An Exact Spectrum Formula for the Maximum Size of Finite Length Block Codes

Ling-Hua Chang*, Carol Wang†, Po-Ning Chen*, Vincent Y. F. Tan‡, and Yunghsiang S. Han§

Abstract

An exact information spectrum-type formula for the maximum size of finite length block codes subject to a minimum pairwise distance constraint is presented. This formula can be applied to codes for a broad class of distance measures. As revealed by the formula, the largest code size is fully characterized by the information spectrum of the distance between two independent and identically distributed (i.i.d.) random codewords drawn from an optimal distribution. A new family of lower bounds to the maximal code size is thus established, and the well-known Gilbert-Varshamov (GV) lower bound is a special case of this family.

By duality, an explicit expression for the largest minimum distance of finite length block codes of a fixed code size is also obtained. Under an arbitrary uniformly bounded symmetric distance measure, the asymptotic largest code rate (in the block length $n$) attainable for a sequence of $(n, M, n\delta)$-codes is given exactly by the maximum large deviation rate function of the normalized distance between two i.i.d. random codewords. The exact information spectrum-type formula also yields bounds on the second-order terms in the asymptotic expansion of the optimum finite length rate for block codes with a fixed normalized minimum distance.

I. INTRODUCTION

The determination of the maximal size $M_n^*(d)$ of a finite length block code with pairwise distance $d$ and block length $n$ has been a long-standing problem in information and coding theory. In this paper, we establish an exact formula for this fundamental quantity for all block lengths $n \geq 1$. More specifically, we show that $M_n^*(d)$ is equal to the maximal reciprocal
of the cumulative distribution function $\Pr[\mu(\hat{X}^n, X^n) < d]$ over all i.i.d. pairs of random vectors $\hat{X}^n$ and $X^n$. This tight characterization gives a new way to construct good codes, and allows us to derive improved bounds on the optimal code size for a given distance.

Before describing our results, we introduce some of the previously known solutions and bounds, both in the asymptotic regime and for finite block length.

A. Related Works on Asymptotic Bounds

A common approach to understanding the maximum size of a block code is to allow the block length $n$ to grow, and to apply asymptotic analysis of the largest code rate $(1/n) \log M_n^*(d)$, subject to a normalized distance constraint $d/n \geq \delta$. Many bounds on this asymptotic quantity have been derived, for which interested readers are referred to [1]–[10] and the references therein and thereafter. The best-known upper and lower bounds are perhaps the linear programming upper bound [1] and the Gilbert-Varshamov (GV) lower bound [7]. Both of these bounds were stated originally for binary alphabets and the Hamming distance, but they can be generalized [11]. The GV lower bound remained the best lower bound for decades until the invention of Goppa codes [12], based on which a better lower bound, called the algebraic geometry (AG) bound, was established for certain range of $\delta$ when the code alphabet size is an even power of a prime and no less than 49 [13], [14]. Later, Zinoviev and Litsyn proved that a better lower bound than the GV lower bound is actually possible for any code alphabet size larger than 45 [15]. Nonetheless, a gap remains between these asymptotic bounds.

In 2000, Chen, Lee and Han [16] used ideas from information spectrum analysis [17], [18] to establish an exact formula for the asymptotic largest minimum distance (in the block length $n$) of deterministic block codes under generalized distance measures. Dually, the largest minimum distance considered in [16] is formulated as $\lim_{n \to \infty} (1/n) d_n^*(e^{nR})$ with

$$d_n^*(M) \triangleq \max_{C \subseteq X^n : |C| \geq M} \min_{\hat{x}^n, x^n \in C \text{ and } \hat{x}^n \neq x^n} \mu(\hat{x}^n, x^n),$$

where $|\cdot|$ denotes the size of a set and $\mu(\hat{x}^n, x^n)$ represents the distance between $\hat{x}^n$ and $x^n$. A major advantage of the exact formula in [16], as with most information spectrum results [18], is that its validity is not limited to the usual Hamming distance but can be applied to arbitrary distance measures. Furthermore, the derivation in [16] can be applied to arbitrary code alphabets, including continuous ones. This exact formula shows that the asymptotic behavior of the largest minimum distance of block codes is completely determined by the statistical properties of the normalized distance function $(1/n)\mu(\hat{X}^n, X^n)$ evaluated under
optimally chosen i.i.d. $\hat{X}^n$ and $X^n$. This result complements what has been obtained for channel capacity [17] and its optimistic version [19], where the information density is shown to be fundamental in the determination of the general formulae for channel capacities.

B. Related Works on Finite Length Bounds

Although the results stated above are for block lengths tending to infinity, in practice, encoding and decoding complexities often grow with the block length. Hence, it may not be practical to use long block lengths in order to approach asymptotic performance guarantees. Most of the asymptotic bounds mentioned at the beginning of this section, such as the linear programming upper bound [1] and the GV lower bound [7], also admit finite-length expressions, and there are many other well-known upper bounds, including the Singleton, Plotkin, and Elias bounds [20]. In general, however, these bounds are not tight.

In recent years, there has been renewed interest in understanding the performance of block codes with a fixed, finite block length (see, e.g., [21]–[27]). As a particular example, the authors in [21] established bounds on the maximal achievable transmission rate subject to a tolerable error probability and a finite block length. As with previous bounds, the results cited are not exact, and gaps generally remain between the upper and lower bounds.

C. Main Contributions

The main contributions of this work are as follows:

1) We provide an exact information spectrum formula for the maximal code size $M_*^n(d)$, given in Theorem 1. This simple formula completely characterizes $M_*^n(d)$ by an information spectrum expression. Our formula stands in stark contrast to existing finite length studies (e.g., [21]) in which the non-asymptotic bounds (e.g., dependence testing, meta-converse) obtained are, in general, not tight. An immediate implication of our result is a new family of lower bounds to $M_*^n(d)$, which includes the well-known finite length GV bound as a special case.

2) The Huffman-coding-like codeword-combining procedure used in the proof of Theorem 1 suggests an iterative algorithm for the construction of a code subject to a minimum pairwise distance constraint. Numerical experiments show that our procedure yields codes that outperform the finite length GV lower bound.

3) We extend our results to the determination of the largest asymptotic code rate and for the largest asymptotic minimum relative distance. These provide alternative, and arguably simpler, expressions compared to the asymptotic exact spectrum formulas in [16]. The
limiting rate and distance are shown to be characterized by the large deviation rate function for $(1/n)\mu(\hat{X}^n, X^n)$ [28], [29]. Novel bounds on the second-order terms in the asymptotic expansion of the optimum finite length rate for block codes with a fixed normalized minimum distance are obtained.

D. Paper Organization

The rest of the paper is organized as follows. The exact spectrum formula for $M^*_n(d)$ is stated and proven in Section II. A generalization of the finite length GV lower bound is proposed in Section III. The development of an iterative algorithm for the construction of a block code satisfying a given minimum distance is also presented in this section. The spectrum expression for the largest minimum distance of finite length block codes is addressed in Section IV. Extensions to the asymptotic regime, including bounds on the second-order terms, are studied in Section V. Open problems are discussed in Section VI.

II. LARGEST CODE SIZE ATTAINABLE UNDER A FIXED MINIMUM PAIRWISE DISTANCE

We introduce the notation in this paper. Let $\mathcal{X}$ be a finite set. An $(n, M)$-code $C \triangleq \{x^n_1, x^n_2, \ldots, x^n_M\}$ over alphabet $\mathcal{X}$ denotes a set of $M$ codewords, where each codeword $x^n_m \triangleq (x_{m,1}, x_{m,2}, \ldots, x_{m,n})$ belongs to $\mathcal{X}^n$ [30]. An $(n, M, d)$-code denotes an $(n, M)$-code with the minimum pairwise distance among codewords equal to $d$, i.e.,

$$d = \min_{x^n_i, x^n_j \in C \text{ and } x^n_i \neq x^n_j} \mu(x^n_i, x^n_j).$$

The maximal code size $M^*_n(d)$ subject to a fixed pairwise distance lower bound $d$ is

$$M^*_n(d) \triangleq \max \{M \in \mathbb{N} : \exists (n, M, d)\text{-code}\},$$

(1)

where $\mathbb{N}$ is the set of positive integers. The distance measure $\mu(\cdot, \cdot)$ is assumed to be symmetric, non-negative, and zero between a point and itself. In other words, for any two elements $\hat{x}^n$ and $x^n$ in $\mathcal{X}^n$, $\mu(\hat{x}^n, x^n) = \mu(x^n, \hat{x}^n)$, and $\mu(\hat{x}^n, x^n) \geq 0$ with equality holding if and only if $\hat{x}^n = x^n$. In order to simplify notation, we will sometimes use $\hat{x}$ and $X$, instead of the conventional notations $x^n$ and $X^n$ in the information spectrum literature [18], to respectively denote a single element and a random variable taking values in $\mathcal{X}^n$.

1The finiteness of $\mathcal{X}$ is not strictly necessary. This assumption implies that $M^*_n(d)$ in (1) is finite. See the Remark after the proof of Theorem 1.
A. An Exact Spectrum Formula for $M_n^*(d)$

**Definition 1:** Define the distance spectrum corresponding to a distance measure $\mu(\cdot, \cdot)$ and a distribution $P_X$ on $\mathcal{X}^n$ as

$$F_X(d) \triangleq \Pr[\mu(\hat{X}, X) < d],$$

where in this definition (and also throughout the paper), $\hat{X}$ and $X$ are used to denote two independent random variables over $\mathcal{X}^n$ having common distribution $P_X$.

If we consider attempting to construct a code of distance $d$ by drawing i.i.d. samples from $X$, the quantity $F_X(d)$ yields the probability that any two of the selected codewords violate the distance constraint. It turns out, perhaps surprisingly, that this quantity provides an exact information spectrum characterization of $M_n^*(d)$. This key result is summarized in the next theorem.

**Theorem 1:** For all $n \geq 1$ and $d > 0$,

$$M_n^*(d) = \sup_X \frac{1}{F_X(d)} = \sup_X \frac{1}{\Pr[\mu(\hat{X}, X) < d]}. \quad (2)$$

Before giving the proof of Theorem 1, which appears in Section II-B, we provide some remarks on the theorem itself, including some of its implications.

The determination of $M_n^*(d)$—that is, finding an “information” expression in terms of $d$ and $\mu(\cdot, \cdot)$ that is exactly equal to the operational quantity $M_n^*(d)$—has been a long-standing problem in information and coding theory. Up until now, only upper and lower bounds on this quantity were known. The above theorem beautifully shows that $M_n^*(d)$ can be fully characterized by a simple distance spectrum formula and confirms definitively that $\mu(\hat{X}^n, X^n)$ is a “sufficient statistic” for the determination of $M_n^*(d)$. It is also interesting to point out that the reciprocal of $F_X(d)$ is often not an integer for a general distribution $P_X$. However, after we have optimized $1/F_X(d)$ over all distributions, the optimal value is indeed an integer for all $n \geq 1$.

An immediate consequence of Theorem 1 is that a family of lower bounds to $M_n^*(d)$ can be obtained in a straightforward manner by evaluating $1/F_X(d)$ for any $X$. We summarize this observation in the following corollary.

**Corollary 1 (Distance spectrum (DS) lower bound):** For all $n \geq 1$ and $d > 0$, for any distribution $P_X$ over $\mathcal{X}^n$,

$$M_n^*(d) \geq L_X(d) \triangleq \frac{1}{F_X(d)}. \quad (3)$$

5
Note that if there does not exist a code of size two, in which the pairwise distance is no less than \(d\), then \(M_n^*(d) = 1\). This observation coincides with (2) and (3) as \(F_X(d) = 1\) for any \(X\) if \(d\) exceeds \(\max_{\hat{x}, x \in \mathcal{X}^n} \mu(\hat{x}, x)\).

The above corollary implies that a good lower bound can be obtained by employing a good distribution \(P_X\) over \(\mathcal{X}^n\). If the optimizer to (2) is used to evaluate \(1/F_X(d)\), the associated bound is indeed tight! A question that naturally follows is what we can possibly say about the set of optimizing distributions. Despite the challenge of determining optimizers to the optimization problem in Theorem 1, extensive numerical studies for binary codes of small block lengths \(n\) leads to the following observations.

First, the optimizer of the optimization problem in Theorem 1 may not be unique. Secondly, we have observed in numerical experiments that the set of optimizers seems to always include (at least) one distribution \(X^* = (X_1^* X_2^* \ldots X_n^*)\), of which all one-dimensional marginal distributions are uniform over \(\mathcal{X}\), i.e., \(P_{X_i^*}(x) = 1/|\mathcal{X}|\) for every \(x \in \mathcal{X}\) and every \(1 \leq i \leq n\). For example, an optimal distribution for codes of block length \(n = 3\) subject to a minimum pairwise distance constraint \(d = 3\) is

\[
P_{X^*}(x) = \begin{cases} 
\frac{1}{2}, & \text{for } x \in \{011, 100\}; \\
0, & \text{otherwise},
\end{cases}
\]

which yields \(M_n^*(3) = L_{X^*}(3) = 2\). It can be verified that

\[
P_{X_1^*}(0) = P_{X_1^*}(1) = \frac{1}{2} \quad \text{for } i = 1, 2, 3.
\]

An optimal distribution \(P_{X^*}\), under which \(L_{X^*}(d)\) achieves \(M_n^*(2) = 4\), is

\[
P_{X^*}(x) = \begin{cases} 
\frac{1}{4}, & \text{for } x \in \{001, 010, 100, 110\}; \\
0, & \text{otherwise},
\end{cases}
\]

which again yields

\[
P_{X_1^*}(0) = P_{X_1^*}(1) = \frac{1}{2} \quad \text{for } i = 1, 2, 3.
\]

Third, the two examples just introduced indicate that the optimizer may vary with \(d\) and is uniformly distributed over a subset of \(\mathcal{X}^n\). These observations motivate the main idea behind the proof of Theorem 1.

**B. Proof of Theorem 1**

The theorem will be proven in two steps:

1. (Converse) \(M_n^*(d) \leq 1/\left[\inf_X F_X(d)\right]\) and
2. (Direct) $M_n^*(d) \geq 1/\inf_X F_X(d)$, which together imply $M_n^*(d) = 1/\inf_X F_X(d)$.

1. $M_n^*(d) \leq 1/\inf_X F_X(d)$.

For an $(n, M, d)$-code $C \triangleq \{x_1, x_2, \ldots, x_M\}$ with $M = M_n^*(d)$, let $Z$ be the distribution which places probability mass $1/M$ on each codeword of $C$. Then, 

$$F_Z(d) = \Pr[\mu(Z, Z) < d] = \sum_{i=1}^M P_Z(x_i)P_Z(x_i) = \sum_{i=1}^M \left(\frac{1}{M}\right)\left(\frac{1}{M}\right) = \frac{1}{M} = \frac{1}{M_n^*(d)}.$$ 

This immediately implies that 

$$\inf_X F_X(d) \leq F_Z(d) = \frac{1}{M_n^*(d)}.$$ 

2. $M_n^*(d) \geq 1/\inf_X F_X(d)$.

Suppose that the support $S(Z)$ of a distribution $P_Z$ consists of $\ell$ distinct elements, i.e., 

$$S(Z) = \{x_1, x_2, \ldots, x_{\ell}\}$$

for some $x_1, x_2, \ldots, x_{\ell} \in X^n$. In other words, $P_Z(x_i) = p_i > 0$ for every $1 \leq i \leq \ell$ and $\sum_{i=1}^\ell P_Z(x_i) = \sum_{i=1}^\ell p_i = 1$. We claim that if $\ell > M_n^*(d)$, there exists another distribution $P_W$, whose support contains only $M_n^*(d)$ elements, satisfying 

$$F_W(d) \leq F_Z(d).$$

As a result of this claim, $\inf_X F_X(d)$ can be simplified to 

$$\inf_X F_X(d) = \inf_{X:|S(X)| \leq M_n^*(d)} F_X(d).$$

We thus continue from (5) to show that for any distribution $P_X$ with $|S(X)| = \ell \leq M_n^*(d)$,

$$F_X(d) \geq \sum_{i=1}^\ell P_X(x_i)P_X(x_i)$$

$$= \sum_{i=1}^\ell p_i^2 \geq \frac{1}{\ell}$$

$$\geq \frac{1}{M_n^*(d)},$$

where (6) follows from the Cauchy-Schwarz inequality, and (7) is due to the assumption that $\ell \leq M_n^*(d)$. We conclude

$$\inf_X F_X(d) = \inf_{X:|S(X)| \leq M_n^*(d)} F_X(d) \geq \frac{1}{M_n^*(d)}.$$ 

The Cauchy-Schwarz inequality states that

$$\left(\sum_{i=1}^\ell p_ir_i\right)^2 \leq \left(\sum_{i=1}^\ell p_i^2\right)\left(\sum_{i=1}^\ell r_i^2\right),$$

which recovers (6) by setting $r_i = 1$ for $1 \leq i \leq \ell$. 

7
It remains to substantiate the claim in (4), which can be proven through a Huffman-coding-
lke [31] codeword-combining procedure as follows. Among the support $S(Z) = \{x_1, x_2, \ldots, x_\ell\}$ of size $\ell > M^*(d)$, there must exist two distinct elements with distance smaller than $d$. Without loss of generality, we let this pair be $x_1$ and $x_2$, and denote for $1 \leq k \leq 2$,

$$q_k \triangleq \sum_{j=1}^{\ell} p_j 1\{\mu(x_k, x_j) < d\},$$

(8)

where $1\{\cdot\}$ is the indicator function. Assume, without loss of generality, that $q_1 \geq q_2$.

Construct $P_{W(1)}$ as

$$P_{W(1)}(x) = \tilde{p}_i = \begin{cases} 
0, & i = 1; \\
p_1 + p_2, & i = 2; \\
p_i, & 3 \leq i \leq \ell. 
\end{cases}$$

(9)

Note that the support of $P_{W(1)}$ consists of only $(\ell - 1)$ elements, i.e.,

$$S(W^{(1)}) = \{x_2, x_3, \ldots, x_\ell\}.$$ 

Now consider,

$$F_Z(d) - F_{W^{(1)}}(d) = \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} p_i p_j 1\{\mu(x_i, x_j) < d\} - \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \tilde{p}_i \tilde{p}_j 1\{\mu(x_i, x_j) < d\}$$

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} (p_i p_j - \tilde{p}_i \tilde{p}_j) 1\{\mu(x_i, x_j) < d\}$$

(10)

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} (p_i p_j + p_i \tilde{p}_j - \tilde{p}_i p_j - \tilde{p}_i \tilde{p}_j) 1\{\mu(x_i, x_j) < d\}$$

(11)

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} (p_i - \tilde{p}_i) (p_j + \tilde{p}_j) 1\{\mu(x_i, x_j) < d\}$$

(12)

$$= 2p_1(q_1 - q_2) \geq 0,$$

(13)

where (10) holds because

$$\sum_{i=1}^{\ell} \sum_{j=1}^{\ell} p_i \tilde{p}_j 1\{\mu(x_i, x_j) < d\} = \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \tilde{p}_i p_j 1\{\mu(x_i, x_j) < d\}.$$

Equalities (11) and (12) follow from the definition of $\tilde{p}_i$ in (9), and the inequality in (13) follows from the assumption that $q_1 \geq q_2$. 

8
If $|S(W^{(1)})| = \ell - 1$ is still larger than $M^*_n(d)$, we can similarly construct $P_{W^{(2)}}$ satisfying

$$|S(W^{(2)})| = \ell - 2 \quad \text{and} \quad F_{W^{(2)}}(d) \leq F_{W^{(1)}}(d).$$

By repeating such a construction $(\ell - M^*_n(d))$ times, we obtain the desired distribution $P_W = P_{W^{(\ell - M^*_n(d))}}$ with

$$|S(W)| = \ell - (\ell - M^*_n(d)) = M^*_n(d) \quad \text{and} \quad F_W(d) \leq F_Z(d).$$

This completes the proof of Theorem 1. ■

**Remarks.** We conclude this section by pointing out that the validity of the proof does not require the assumption that the alphabet $X$ is finite but only requires the finiteness of $M^*_n(d)$. As long as $M^*_n(d) < \infty$, the claim in (4) holds. In addition, the non-negativity assumption on the distance measure $\mu(\cdot, \cdot)$ is not necessary. We only require the property that

$$\mu(x_1, x_2) = \min_{\hat{x} \in \mathbb{Z}^n_X} \mu(\hat{x}, x) \triangleq \mu_0 \quad \text{if} \quad x_1 = x_2. \quad (14)$$

As a result, Theorem 1 can be extended to arbitrary code alphabets and arbitrary symmetric distance measure satisfying (14), provided that $d > \mu_0$ and $M^*_n(d)$ is finite.

### III. Implications of the Exact Spectrum Formula for $M^*_n(d)$

In this section, further explorations based on the theoretical result in the previous section are conducted. Specifically, Section III-A verifies that the finite length GV lower bound is a special case of the DS lower bound, Section III-B explores the achievability (to $M^*_n(d)$) of a subclass of the DS lower bound under uniform $X$, and Section III-C presents an algorithmic construction of block codes with code sizes exceeding the finite length GV bound.

#### A. Special Cases of the DS Lower Bound for $M^*_n(d)$

In the literature, the most well-known lower bound to $M^*_n(d)$ is perhaps the GV lower bound [7]. Under a finite code alphabet with $|X| = Q$ and the Hamming distance measure, the bound states that

$$M^*_n(d) \geq G_n(d) = \frac{Q^n}{\sum_{i=0}^{d-1} \binom{n}{i}}(Q - 1)^i. \quad (15)$$

When $Q$ is a prime power and the code is linear with respect to modulo-$Q$ addition and modulo-$Q$ multiplication, $G_n(d)$ has been improved [7] to

$$G^\text{lin}_n(d) = \frac{Q^n - 1}{Q^{\log_Q(\sum_{i=0}^{d-2} \binom{n-1}{i} (Q-1)^i)}}. \quad (16)$$
The next two examples show that both $G_n(d)$ and $G_{n|}(d)$ are special cases of the DS lower bound in Corollary 1.

**Example 1:** Consider a finite code alphabet $\mathcal{X}$ with $|\mathcal{X}| = Q$ and the Hamming distance measure. Let the components of $\underline{X} = (X_1 X_2 \ldots X_n)$ be i.i.d. random variables with a common distribution $P_X$. This choice yields

$$F_X(d) = \Pr \left[ \mu(\hat{X}, X) < d \right] = \Pr \left[ \sum_{i=1}^{n} \mu(\hat{X}_i, X_i) < d \right] = \sum_{i=0}^{d-1} \binom{n}{i} (1 - \beta)^i \beta^{n-i}, \quad (17)$$

where $\beta \triangleq \Pr[\mu(\hat{X}_i, X_i) = 0] = \Pr[\hat{X}_i = X_i]$. Similar to (6), we have

$$\beta = \sum_{x \in \mathcal{X}} P_X^2(x) \geq \frac{1}{Q},$$

according to the Cauchy-Schwarz inequality. Since (17) is monotonically increasing in $\beta$, the best DS bound $L_X(d)$ is to take the smallest $\beta$, i.e., $\beta = 1/Q$. Hence, $\underline{X}$ is uniformly distributed over $\mathcal{X}^n$ and

$$L_X(d) = \frac{1}{\sum_{i=0}^{d-1} \binom{n}{i} (1 - \frac{1}{Q})^i (\frac{1}{Q})^{n-i}} = \frac{Q^n}{\sum_{i=0}^{d-1} \binom{n}{i} (Q - 1)^i} \quad (18)$$

which is exactly $G_n(d)$ in (15). \qed

**Example 2:** Continue from Example 1 and consider a binary code alphabet $\mathcal{X} = \{0, 1\}$. Instead of taking $\underline{X}$ to be uniformly distributed over $\mathcal{X}^n$ as was done in the previous example, we now let it be uniformly distributed over a linear space spanned by $k$ linearly independent vectors $\{z_i\}_{i=1}^k$. Hence, $\underline{X}$ can be written as

$$\underline{X} \triangleq B_1 z_1 \oplus B_2 z_2 \oplus \cdots \oplus B_k z_k,$$

where $\{B_i\}_{i=1}^k$ are i.i.d. zero-one random variables with $\Pr[B_1 = 0] = \Pr[B_1 = 1] = 1/2$, and “$\oplus$” are component-wise XOR operation. Let $d$ be the minimum Hamming weight among all non-zero linear combinations of $\{z_i\}_{i=1}^k$. Then,

$$\mu(\hat{X}, X) = w_H(\hat{X} \oplus X) = w_H((\hat{B}_1 z_1 \oplus \cdots \oplus \hat{B}_k z_k) \oplus (B_1 z_1 \oplus \cdots \oplus B_k z_k))$$

$$= w_H((\hat{B}_1 \oplus B_1) z_1 \oplus \cdots \oplus (\hat{B}_k \oplus B_k) z_k)$$

$$= w_H(U_1 z_1 \oplus \cdots \oplus U_k z_k),$$

where $w_H(x)$ is the Hamming weight of vector $x$, and $U_i \triangleq \hat{B}_i \oplus B_i$. Accordingly,

$$\Pr[\mu(\hat{X}, X) < d] = \Pr \left[ w_H(U_1 z_1 \oplus \cdots \oplus U_k z_k) < d \right],$$
from which we can see that the distance spectrum is reduced to the probability that the
Hamming weights of linear codewords are less than $d$. Accordingly,
\[
\Pr[\mu(\hat{X}, X) < d] = \Pr[w_H(U_1 \tilde{z}_1 \oplus \cdots \oplus U_k \tilde{z}_k) < d]
\]
\[
= \Pr[U_1 = U_2 = \cdots = U_k = 0]
\]
\[
= (\Pr[U_1 = 0])^k
\]
\[
= 2^{-k}.
\]
As anticipated, $L_X(d) = 2^k$.

As a specific example, when taking \(\{\tilde{z}_1, \tilde{z}_2\} = \{0011, 1100\}\) for $n = 4$, we have $d = 2$ and
\[
\Pr[\mu(\hat{X}, X) < 2] = \Pr[w_H(U_1 \tilde{z}_1 \oplus U_2 \tilde{z}_2) < 2]
\]
\[
= \Pr[U_1 = U_2 = 0]
\]
\[
= (\Pr[U_1 = 0])^2 = \frac{1}{4}.
\]
This yields $L_X(2) = G_{4\text{in}}(2) = 4$, which is an improvement over $G_4(2) = 16/5 = 3.2$ from (18).

The two examples indicate that the finite length GV lower bounds could be regarded as
special cases of the DS lower bound by adopting different uniform $X$. While (15) is based
on uniform $X$ over the entire $\mathcal{X}^n$, the improved finite length GV lower bound in (16) can
be obtained by adopting uniform $X$ over a linear subspace of $\mathcal{X}^n$. The former observation
in (15) has actually been stated by Kolesnik and Krachkovsky in [5, pp. 1446]. As we will
see in the next subsection, this may help determining distributions $P_X$ which yield tighter
bounds on $M_n^*(d)$.

B. Uniform Distribution (UD) Lower Bounds

Theorem 1 implies the sufficiency of employing uniform $X$ (over an appropriate subset of
$\mathcal{X}^n$) to achieve $M_n^*(d)$. The examples in previous subsection suggest that adopting uniform
$X$ may result in a good lower bound. Thus, $L_X(d)$ based on uniform $X$ forms an important
subclass of the DS lower bounds, which is referred to as the uniform distribution (UD)
lower bound for convenience. The examples below for specific distance measures confirm
the significance of the UD lower bound.
Example 3: For a finite code alphabet $\mathcal{X}$ with $|\mathcal{X}| = Q$ and a so-called “probability-of-error” distance measure [16], defined as

$$
\mu(\hat{x}, x) = \begin{cases} 
0, & \hat{x} = x; \\
q, & \hat{x} \neq x.
\end{cases}
$$

we obtain

$$
M^*_n(d) = \begin{cases} 
Q^n, & 0 < d \leq n; \\
1, & d > n,
\end{cases}
$$

which can be validated by

$$
M^*_n(d) \geq \frac{1}{F_X(d)} = \begin{cases} 
Q^n, & 0 < d \leq n; \\
1, & d > n,
\end{cases}
$$

where $X$ is the uniform distribution on $\mathcal{X}^n$.

Example 4: Let the distance measure be given by

$$
\mu(\hat{x}, x) = |\kappa_n(\hat{x}) - \kappa_n(x)|,
$$

where $\hat{x}$ and $x$ are in $\{0, 1\}^n$, and $\kappa_n(x) \triangleq x_n2^{n-1} + x_{n-1}2^{n-2} + \ldots + x_22^1 + x_1$ is the binary representation of $x = (x_1 \ x_2 \ \ldots \ x_n)$. By its definition, $\mu(\hat{x}, x)$ is the absolute difference between two decimal numbers $\kappa_n(\hat{x})$ and $\kappa_n(x)$, and is a separable distance measure [16, Def. 1].

Since $\kappa_n(x)$ is an integer in $\{0, 1, 2, \ldots, 2^n - 1\}$, it can be easily seen that for $d > 0$,

$$
M^*_n(d) = \left\lceil \frac{2^n}{\lceil d \rceil} \right\rceil,
$$

where $\lceil \cdot \rceil$ is the ceiling function. One of those uniform $X$’s that equate $L_X(d)$ with $M^*_n(d)$ has support $\{0, [d], 2[d], \ldots, (M^*_n(d) - 1)[d]\}$, and there are exactly $\lceil d \rceil$ optimizers for $M^*_n(d) = \sup_X(1/F_X(d))$.

We then recall that Example 1 has illustrated that $G_n(d)$ can be regarded as a special case of the DS lower bound under uniform $X$ over $\mathcal{X}^n$. Along this perspective, the “GV lower
bound" for this separable distance measure is given by

\[ G_n(d) = \frac{1}{\Pr\{|\kappa_n(\hat{X}) - \kappa_n(X)| < d\}} = \frac{1}{\Pr\{|\kappa_n(\hat{X}) - \kappa_n(X)| < [d]\}} \]

\[ = \begin{cases} 
\frac{2^n}{(2n - 1)(2^n - 2^n + 1)}, & 0 < [d] \leq 2^{n-1} \\
\frac{2^n + ([d] - 2^n)(2^n - [d] + 1)}{2^n}, & 2^{n-1} < [d] \leq 2^n - 1; \\
1, & [d] > 2^n - 1.
\end{cases} \]

As a consequence, \( G_n(d) \) is strictly less than \( M^*_n(d) \) except when \([d] = 1\) and \([d] \geq 2^n\). This result confirms that the finite length GV lower bound is not tight in general.

In this special example, an upper bound \( U_n(d) \) for \( M^*_n(d) \) can also be provided based on Theorem 1. If there exists \( U_n(d) \) such that \( U_n(d) \geq 1/F_X(d) \) for all \( X \)'s, then

\[ U_n(d) \geq \sup_X \frac{1}{F_X(d)} = M^*_n(d). \]

Now setting \( j = j(n, d) \triangleq 2^n/[d] \), we derive

\[ F_X(d) = \Pr\left\{ \left| \frac{\kappa_n(\hat{X})}{2^n} - \frac{\kappa_n(X)}{2^n} \right| < \frac{[d]}{2^n} \right\} \]

\[ = \Pr\left\{ \left| \frac{\kappa_n(\hat{X})}{2^n} - \frac{\kappa_n(X)}{2^n} \right| < \frac{1}{j} \right\} \]

\[ \geq \sum_{i=0}^{[j] - 1} \Pr\left\{ \frac{i}{[j]} \leq \frac{\kappa_n(\hat{X})}{2^n} < \frac{i + 1}{[j]} \right\} \]

\[ = \sum_{i=0}^{[j] - 1} \Pr\left\{ \frac{i}{[j]} \leq \frac{\kappa_n(\hat{X})}{2^n} < \frac{i + 1}{[j]} \right\} \Pr\left\{ \frac{i}{[j]} \leq \frac{\kappa_n(X)}{2^n} < \frac{i + 1}{[j]} \right\} \]

\[ = \sum_{i=0}^{[j] - 1} \left( \Pr\left\{ \frac{i}{[j]} \leq \frac{\kappa_n(X)}{2^n} < \frac{i + 1}{[j]} \right\} \right)^2 \]

\[ \geq \frac{1}{[j]^2}, \]

where (19) and (20) hold because \( \hat{X} \) and \( X \) are i.i.d., and (21) follows from the Cauchy-Schwarz inequality (similar to (6)). This immediately gives

\[ U_n(d) = [j] = \left\lfloor \frac{2^n}{[d]} \right\rfloor, \]

which is precisely \( M^*_n(d) \).

\[ \square \]
C. Algorithmic Construction of Block Codes Subject to a Given Minimum Distance Criterion

The direct proof (i.e., step 2) of Theorem 1 suggests an iterative procedure to remove elements from an initial support of size larger than \( M^* n(d) \). Between any two elements selected from the current support, the one with a larger \( q \)-value (cf. (8)) should be removed, as implied from (13). Through such an iterative procedure, it can be anticipated that

\[
F_{\mathcal{Z}}(d) \geq F_{W(1)}(d) \geq F_{W(2)}(d) \geq F_{W(3)}(d) \geq \ldots
\]

and the DS lower bound is iteratively improved until either \( M^* n(d) \) is reached or no further improvement can be obtained. An experiment on this algorithmic construction is thus conducted via the procedures outlined below. We observe that a number of block codes of sizes exceeding the GV lower bound are obtained.

Algorithm 1 (Algorithmic lower bound \( A_n(d) \) to \( M^* n(d) \)):

Step 1. Set \( i = 1 \). Let \( P_{\mathcal{X}^{(i)}} \) be the uniform distribution over (finite) support \( \mathcal{X}^n \). Initialize for every \( \underline{z} \) and \( \underline{x} \) in \( \mathcal{X}^n \),

\[
\eta(\underline{z}, \underline{x}) = 1\{\mu(\underline{z}, \underline{x}) < d\},
\]

which indicates whether or not one of \( \underline{z} \) and \( \underline{x} \) needs to be removed. Record the number of such pairs as

\[
\Gamma^{(i)} = \sum_{\underline{z}, \underline{x} \in S(\mathcal{X}^{(i)}) : \underline{z} \neq \underline{x}} \eta(\underline{z}, \underline{x}).
\]

Step 2. If \( \Gamma^{(i)} = 0 \), then output

\[
A_n(d) = |S(\mathcal{X}^{(i)})|
\]

and stop the algorithm.

Step 3. Select arbitrarily two distinct elements \( \underline{z}^{(i)} \) and \( \underline{x}^{(i)} \) from \( S(\mathcal{X}^{(i)}) \), satisfying

\[
\eta(\underline{z}^{(i)}, \underline{x}^{(i)}) = 1 \quad \text{and} \quad q_{\underline{z}^{(i)}} \geq q_{\underline{x}^{(i)}},
\]

where

\[
q_{\underline{z}^{(i)}} \triangleq \sum_{\underline{x} \in \mathcal{X}^n} P_{\mathcal{X}^{(i)}}(\underline{x}) \cdot \eta(\underline{z}^{(i)}, \underline{x})
\]

and

\[
q_{\underline{x}^{(i)}} \triangleq \sum_{\underline{z} \in \mathcal{X}^n} P_{\mathcal{X}^{(i)}}(\underline{z}) \cdot \eta(\underline{z}^{(i)}, \underline{x}).
\]
Step 4. Construct $P_{X^{(i+1)}}$ as

$$P_{X^{(i+1)}}(x) = \begin{cases} 0, & x = z^{(i)}; \\ P_{X^{(i)}}(z^{(i)}) + P_{X^{(i)}}(x^{(i)}), & x = x^{(i)}; \\ P_{X^{(i)}}(x), & \text{otherwise} \end{cases}$$

and update

$$\Gamma^{(i+1)} = \sum_{z, x \in S(X^{(i+1)}): z \neq x} \eta(z, x).$$

Step 5. Set $i = i + 1$ and go to Step 2.

For clarity, an example is provided to demonstrate the algorithmic establishment of $A_n(d)$.

**Example 5:** Let $\mu(\cdot, \cdot)$ be the Hamming distance measure, and initialize $P_{X^{(1)}}$ as the uniform distribution with support $\{0, 1\}^3$. Then, a lower bound $A_3(2)$ to $M_3^*(2)$ can be established through the proposed algorithmic procedure as follows.

1) First, we choose 000 and 001 in Step 3 as $\eta(000, 001) = 1$. Then,

$$q_{000} = P_{X^{(1)}}(000) + P_{X^{(1)}}(001) + P_{X^{(1)}}(010) + P_{X^{(1)}}(100) = \frac{1}{2},$$

and

$$q_{001} = P_{X^{(1)}}(000) + P_{X^{(1)}}(001) + P_{X^{(1)}}(011) + P_{X^{(1)}}(101) = \frac{1}{2}.$$  

Because $q_{000}$ and $q_{001}$ are equal, we just choose to keep 001 with its probability being updated as $1/8 + 1/8 = 1/4$, and remove 000 by letting the corresponding probability mass be zero.

2) Next, we pick 010 and 011 as $\eta(010, 011) = 1$. Then,

$$q_{010} = P_{X^{(2)}}(010) + P_{X^{(2)}}(011) + P_{X^{(2)}}(110) = \frac{3}{8},$$

and

$$q_{011} = P_{X^{(2)}}(010) + P_{X^{(2)}}(011) + P_{X^{(2)}}(001) + P_{X^{(2)}}(111) = \frac{5}{8}.$$  

Thus, we should keep 010 with $P_{X^{(2)}}(010)$ becoming $1/4$, and remove 011.

3) We next pick 100 and 101 since $\eta(100, 101) = 1$. Then,

$$q_{100} = P_{X^{(3)}}(100) + P_{X^{(3)}}(101) + P_{X^{(3)}}(110) = \frac{3}{8},$$

and

$$q_{101} = P_{X^{(3)}}(100) + P_{X^{(3)}}(101) + P_{X^{(3)}}(111) + P_{X^{(3)}}(001) = \frac{5}{8}.$$
Thus, we keep 100, update $P_{X(4)}(100)$ as 1/4, and remove 101.

4) Last, we pick 110 and 111, and compute

$$q_{110} = P_{X(4)}(110) + P_{X(4)}(111) + P_{X(4)}(100) + P_{X(4)}(010) = \frac{3}{4}$$

and

$$q_{111} = P_{X(4)}(110) + P_{X(4)}(111) = \frac{1}{4}.$$

So, we keep 111, update $P_{X(5)}(111)$ as 1/4, and remove 110.

5) As $\Gamma^{(5)} = 0$, the algorithm stops with $A_3(2) = 4$, which is exactly $M^*_3(2)$! \hfill \Box

Numerical results show that the algorithmic bound $A_n(d)$ improves considerably the finite length GV bound in (15) for most $d$ between 3 and $n/2$, particularly at larger $n$ such as $n = 13$, as shown in Fig. 1. Such an improvement was obtained merely based on picking up the next two elements lexicographically from the current support in Step 3 (cf. Example 5).

A key to the success of the above algorithm, as indicated by Example 5, is that the one to be removed is not selected by a local distance property, but by a global $q$-value that considers the relation of the selected vector with all other elements in the current support. Investigating how to select a good pair of elements to further improve $A_n(d)$ could be a future work of practical interest.

IV. LARGEST MINIMUM DISTANCE OF FINITE LENGTH BLOCK CODES

Dual to the definition of $M^*_n(d)$ in (1), we define the largest minimum distance of finite length block codes as

$$d^*_n(M) \triangleq \sup \{d > 0 : \exists (n, M, d)-code\}.$$

By leveraging duality and Theorem 1, this quantity is characterized in the following theorem.

**Theorem 2:** For all $n \geq 1$ and $M \geq 2$,

$$d^*_n(M) = \max \left\{ a \in \mathbb{R}^+ : M^*_n(a) \geq M \right\}$$

$$= \max \left\{ a \in \mathbb{R}^+ : \inf \sum_{X} F_X(a) \leq \frac{1}{M} \right\}, \tag{22}$$

where $\mathbb{R}^+$ is the set of positive real numbers.

Similar to the remarks made at the end of Section II, if $\mu(\cdot, \cdot)$ is not a non-negative distance measure, the range of $a$ in (22) should be changed to $(\mu_0, \infty)$ (where $\mu_0$ is defined in (14)).

An immediate consequence of Theorem 2 is that any distribution $P_X$ provides a lower bound to $d^*_n(M)$. We summarize this observation in the next corollary.
Fig. 1: Logarithmic function values $k = \log_2(\cdot)$ of the algorithmic bound $A_n(d)$ and the finite length GV lower bound $G_n(d)$ in (15).

**Corollary 2:** For $2 \leq M \leq |\mathcal{X}|^n$,

$$d_n^*(M) \geq \sup\left\{ a \in \mathbb{R}^+ : F_X(a) \leq \frac{1}{M} \right\}$$

for any $n$-letter distribution $P_X$.

Based on this corollary, lower bounds to $d_n^*(M)$ for binary block codes have been computed for $n = 6, 7, 8, 11$ and $M = 2^k$ for any integer $1 \leq k < n$; this was presented in the conference version of the present work [32, Fig. 1]. In contrast to the algorithmic construction in Section III-C, the distributions employed therein are uniformly distributed.
over a $k$-dimensional linear subspace with $k$ properly chosen length-$n$ vectors as their bases, which satisfy the condition that all pairwise distances among these $k$ basis vectors are no less than $d$ (as was similarly done in Example 2). The results confirm, again, the lower bound obtained via Corollary 2 can considerably outperform the finite length GV lower bound.

V. EXTENSIONS TO THE ASYMPTOTIC REGIME

A. The Maximal Code Rate and the Largest Minimum Relative Distance

In this section, we extend the results in Theorems 1 and 2 to a form that is amenable to asymptotic analyses. Distance spectrum-based formulae for the relative distance $\delta = d/n$ and the code rate $R = \log(M)/n$ are provided in the next theorem. We adopt the natural logarithm in the following. In order to reflect clearly its dependence on block length $n$, $X^n$ rather than $X$ will be used to denote an $n$-tuple of random variables taking values in $X^n$.

Theorem 3: The largest code rate attainable for an $(n, M, n\delta)$-code is equal to

$$R^*_n(\delta) \triangleq \frac{1}{n} \log M^*_n(n\delta) = \sup_{X^n} \left( -\frac{1}{n} \log F_{X^n}(n\delta) \right)$$

$$= \sup_{X^n} \left( -\frac{1}{n} \log \Pr \left[ \frac{1}{n} \mu(\hat{X}^n, X^n) < \delta \right] \right). \tag{23}$$

By duality, the largest minimum relative distance $\delta^*_n(R)$ for codes with rate $R = (1/n) \log(M)$ is given by

$$\delta^*_n(R) = \max \left\{ a \in \mathbb{R}^+ : \sup_{X^n} \left( -\frac{1}{n} \log \Pr \left[ \frac{1}{n} \mu(\hat{X}^n, X^n) < a \right] \right) \geq R \right\}. \tag{24}$$

Proof: Equation (23) follows from Theorem 1. By duality to (23), we have $\delta^*_n(R) = \max \{ a \in \mathbb{R}^+ : R^*_n(a) \geq R \}$. \qed

The formula in (24) provides a different form compared to the general asymptotic formula of $\delta^*_n(R)$ from [16], which we recapitulate here to facilitate subsequent comparison:

$$\limsup_{n \to \infty} \delta^*_n(R) \triangleq \sup X \inf a \in \mathbb{R} \limsup_{n \to \infty} \left( \Pr \left[ \frac{1}{n} \mu(\hat{X}^n, X^n) > a \right] \right)^{\exp(nR)} = 0 \} \tag{25}$$

and

$$\liminf_{n \to \infty} \delta^*_n(R) \triangleq \sup X \inf a \in \mathbb{R} \liminf_{n \to \infty} \left( \Pr \left[ \frac{1}{n} \mu(\hat{X}^n, X^n) > a \right] \right)^{\exp(nR)} = 0 \}, \tag{26}$$

where $X \triangleq \{X^n\}_{n=1}^\infty$ and $\mathbb{R}$ denotes the set of real numbers.

When compared to (25) and (26), the new way to characterize the largest minimum relative distance in (24) exhibits the following advantages.

\[\text{What was proven in [16] is that (25) and (26) hold except possibly at countably many points of discontinuities in } R.\]
1) The cumbersome exponent of $e^{nR}$ in (25) and (26) no longer exists.

2) The task of taking the supremum over all distributions is now done before the functional optimization of parameter $a$ over the real line. This reduces the set of all distributions to be optimized over from being of infinite dimension for $X$ to being of finite dimension for $X^n$, and hence may help in evaluating $R_n^*(\delta)$ via the numerical characterization of the optimal distribution $P^*_{X^n}$, particularly when $n$ is small.

3) With the exact formula for every $n$, we exclude the necessity of sandwiching the quantities of (25) and (26), and remove the mathematical peculiarity, such as equality holding for all rates $R$, except for (at most) countably many points.

B. Bounds for the largest code rate attainable for an $(n, M, n\delta)$-code

The formula of $R_n^*(\delta)$ in Theorem 3 provides a new quantitative characterization of the largest code rate attainable for an $(n, M, n\delta)$-code. As a result, $R_n^*(\delta)$ can be lower-bounded by the large deviation rate function of $(1/n)\mu(X^n, X^n)$.

**Theorem 4:** The largest code rate $R_n^*(\delta)$ and the largest minimum relative distance $\delta_n^*(R)$, are respectively lower bounded as

$$R_n^*(\delta) \geq \sup_{X^n} J_{X^n}(\delta)$$

and

$$\delta_n^*(R) \geq \max \left\{ a \in \mathbb{R}^+ : \sup_{X^n} J_{X^n}(a) \geq R \right\},$$

where

$$J_{X^n}(\delta) \triangleq \inf_{0 \leq a \leq \delta} I_{X^n}(a),$$

$$I_{X^n}(a) \triangleq \sup_{\theta \in \mathbb{R}} \{ a\theta - \varphi_{X^n}(\theta) \},$$

and

$$\varphi_{X^n}(\theta) \triangleq \frac{1}{n} \log \mathbb{E} \left[ e^{\theta \mu(\hat{X}^n, X^n)} \right].$$

**Proof:** First, we note from basic properties of the large deviation rate function [29] that

$$J_{X^n}(\delta) = \begin{cases} I_{X^n}(\delta), & 0 < \delta < \frac{1}{n} \mathbb{E}[\mu(\hat{X}^n, X^n)]; \\ 0, & \text{otherwise.} \end{cases}$$

The large deviation rate function $I_{X^n}(a)$ is convex, and admits its global minimum $\min_{a \in \mathbb{R}} I_{X^n}(a) = 0$ at $a = (1/n)\mathbb{E}[\mu(\hat{X}^n, X^n)]$. 

19
Therefore, it suffices to prove (27) under the condition that
\[ n\delta < \mathbb{E}[\mu(\hat{X}^n, X^n)]. \] (29)

Let \( Y \triangleq n\delta - \mu(\hat{X}^n, X^n) \) and note that \( \mathbb{E}[Y] < 0 \). It can then be derived from Markov’s inequality that for \( \theta > 0 \),
\[
\Pr\left[ \frac{1}{n} \mu(\hat{X}^n, X^n) < \delta \right] = \Pr[Y > 0] \leq \mathbb{E}[e^{\theta Y}] \triangleq M_Y(\theta). \] (30)

By using the fact that \[ \frac{\partial}{\partial \theta} M_Y(\theta) \bigg|_{\theta=0} = \mathbb{E}[Y] < 0 \]
and the convexity of \( M_Y(\theta) \) over \( \theta \in \mathbb{R} \), we obtain
\[
\Pr[Y > 0] \leq \inf_{\theta > 0} M_Y(\theta) = \inf_{\theta \in \mathbb{R}} M_Y(\theta) = \inf_{\theta \in \mathbb{R}} M_Y(-\theta)
= \exp \left\{ -n \sup_{\theta \in \mathbb{R}} \left( \delta \theta - \frac{1}{n} \log \mathbb{E} \left[ e^{\theta \mu(\hat{X}^n, X^n)} \right] \right) \right\}
= \exp \left\{ -n \cdot I_X(\delta) \right\}.
\]

This completes the proof of (27). The bound in (28) can be easily obtained by duality.

When the distance measure \( \mu(\cdot, \cdot) \) is additive and the components of \( X^n \) are i.i.d. with generic distribution \( P_X \), \( I_X(a) \) exhibits a single-letter expression for all block lengths \( n \) as:
\[
I_X(a) = I_X(a) = \sup_{\theta \in \mathbb{R}} \{ a\theta - \varphi_X(\theta) \},
\]
where
\[
\varphi_X(\theta) = \log \mathbb{E} \left[ e^{\theta \mu(X, X)} \right].
\]

This reduces (27) to
\[
R^*_n(\delta) \geq \sup_X J_X(a).
\]

Further assuming that \( X = \{ \alpha_1, \alpha_2, \ldots, \alpha_Q \} \) is finite and \( \mu(\cdot, \cdot) \) is the Hamming distance measure, we obtain
\[
\varphi_X(\theta) = \log \left( \sum_{i=1}^{Q} \sum_{j=1}^{Q} P_X(\alpha_i) P_X(\alpha_j) e^{\theta \mu(\alpha_i, \alpha_j)} \right)
= \log \left( 1 - b_X + b_X e^\theta \right),
\]
and
\[
I_X(a) = \sup_{\theta \in \mathbb{R}} \{ a\theta - \varphi_X(\theta) \} = D(a || b_X),
\]

20
where $b_x \triangleq 1 - \sum_{i=1}^{Q} P_{X}^2 (\alpha_i)$ and

$$D(a \| b_x) \triangleq a \log \left( \frac{a}{b_x} \right) + (1 - a) \log \left( \frac{1 - a}{1 - b_x} \right)$$

is the binary Kullback-Leibler divergence [33]. This leads to

$$J_X(\delta) = I_X(\delta) = \begin{cases} D(\delta \| b_x), & 0 < \delta < b_x; \\ 0, & \text{otherwise}. \end{cases}$$

The same argument used to verify (6) implies that $0 \leq b_x \leq (Q - 1)/Q$. Consequently, for $\delta > 0$,

$$R^*_n(\delta) \geq \sup_X J_X(\delta)$$

$$= \sup_{0 \leq b_x \leq (Q - 1)/Q} J_X(\delta)$$

$$= \begin{cases} D \left( \delta \left\| \frac{Q - 1}{Q} \right\| \right), & 0 < \delta < \frac{Q - 1}{Q}; \\ 0, & \delta \geq \frac{Q - 1}{Q}. \end{cases} \quad (31)$$

and for $0 < R < \log Q$,

$$d^*_n(R) \geq \max \left\{ a \in \mathbb{R} : \sup_X J_X(a) \geq R \right\}$$

$$= \max \left\{ a \in \mathbb{R} : D \left( a \left\| \frac{Q - 1}{Q} \right\| \right) \geq R \right\}. \quad (32)$$

As a result, when we take $X^n = (X_1, X_2, \ldots, X_n)$ to have i.i.d. components, (27) is reduced to (31) under the Hamming distance measure, which yields exactly the asymptotic GV lower bound [7], [34]. This suggests that the asymptotic GV lower bound may be improved with more general $X^n$, particularly $X^n$ with dependent components.

In fact, by using a more sophisticated large deviation technique, we can obtain slight but non-trivial improvements to (31) and (32). As above, we assume that $\mu(\cdot, \cdot)$ is the Hamming distance measure but for the sake of simplicity, we let $Q = 2$; thus, $X = \{\alpha_1, \alpha_2\}$ is a binary code alphabet. Furthermore, we let $\delta_{GV}(R)$ (for $0 \leq R \leq \log 2$) denote the GV lower bound in (32). This function has the property that $\delta_{GV}(0) = 1/2$ and $\delta_{GV}(R)$ decreases monotonically to 0 as $R \uparrow \log 2$. When $R > \log 2$, we define $\delta_{GV}(R) = 0$. Furthermore, $\delta_{GV}(R)$ is continuously differentiable on $(0, \log 2)$.

We then have the following result, which strengthens (31) and (32) by the addition of a logarithmic term.

**Theorem 5:** For any $0 < \delta < 1/2$, one has

$$R^*_n(\delta) \geq D \left( \delta \left\| \frac{1}{2} \right\| \right) + \log n + \Theta \left( \frac{1}{n} \right) \quad (33)$$
as $n \to \infty$. In a similar manner, for any $0 < R < \log 2$, one has
\[
\delta_n^*(R) \geq \delta_{GV}(R) + \left| \frac{d \delta_{GV}(R)}{dR} \log n \right| 2n + \Theta\left( \frac{1}{n} \right)
\]
as $n \to \infty$.

Before we prove this result, we remark that Jiang and Vardy [35, Thm. 1] proved an asymptotic improvement to the GV lower bound. They showed, by using a graph-theoretic framework, that the achievable second-order term in (33) is at least $(\log n)/n$. This is slightly stronger than our result in (33) because we showed that the achievable second-order term in (33) is at least $(\log n)/(2n)$ but our methods, based primarily on information spectrum analysis [18], are completely different from Jiang and Vardy’s [35] and provide some additional insight into the suboptimality of choosing $X^n$ with i.i.d. components since our evaluation of the relevant distance spectrum is asymptotically tight. See (35) to follow.

**Proof:** Per the proof of Theorem 4, to lower bound $R_n^*(\delta)$, it suffices to provide an upper bound on $\Pr[\mu(\hat{X}^n, X^n) < n\delta]$. However, instead of using Markov’s inequality in (30), we will evaluate this using exact asymptotics. Choose $X^n$ (and also $\hat{X}^n$) to also be i.i.d. with generic distribution Bernoulli$(1/2)$. Then it is clear that $K_i \overset{\text{d}}{=} \mu(X_i, X_i)$ is also a sequence of i.i.d. Bernoulli$(1/2)$ random variables. Since the span of $K_i$ is 1, by the Bahadur-Rao theorem [36] (see also [28, Thm. 3.7.4]) applied to lattice random variables, we know that
\[
\Pr\left[\frac{1}{n} \mu(\hat{X}^n, X^n) < \delta \right] = \Pr\left[\frac{1}{n} \sum_{i=1}^{n} K_i < \delta \right] \sim \left( \frac{1 - \delta}{1 - 2\delta} \right) \left( \frac{e^{-nD(\delta \parallel \frac{1}{2})}}{\sqrt{2\pi \delta(1 - \delta)n}} \right)
\]
where $f_n \sim g_n$ means that $\lim_{n \to \infty} f_n/g_n = 1$. See [37, Eqn. (5.41)] for a detailed derivation of (35). Uniting (23) and (35), we obtain
\[
R_n^*(\delta) \geq D\left( \delta \parallel \frac{1}{2} \right) + \frac{\log n}{2n} + \frac{1}{n} \log \left( \frac{1 - 2\delta}{1 - \delta} \cdot \sqrt{2\pi \delta(1 - \delta)} \right) + o\left( \frac{1}{n} \right),
\]
which is (33).

For (34), we simply have to “invert” the former result carefully. From (24) and (35), we note that for $n$ sufficiently large,
\[
\delta_n^*(R) \geq \max \left\{ a \in \mathbb{R}^+ : \frac{1}{g(a)} \cdot e^{-nD(a \parallel \frac{1}{2}) - \frac{1}{2} \log n} \leq e^{-nR} \right\}
\]
where
\[
g(a) \overset{\text{d}}{=} 2 \cdot \left( \frac{1 - 2a}{1 - a} \right) \cdot \sqrt{2\pi a(1 - a)}
\]

The random variable $K$ is lattice if for some $k_0, r \in \mathbb{R}$, the random variable $(K - k_0)/r$ is almost surely an integer, and $r$, the span of $K$, is the largest number with this property.
is some continuous function of \( a \). Let \( \{a^*_n\}_{n \in \mathbb{N}} \) be the sequence of maximizers in (36). Then because \( 0 < R < \log 2 \), the sequence \( \{a^*_n\}_{n \in \mathbb{N}} \) converges to some limit \( a = \delta_{GV}(R) \in (0, 1/2) \); so, \( g(a) \in (0, \infty) \). Due to the additional term \(-(1/2) \log n\) in the exponent in (36), we have
\[
\delta^*_n(R) \geq \delta_{GV}\left(R - \frac{\log n}{2n}\right) + \Theta\left(\frac{1}{n}\right).
\]
The final result in (34) follows by Taylor expanding \( \delta_{GV}() \).

In parallel to Theorems 4 and 5, we also provide an upper bound to \( R^*_n(\delta) \) with the aid of the so-called “twisted distributions” technique [38]. In the following, we again do not impose any structure (e.g., Hamming or additive) on the distance measure \( \mu(\cdot, \cdot) \).

**Theorem 6:** Suppose that \( \mu(\cdot, \cdot) \) is a bounded distance measure, i.e.,
\[
\max_{x^n, x^n' \in X^n} \mu(\hat{x}^n, x^n) < \infty.
\]
Given that \( \theta^* \) is the maximizer to achieve \( I_X^n(\delta) = \sup_{\theta \in \mathbb{R}} \{a \theta - \varphi_X^n(\theta)\} \) with \( P_X^n \) being the optimizer of \( \sup_{X^n} J_X^n(\delta) \), we have that under (29),
\[
R^*_n(\delta) \leq I_X^n(\delta) + \frac{4}{\sqrt{n}} \left[ \frac{\varphi_X^n(\theta^*)}{n (\varphi_X^n(\theta^*))^2} + 3 \right] \sqrt{\varphi_X^n(\theta^*)} + \frac{1}{n} \log \left( \frac{4}{n (\varphi_X^n(\theta^*))^2} + 12 \right), \tag{37}
\]
provided that \( \mu(\hat{X}^n, X^n) \) has no point mass at \( n\delta \), i.e., \( \Pr[\mu(\hat{X}^n, X^n) = n\delta] = 0 \).

**Proof:** Following the proof of Theorem 4, we define the twisted (or tilted) distribution of \( Y \) as
\[
dP_{Y(\theta)}(y) \triangleq \frac{e^{\theta y} dP_Y(y)}{M_Y(\theta)}.
\]
Then,
\[
\Pr[Y > 0] = \Pr[Y \geq 0] = \int_0^\infty dP_Y(y) = \int_0^\infty M_Y(\theta^*) e^{-\theta^* y} dP_{Y(\theta^*)} = M_Y(\theta^*) \int_0^\infty e^{-\theta^* y} dP_{Y(\theta^*)}(y), \tag{38}
\]
where \( \theta^* \) is the minimizer of \( \inf_{\theta \in \mathbb{R}} M_Y(\theta) \), satisfying \( 0 < \theta^* < \infty \) (since \( -\infty < \mathbb{E}[Y] < 0 \)). By noting that \( \Pr[Y(\theta^*) \geq 0] \) is positive,\(^6\) we let \( W \) be a nonnegative random variable with distribution

\[
dP_W(y) = \frac{dP_{Y(\theta^*)}(y)}{Pr[Y(\theta) \geq 0]}.
\]

Then, (38) can be rewritten as

\[
Pr[Y \geq 0] = M_Y(\theta^*) \int_0^\infty e^{-\theta^* y} dP_{Y(\theta^*)}(y)
= M_Y(\theta^*) \cdot Pr[Y(\theta^*) \geq 0] \int_0^\infty e^{-\theta^* y} dP_W(y)
= M_Y(\theta^*) \cdot Pr[Y(\theta^*) \geq 0] \cdot \mathbb{E}[e^{-\theta^* W}].
\]

Using the fact that \( \mathbb{E}[Y(\theta^*)] = 0 \) [29, Thm. 9.2], we obtain\(^7\)

\[
Pr[Y(\theta^*) \geq 0] \geq \frac{\mathbb{E}^2[(Y(\theta^*))^2]}{4 \mathbb{E}[(Y(\theta^*))^4]}
= \frac{(M_Y''(\theta^*))^2}{4M_Y(\theta^*)M_Y^{(4)}(\theta^*)}
= \frac{(C_Y''(\theta^*))^2}{4[C_Y''(\theta^*) + 3(C_Y''(\theta^*))^2]}
= \frac{1}{4\left[\frac{1}{n} \frac{\varphi_Y^{(4)}(-\theta)}{\varphi^{(4)}_{X_n}(-\theta)} + 3\right]}
\]

where

\[
C_Y(\theta) \triangleq \log M_Y(\theta) = n\theta a + \log \mathbb{E}[e^{-\theta \mu(X^n, X^n)}] = n (\theta a + \varphi_{X^n}(-\theta)),
\]

\[C_Y''(\theta) = n \cdot \varphi''_{X^n}(-\theta),\]

and

\[C_Y^{(4)}(\theta) = n \cdot \varphi^{(4)}_{X^n}(-\theta).\]

Notably, for a bounded distance measure \( \mu(\cdot, \cdot) \), \( \varphi_{X^n}(\theta) \) is guaranteed to be fourth-order differentiable. Using Jensen’s inequality, we derive

\[
\mathbb{E}[e^{-\theta^* W}] \geq e^{-\theta^* \mathbb{E}[W]},
\]

\(^6\)\(\Pr[Y(\theta^*) > 0] \) must be positive because if \( \Pr[Y(\theta^*) > 0] = 0 \), then together with \( \Pr[-Y(\theta^*) \geq \epsilon] \leq \mathbb{E}[-Y(\theta^*)] / \epsilon = 0 \) for arbitrary \( \epsilon > 0 \), we infer that \( \Pr[-\epsilon < Y(\theta^*) \leq 0] = 1 \) for arbitrary \( \epsilon > 0 \), which implies \( \Pr[Y(\theta^*) = 0] = 1 \). Since \( Y(\theta^*) \) and \( Y \) must have the same support, we obtain \( \Pr[Y = 0] = 1 \), and hence, a contradiction to (29), i.e., \( \mathbb{E}[Y] < 0 \), is resulted.

\(^7\)\(\Pr[Y(\theta^*) > 0] > 0 \) and \( \mathbb{E}[Y(\theta^*)] = 0 \) jointly imply \( \mathbb{E}[(Y(\theta^*))^2] > 0 \) and \( \mathbb{E}[(Y(\theta^*))^4] > 0 \), which justifies (39).
and hence
\[
E[W] = \int_{0}^{\infty} w \, dP_W(w) \\
= \int_{0}^{\infty} y \frac{dP_Y(\theta^*)}{\Pr[Y(\theta^*) \geq 0]} \\
\leq \frac{1}{\Pr[Y(\theta^*) \geq 0]} \int_{-\infty}^{\infty} |y| \, dP_Y(y) \\
= \frac{1}{\Pr[Y(\theta^*) \geq 0]} \mathbb{E}[|Y(\theta^*)|] \\
\leq \frac{1}{\Pr[Y(\theta^*) \geq 0]} \sqrt{\mathbb{E}[(Y(\theta^*))^2]} \\
= \frac{1}{\Pr[Y(\theta^*) \geq 0]} \sqrt{M_Y'(\theta^*)} \\
= \frac{1}{\Pr[Y(\theta^*) \geq 0]} \sqrt{C_Y''(\theta^*)} \\
= \frac{1}{\Pr[Y(\theta^*) \geq 0]} \sqrt{n \cdot \varphi'_{X^n}(-\theta^*)}.
\]

We conclude from all the above derivations that
\[
\Pr[Y > 0] \geq e^{-nI_X^n(\delta)} \times \frac{-4 \left[ \frac{1}{n} \varphi^{(4)}_{X^n}(-\theta^*) \varphi_{X^n}(-\theta^*)^2 + 3 \right] \sqrt{n \varphi''_{X^n}(-\theta^*)}}{4 \left[ \frac{1}{n} \varphi^{(4)}_{X^n}(-\theta^*) \varphi_{X^n}(-\theta^*)^2 + 3 \right]},
\]
which completes the proof of (37).

\[\blacksquare\]

\textbf{Remarks.} When \(\mu(\cdot, \cdot)\) is the Hamming distance measure, we have \(0 \leq \mu(\hat{x}^n, x^n) \leq n\) for \(\hat{x}^n, x^n \in X^n\) and for every \(n\). Thus, with probability one, \((1/n)\mu(\hat{X}^n, X^n)\) is not only bounded, but uniformly upper bounded in the block length \(n\), and so are its moments and cumulants. Since a twisted random variable generated from \((1/n)\mu(\hat{X}^n, X^n)\) must have the same support as \((1/n)\mu(\hat{X}^n, X^n)\), its twisted moments as well as twist cumulants are also uniformly bounded. Accordingly, \(\varphi^{(4)}_{X^n}(-\theta^*) = O(1)\) and \(\varphi''_{X^n}(-\theta^*) = O(1)\), based on which Theorems 4 and 6 yield
\[
\limsup_{n \to \infty} R_n^*(\delta) = \limsup_{n \to \infty} \sup_{X^n} J_{X^n}(\delta)
\]
and
\[
\liminf_{n \to \infty} R_n^*(\delta) = \liminf_{n \to \infty} \sup_{X^n} J_{X^n}(\delta).
\]

A new first-order characterization for the largest asymptotic code rate attainable for a sequence of \((n, M, n\delta)\)-codes under an arbitrary symmetric distance measure satisfying that \((1/n)\mu(\hat{X}^n, X^n)\) that is uniformly bounded in the block length \(n\) can thus be obtained.
In fact, the preceding theorems give us much more, assuming the conditions for them are satisfied. Theorem 5 refines the lower bound to $R^*_n(\delta)$ in Theorem 4 yielding

$$R^*_n(\delta) \geq \max \left\{ \sup_{X^n} J_X^n(\delta), D\left(\delta \left\| \frac{1}{2}\right.\right) + \frac{\log n}{2n} + \Theta\left(\frac{1}{n}\right) \right\}. $$

Theorem 6 asserts additionally that

$$R^*_n(\delta) \leq \sup_{X^n} J_X^n(\delta) + \Theta\left(\frac{1}{\sqrt{n}}\right).$$

Hence, we obtain bounds on the “second-order terms” [21], [27], [39] in the asymptotic expansion of $R^*_n(\delta)$.

VI. CONCLUSION AND OPEN PROBLEMS

This paper provides an exact formula for maximal size of a code with pairwise distance $d$. As a result, we have made progress towards resolving a long-standing open problem in information and coding theory, albeit the formula may be difficult to compute for large block lengths. Future work includes:

1) Section III-C presents an algorithmic construction of block codes subject to a given minimum distance. It would be fruitful to conduct systematic theoretical studies to examine properties of the resultant code produced by this algorithm.

2) In [40], the covering analogue of the minimum distance (packing) problem was considered from an information spectrum perspective. Naturally, one would wonder whether it is possible to prove a similar finite length formula for the covering analogue.

Acknowledgements

The work of Ling-Hua Chang and Po-Ning Chen is supported by the Ministry of Science and Technology (MoST), Taiwan, R.O.C., under grant 105-2221-E-009-009-MY3. The work of Carol Wang and Vincent Tan is supported by a Singapore Ministry of Education (MoE) Tier 2 grant (R-263-000-B61-112). The work of Yungkin S. Han is support by the National Natural Science Foundation of China (Grant No. 61671007).

A special acknowledgement is given to Dr. Mladen Kovačević, who brought [5] to the authors’ attention.

REFERENCES

[1] R. J. McEliece, E. R. Rodemich, H. Rumsey, Jr., and L. R. Welch, “New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities,” IEEE Trans. Inform. Theory, vol. 23, no. 2, pp. 157–166, March 1977.
[2] S. N. Litsyn and M. A. Tsfasman, “A note on lower bounds,” *IEEE Trans. Inform. Theory*, vol. IT-32, no. 5, pp. 705–706, September 1986.

[3] 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, September 1987.

[4] S. G. Vladut, “An exhaustion bound for algebraic-geometric modular codes,” *Probl. Inform. Transm.*, vol. 23, pp. 22–34, September 1987.

[5] V. D. Kolesnik and V. Y. Krachkovsky, “Lower bounds on achievable rates for limited bitshift correcting codes,” *IEEE Trans. Inform. Theory*, vol. 40, no. 5, pp. 1443–1458, September 1994.

[6] M. Svanstrom, “A lower bound for ternary constant weight codes,” *IEEE Trans. Inform. Theory*, vol. 43, no. 5, pp. 1630–1632, September 1997.

[7] T. K. Moon, *Error Correction Coding: Mathematical Methods and Algorithms*. Wiley, 2005.

[8] H. Stichtenoth and C. Xing, “Excellent nonlinear codes from algebraic function fields,” *IEEE Trans. Inform. Theory*, vol. 51, no. 11, pp. 4044–4046, October 2005.

[9] P. Gaborit and G. Zemor, “Asymptotic improvement of the Gilbert-Varshamov bound for linear codes,” *IEEE Trans. Inform. Theory*, vol. 54, no. 9, pp. 3865–3872, August 2008.

[10] A. Bassa, P. Beelen, A. Garcia, and H. Stichtenoth, “An improvement of the Gilbert-Varshamov bound over nonprime fields,” *IEEE Trans. Inform. Theory*, vol. 60, no. 7, pp. 3859–3861, April 2014.

[11] J. H. van Lint, *Introduction to Coding Theory*, 2nd ed. New York: Springer Verlag, 1992.

[12] V. D. Goppa, “Codes on algebraic curves,” *Soviet Mathematics Doklady*, vol. 1, no. 24, pp. 170–172, 1981.

[13] M. A. Tsfasman, S. G. Vladut, and T. Zink, “Modular curves, Shimura curves, and Goppa codes better than the Varshamov-Gilbert bound,” *Weinheim: Wiley*, vol. 109, no. 1, pp. 21–28, 1982.

[14] G. van der Geer and J. H. van Lint, *Introduction to Coding Theory and Algebraic Geometry*. Basel, Switzerland: Birkhauser, 1988.

[15] V. A. Zinoviev and S. N. Litsyn, “Codes that exceed the Gilbert bound,” *Probl. Pered. Inform*, vol. 21, no. 1, pp. 105–108, 1985.

[16] P.-N. Chen, T.-Y. Lee, and Y. S. Han, “Distance-spectrum formulas on the largest minimum distance of block codes,” *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 869–885, May 2000.

[17] S. Verdú and T.-S. Han, “A general formula for channel capacity,” *IEEE Trans. Inform. Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.

[18] T.-S. Han, *Information-Spectrum Method in Information Theory*. Springer Verlag, 2003.

[19] P.-N. Chen and F. Alajaji, “Optimistic shannon coding theorems for arbitrary single-user systems,” *IEEE Trans. Inform. Theory*, vol. 45, no. 7, pp. 2623–2629, November 1999.

[20] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, New York, and North-Holland: Oxford University Press, 1983.

[21] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[22] V. Kostina and S. Verdú, “Fixed-length lossy compression in the finite blocklength regime,” *IEEE Trans. Inform. Theory*, vol. 58, no. 6, pp. 3309–3338, 2012.

[23] ———, “Lossy joint source-channel coding in the finite blocklength regime,” *IEEE Trans. Inform. Theory*, vol. 59, no. 5, pp. 2545–2575, 2013.

[24] M. C. Gursoy, “Throughput analysis of buffer-constrained wireless systems in the finite blocklength regime,” in *Proc. IEEE Int. Conf. Commun.*, Kyoto, Japan, June 5–9, 2011, pp. 1–5.

[25] T. J. Riedl, T. P. Coleman, and A. C. Singer, “Finite block-length achievable rates for queuing timing channels,” in *Proc. IEEE Inf. Theory Workshop*, Paraty, Brazil, October 16–20, 2011, pp. 200–204.
[26] J. Scarlett, A. Martinez, and A. Guillen i Fàbregas, “Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations,” *IEEE Trans. Inform. Theory*, vol. 60, no. 5, pp. 2647–2666, 2014.

[27] V. Y. F. Tan, “Asymptotic estimates in information theory with non-vanishing error probabilities,” *Foundations and Trends*® in Communications and Information Theory, vol. 11, no. 1–2, pp. 1–184, September 2014.

[28] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. Springer, 1998.

[29] G. van der Geer and J. H. van Lint, *Large Deviation Techniques in Decision, Simulation, and Estimation*. New York, NY, USA: Wiley, 1990.

[30] S. Lin and D. J. Costello, Jr., *Error Control Coding*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2004.

[31] D. Huffman, “A method for the construction of minimum-redundancy codes,” *Proceedings of the IRE*, vol. 40, no. 9, pp. 1098–1101, 1952.

[32] L.-H. Chang, C. Wang, P.-N. Chen, Y. S. Han, and V. Y. F. Tan, “Distance spectrum formula for the largest minimum hamming distance of finite-length binary block codes,” in *Proc. IEEE Inf. Theory Workshop*, Kaohsiung, Taiwan, November 6–10, 2017, submitted.

[33] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.

[34] A. Barg and G. D. Forney, “Random codes: Minimum distances and error exponents,” *IEEE Trans. Inform. Theory*, vol. 49, no. 9, pp. 2568–2573, September 2002.

[35] T. Jiang and A. Vardy, “Asymptotic improvement of the Gilbert-Varshamov bound on the size of binary codes,” *IEEE Trans. Inform. Theory*, vol. 50, no. 8, pp. 1655–1664, November 2004.

[36] R. Bahadur and R. R. Rao, “On deviations of the sample mean,” *Ann. Math. Statist.*, vol. 31, no. 4, pp. 1015–1027, 1960.

[37] P. Moulin, “The log-volume of optimal codes for memoryless channels, asymptotically within a few nats,” *IEEE Trans. Inform. Theory*, vol. 63, no. 4, pp. 2278–2313, April 2017.

[38] P.-N. Chen, “Generalization of Gartner-Ellis theorem,” *IEEE Trans. Inform. Theory*, vol. 46, no. 7, pp. 2752–2760, 2000.

[39] V. Strassen, “Asymptotische Abschätzungen in Shannons Informationstheorie,” in *Trans. Third Prague Conf. Inf. Theory*, Prague, 1962, pp. 689–723, http://www.math.cornell.edu/~pmlut/strassen.pdf.

[40] P.-N. Chen and Y. S. Han, “Asymptotic minimum covering radius of block codes,” *SIAM J. Discrete Math.*, vol. 14, no. 4, pp. 549–564, 2001.