a given triangular-array distribution.

Proof: Let $\boldsymbol{C}_m^{(n)}$ denote the mth randomly selected codeword for block length n. From the definition of D_m, we have

$$\begin{aligned}
E[e^{-D_m/s}] &= E\left[\exp\left\{-\frac{1}{s} \min_{\substack{0\leq \hat{m}<M-1 \\ \hat{m}\neq m}} \mu_n\left(\boldsymbol{C}_{\hat{m}}^{(n)}, \boldsymbol{C}_m^{(n)}\right)\right\}\right] \\
&= E\left[\max_{\substack{0\leq \hat{m}<M-1 \\ \hat{m}\neq m}} \exp\left\{-\mu_n\left(\boldsymbol{C}_{\hat{m}}^{(n)}, \boldsymbol{C}_m^{(n)}\right)\Big/ s\right\}\right] \\
&\leq E\left[\sum_{\substack{0\leq \hat{m}<M-1 \\ \hat{m}\neq m}} \exp\left\{-\mu_n\left(\boldsymbol{C}_{\hat{m}}^{(n)}, \boldsymbol{C}_m^{(n)}\right)\Big/ s\right\}\right] \\
&= \sum_{\substack{0\leq \hat{m}<M-1 \\ \hat{m}\neq m}} E\left[\exp\left\{-\mu_n\left(\boldsymbol{C}_{\hat{m}}^{(n)}, \boldsymbol{C}_m^{(n)}\right)\Big/ s\right\}\right] \\
&= (M-1)E\left[e^{-\mu_n(\hat{X}^n, X^n)/s}\right].
\end{aligned}$$

Lemma A.2 then follows. □

Theorem A.1:

$$\limsup_{n\to\infty} \frac{d_{n,M}}{n} \geq \sup_{\boldsymbol{X}} \bar{G}_{\boldsymbol{X}}(R) \text{ and } \liminf_{n\to\infty} \frac{d_{n,M}}{n} \geq \sup_{\boldsymbol{X}} \underline{G}_{\boldsymbol{X}}(R)$$

where

$$\bar{G}_{\boldsymbol{X}}(R) \triangleq \sup_{s>0}[-sR - s\cdot \underline{\varphi}_{\boldsymbol{X}}(-1/s)]$$

$$\text{and } \underline{G}_{\boldsymbol{X}}(R) \triangleq \sup_{s>0}[-sR - s\cdot \bar{\varphi}_{\boldsymbol{X}}(-1/s)].$$

Proof: By letting $s = -1/t$ in Lemma A.2, the theorem then follows by noting that the rate only decreases by the amount $(\log 2)/n$ when employing a code $\mathfrak{C}_{n,M/2}$ from Lemma A.1, and both $\bar{G}_{\boldsymbol{X}}(R)$ and $\underline{G}_{\boldsymbol{X}}(R)$ are convex and hence continuous. □

The estimate in the proof of Lemma A.2 may make one suspicious that Lemma A.2 can be improved. Yet, it is shown in our previous work ([9], Theorem 4) that when the distance is additive and $\boldsymbol{X}$ is a product measure with identically distributed marginal, which is independent of n, the bound in Lemma A.2 is actually tight.

Appendix B
General Plotkin Upper Bound

Lemma B.1:

$$\limsup_{n\to\infty} \frac{d_{n,M}}{n} \leq \sup_{\boldsymbol{X}} \limsup_{n\to\infty} \frac{1}{n} E[\mu_n(\hat{X}^n, X^n)]$$

and

$$\liminf_{n\to\infty} \frac{d_{n,M}}{n} \leq \sup_{\boldsymbol{X}} \liminf_{n\to\infty} \frac{1}{n} E[\mu_n(\hat{X}^n, X^n)].$$

Proof: Suppose

$$\mathfrak{C}_{n,M}^* \triangleq \left\{\boldsymbol{c}_0^{(n)}, \cdots, \boldsymbol{c}_{M-1}^{(n)}\right\}$$

is one of the optimal code books of block length n. Then

$$\begin{aligned}
\frac{d_{n,M}}{n} &\leq \frac{1}{n} \frac{1}{M(M-1)} \sum_{m=0}^{M-1} \sum_{\hat{m}=0, \hat{m}\neq m}^{M-1} \mu_n\left(\boldsymbol{c}_{\hat{m}}^{(n)}, \boldsymbol{c}_m^{(n)}\right) \\
&\leq \frac{M}{M-1}\left(\frac{1}{n}\frac{1}{M^2} \sum_{m=0}^{M-1} \sum_{\hat{m}=0}^{M-1} \mu_n\left(\boldsymbol{c}_{\hat{m}}^{(n)}, \boldsymbol{c}_m^{(n)}\right)\right) \\
&\leq \frac{M}{M-1} \sup_{X^n} \frac{1}{n} E[\mu_n(\hat{X}^n, X^n)].
\end{aligned}$$

Hence

$$\begin{aligned}
\limsup_{n\to\infty} \frac{d_{n,M}}{n} &\leq \limsup_{n\to\infty} \sup_{X^n} \frac{1}{n} E[\mu_n(\hat{X}^n, X^n)] \\
&= \limsup_{n\to\infty} \frac{1}{n} E[\mu_n(\hat{X}_*^n, X_*^n)] \\
&\leq \sup_{\boldsymbol{X}} \limsup_{n\to\infty} \frac{1}{n} E[\mu_n(\hat{X}^n, X^n)],
\end{aligned}$$

where X_*^n represents one of the optimizers of

$$\sup_{X^n} (1/n) E[\mu_n(\hat{X}^n, X^n)].$$

Similar procedure can be used to prove that

$$\liminf_{n\to\infty} \frac{d_{n,M}}{n} \leq \sup_{\boldsymbol{X}} \liminf_{n\to\infty} \frac{1}{n} E[\mu_n(\hat{X}^n, X^n)].$$ □

APPENDIX C
PROOF OF THE GENERAL SPHERE-PACKING BOUND

Step 1) **Hypothesis testing.** For any code

$$\{\boldsymbol{c}_0, \boldsymbol{c}_1, \cdots, \boldsymbol{c}_{M-1}\}$$

given, we can form a maximum-likelihood partitions at the output as $\mathcal{U}_1, \mathcal{U}_2, \cdots, \mathcal{U}_M$, which is known to be optimal. Let $\mathcal{A}_{\hat{m},m}$ be the optimal acceptance region for alternative hypothesis under equal prior for

testing $H_0 : P_{Y^n \mid X^n}(\cdot \mid \boldsymbol{c}_m)$ against $H_1 : P_{Y^n \mid X^n}(\cdot \mid \boldsymbol{c}_{\hat{m}})$

and denote by $P_{e \mid m}$ the error probability given codeword m is transmitted. Then

$$P_{e \mid m} = P_{Y^n \mid X^n}(\mathcal{U}_m^c \mid \boldsymbol{c}_m) \geq P_{Y^n \mid X^n}(\mathcal{A}_{\hat{m},m}^c \mid \boldsymbol{c}_m)$$

where the superscript "c" represents the set complementary operation. Consequently,

$$\begin{aligned} P_{e \mid m} &\geq P_{Y^n \mid X^n}(\mathcal{A}_{\hat{m},m}^c \mid \boldsymbol{c}_m) \\ &\geq \exp\left\{-D\left(P_{Y_\lambda^n} \| P_{Y^n \mid X^n}(\cdot \mid \boldsymbol{c}_m)\right) + o(n)\right\} \end{aligned}$$

and

$$\begin{aligned} P_{e \mid \hat{m}} &\geq P_{Y^n \mid X^n}(\mathcal{A}_{\hat{m},m} \mid \boldsymbol{c}_{\hat{m}}) \\ &\geq \exp\left\{-D\left(P_{Y_\lambda^n} \| P_{Y^n \mid X^n}(\cdot \mid \boldsymbol{c}_{\hat{m}})\right) + o(n)\right\} \end{aligned}$$

where $P_{Y_\lambda^n}$ is the tilted distribution between $P_{Y^n \mid X^n}(\cdot \mid \boldsymbol{c}_m)$ and $P_{Y^n \mid X^n}(\cdot \mid \boldsymbol{c}_{\hat{m}})$ with

$$\begin{aligned} D\left(P_{Y_\lambda^n} \| P_{Y^n \mid X^n}(\cdot \mid \boldsymbol{c}_m)\right) &= D\left(P_{Y_\lambda^n} \| P_{Y^n \mid X^n}(\cdot \mid \boldsymbol{c}_{\hat{m}})\right) \\ &= \mu_n(\boldsymbol{c}_{\hat{m}}, \boldsymbol{c}_m) \end{aligned}$$

$D(\cdot \| \cdot)$ is the Kullback–Leibler divergence, and $\mu_n(\boldsymbol{c}_{\hat{m}}, \boldsymbol{c}_m)$ is the Bhattacharya distance between $\boldsymbol{c}_{\hat{m}}$ and $\boldsymbol{c}_m$. We thus have

$$P_{e \mid \hat{m}} + P_{e \mid m} \geq 2e^{-\mu_n(\boldsymbol{c}_{\hat{m}}, \boldsymbol{c}_m) + o(n)}.$$

Note that the above inequality holds for any $\hat{m}$ and m with $\hat{m} \neq m$.

Step 2) **Largest minimum distance.** By the definition of $d_{n,M}$, there exists an $(\hat{m}, m)$ pair for the above code such that

$$\mu_n(\boldsymbol{c}_{\hat{m}}, \boldsymbol{c}_m) \leq d_{n,M}$$

which implies

$$(\exists(\hat{m}, m)) \quad P_{e \mid \hat{m}} + P_{e \mid m} \geq 2e^{-d_{n,M} + o(n)}.$$

Step 3) **Probability of error.** Suppose we have found the optimal code with size $2M = e^{n(R + \log 2/n)}$, which minimizes the error probability. Index the codewords in ascending order of $P_{e \mid m}$, namely,

$$P_{e \mid 0} \leq P_{e \mid 1} \leq \cdots \leq P_{e \mid 2M-1}.$$

Form two new codebooks as

$$\mathfrak{C}_1 = \{\boldsymbol{c}_0, \boldsymbol{c}_1, \cdots, \boldsymbol{c}_{M-1}\} \text{ and } \mathfrak{C}_2 = \{\boldsymbol{c}_M, \cdots, \boldsymbol{c}_{2M-1}\}.$$

Then, from Steps 1 and 2, there exists at least one pair of codewords $(\boldsymbol{c}_{\hat{m}}, \boldsymbol{c}_m)$ in $\mathfrak{C}_1$ such that

$$P_{e \mid \hat{m}} + P_{e \mid m} \geq 2e^{-d_{n,M} + o(n)}.$$

Since for all $\boldsymbol{c}_i$ in $\mathfrak{C}_2$

$$P_{e \mid i} \geq \max\{P_{e \mid m}, P_{e \mid \hat{m}}\}$$

and hence

$$P_{e \mid i} + P_{e \mid j} \geq P_{e \mid \hat{m}} + P_{e \mid m} \geq 2e^{-d_{n,M} + o(n)}$$

for any $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ in $\mathfrak{C}_2$. Accordingly

$$\begin{aligned} P_e(n, 2M) &= \frac{1}{8M^2} \sum_{i=0}^{2M-1} \sum_{j=0}^{2M-1} (P_{e \mid i} + P_{e \mid j}) \\ &\geq \frac{1}{8M^2} \sum_{i=M}^{2M-1} \sum_{j=M, j \neq i}^{2M-1} (P_{e \mid i} + P_{e \mid j}) \\ &\geq \frac{1}{8M^2} \sum_{i=M}^{2M-1} \sum_{j=M, j \neq i}^{2M-1} (P_{e \mid \hat{m}} + P_{e \mid m}) \\ &\geq \frac{1}{8M^2} \sum_{i=M}^{2M-1} \sum_{j=M, j \neq i}^{2M-1} 2e^{-d_{n,M} + o(n)} \\ &= \frac{M-1}{4M} e^{-d_{n,M} + o(n)}. \end{aligned}$$

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers and Prof. S. Shamai for their valuable suggestions and comments that greatly helped to improve the paper. The warm encouragement from Prof. S. Verdú is also gratefully appreciated.

REFERENCES

[1] P. Billingsley, *Probability and Measure*. New York: Wiley, 1986.
[2] R. E. Blahut, *Principles and Practice of Information Theory*. Reading, MA: Addison-Wesley, 1988.
[3] V. Blinovsky, *Asymptotic Combinatorial Coding Theory*. Norwell, MA: Kluwer, 1997.
[4] J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation*. New York: Wiley, 1990.
[5] P.-N. Chen, "General formulas for the Neyman-Pearson type-II error exponent subject to fixed and exponential type-I error bounds," *IEEE Trans. Inform. Theory*, vol. 42, pp. 316–323, Jan. 1996.
[6] ——, "Generalization of Gártner-Ellis theorem," *IEEE Trans. Inform. Theory*, Feb. 1998, submitted for publication.
[7] P.-N. Chen and F. Alajaji, "Optimistic shannon coding theorems for arbitrary single-user systems," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2623–2629, Nov. 1999.
[8] ——, "On the optimistic capacity of arbitrary channels," in *Proc. 1998 IEEE Int. Symp. Information Theory*, MIT, Cambridge, MA, Aug. 16–21, 1998, p. 236.
[9] P.-N. Chen and T.-Y. Lee, "A new method to derive a lower bound of the largest minimum distance for block codes," in *Proc. Int. Symp, Communications*, National Chiao Tung Univ., Hsinchu, Taiwan, R.O.C., 1993.
[10] M. Deza, "Correction of arbitrary and worst noise," *Probl. Pered. Inform.*, vol. 6, pp. 24–30, Sept. 1964.
[11] M. Deza and F. Hoffman, "Some results related to generalized Varshamov-Gilbert bounds," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 517–518, July 1977.
[12] T. Ericson and V. A. Zinoviev, "An improvement of the Gilbert bound for constant weight codes," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 721–723, Sept. 1987.
[13] R. G. Gallager, *Information Theory and Reliable Communications*. New York: Wiley, 1968.

[14] G. van der Geer and J. H. van Lint, *Introduction to Coding Theory and Algebraic Geometry*. Basel, Switzerland: Birkhauser, 1988.
[15] R. L. Graham and N. J. A. Sloane, "Lower bounds for constant weight codes," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 37–43, Jan. 1980.
[16] J. Gu and T. Fuja, "A generalized Gilbert-Varshamov bound derived via analysis of a code-search algorithm," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1089–1093, May 1993.
[17] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.
[18] V. D. Kolesnik and V. Y. Krachkovsky, "Generating functions and lower bounds on rates for limited error-correcting codes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 778–788, May 1991.
[19] A. N. Kolmogorov and S. V. Fomin, *Introductory Real Analysis*. New York: Dover, 1970.
[20] J. H. van Lint, *Introduction to Coding Theory*, 2nd ed. New York: Springer-Verlag, 1992.
[21] S. N. Litsyn and M. A. Tsfasman, "A note on lower bounds," *IEEE Trans. Inform. Theory*, vol. IT-32, no. 5, pp. 705–706, Sept. 1986.
[22] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. New York: Elsevier, 1977.
[23] B. H. Marcus and R. M. Roth, "Improved Gilbert-Varshamov bound for constrained systems," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1213–1221, July 1992.
[24] J. K. Omura, "On general Gilbert bounds," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 661–666, Sept. 1973.
[25] H. V. Poor and S. Verdú, "A lower bound on the probability of error in multihypothesis testing," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1992–1994, Nov. 1995.
[26] H. L. Royden, *Real Analysis*, 3rd ed. New York: Macmillan, 1988.
[27] L. M. G. M. Tolhuizen, "The generalized Gibert-Varshamov bound is implied by Turán's theorem," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1605–1606, Sept. 1997.
[28] M. A. Tsfasman and S. G. Vladut, *Algebraic-Geometric Codes*. Amsterdam, The Netherlands: Kluwer, 1991.
[29] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, July 1994.
[30] S. G. Vladut, "An exhaustion bound for algebraic-geometric modular codes," *Probl. Inform. Transm.*, vol. 23, pp. 22–34, 1987.
[31] E. Zehavi and J. K. Wolf, "On runlength codes," *IEEE Trans. Inform. Theory*, vol. 34, pp. 45–54, Jan. 1988.
[32] V. A. Zinoviev and S. N. Litsyn, "Codes that exceed the Gilbert bound," *Probl. Pered. Inform.*, vol. 21, no. 1, pp. 105–108, 1985.