SIAM J. DISCRETE MATH.
Vol. 14, No. 4, pp. 549–564

ASYMPTOTIC MINIMUM COVERING RADIUS OF BLOCK CODES*

PO-NING CHEN† AND YUNGHSIANG S. HAN‡

Abstract. In this paper, we restudy the covering radius of block codes from an information theoretic point of view by ignoring the combinatorial formulation of the problem. In the new setting, the formula of the statistically defined *minimum covering radius*, for which the probability mass of uncovered space by M spheres can be made arbitrarily small, is reduced to a minimization of a statistically defined spectrum formula among codeword-selecting distributions. The advantage of the new view is that no assumptions need to be made on the code alphabet (such as finite, countable, etc.) and the distance measure (such as additive, symmetric, bounded, etc.) in the problem transformation, and hence the spectrum formula can be applied in most general situations. We next address a sufficient condition under which uniform codeword-selecting distribution minimizes the spectrum formula. With the condition, the asymptotic minimum covering radius for *block codes under J-ary quantized channels* and *constant weight codes under Hamming distance measure* are determined to display the usage of the spectrum formula.

Key words. covering radius, block codes, information spectrum

AMS subject classifications. 94B65, 94A24

PII. S0895480100379993

1. Introduction.

We first introduce the notations used in this paper. We denote the n-tuple alphabet by $\mathcal{X}^n = \mathcal{X} \times \mathcal{X} \times \cdots \times \mathcal{X}$. For any two elements $x^n = (x_0, x_1, \ldots, x_{n-1})$ and $y^n = (y_0, y_1, \ldots, y_{n-1})$ in $\mathcal{X}^n$, we use $\mu_n(x^n, y^n)$ to denote the n-fold measure[1] on the "distance" between them. In our study, the codewords are drawn from a pregiven codeword set $\mathcal{S}_n$, which can be either the entire space $\mathcal{X}^n$ or its proper subset. Such a generalization will be useful for some specific applications, such as constant weight codes, where the codewords are drawn from a subset (of $\mathcal{X}^n$) containing only words of fixed weight.

Based on the above notations, the problem on the minimum covering radius becomes the following: for given M and $\mathcal{S}_n$, determine the minimum radius $\rho(M, \mathcal{S}_n)$ for which M spheres that center at M selected elements from $\mathcal{S}_n$ jointly cover the entire space $\mathcal{X}^n$. Specifically,

$$\rho(M, \mathcal{S}_n) \stackrel{\triangle}{=} \min_{\substack{\mathcal{C} \subset \mathcal{S}_n \\ |\mathcal{C}| = M}} \max_{x^n \in \mathcal{X}^n} \min_{y^n \in \mathcal{C}} \mu_n(x^n, y^n), \tag{1}$$

*Received by the editors October 23, 2000; accepted for publication (in revised form) August 16, 2001; published electronically October 23, 2001.
http://www.siam.org/journals/sidma/14-4/37999.html

†Department of Communications Engineering, National Chiao Tung University, Hsin Chu, Taiwan 30050, ROC (poning@cc.nctu.edu.tw). The work of this author was supported by the National Science Council of Taiwan under project code NSC 89-2213-E-009-106.

‡Department of Computer Science and Information Engineering, National Chi Nan University, Nan Tou, Taiwan 545, ROC (yshan@csie.ncnu.edu.tw). The work of this author was supported by the National Science Council of Taiwan under project code NSC 89-2213-E-260-002.

[1]Conventionally, a *distance* [16, p. 139] should satisfy the properties of (i) nonnegativity, (ii) being zero iff two points coincide, (iii) symmetry, and (iv) triangle inequality. The spectrum formula derived in this paper, however, is applicable to any measurable function defined over the 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. For simplicity, we will abbreviate the *generalized distance* function simply as the *distance* function in the remaining part of the paper.

where $|\mathcal{C}|$ denotes the size of the set $\mathcal{C}$. Here we do not assume that the M elements drawn from $\mathcal{S}_n$ must be distinct. In other words, one can choose M identical elements from $\mathcal{S}_n$ as long as the resultant codebook gives the minimum covering radius. In addition, we implicitly assume that $|\mathcal{S}_n| > 0$.

The problem of determining the covering radius has been studied by many researchers [2], [5], [6], [7], [8], [9], [10], [11], [12], [14], [15], [17], [18], [19], [20], [21] among which [2], [5], [6], [11], [17], and [19] focused on its asymptotic behavior with respect to block length n under an exponentially increasing size $M = e^{nR}$ and a fixed rate R. Specifically, [5], [17], and [19] investigate this problem based on combinatorial techniques, while the studies in [2] and [11] introduce probabilistic approaches. All the mentioned works concentrated on codes transmitted over binary symmetric channel, where Hamming distance is the only distance measure.

In this paper, we employ a new notion from information-spectrum methodologies [3], [13] to determine the asymptotic minimum covering radius among (a prespecified class of) block codes. As a result, the asymptotic minimum covering radius formula can be established under any alphabet $\mathcal{X}^n$ and any measure on the "distance" between elements in $\mathcal{X}^n$. With the new expression, we can now, for example, investigate the asymptotic minimum covering radius not only for Hamming distance but also for the "quantized distance measure" defined for codes transmitted over quantized channels.

The rest of the paper is organized as follows. In section 2, we transform the problem of determining the *asymptotic minimum covering radius* among block codes into one that minimizes a spectrum function $\Omega_{\boldsymbol{Y}\|\boldsymbol{X}}(R)$ among all codeword-selecting distributions $\boldsymbol{Y}$. A sufficient condition under which uniform $\boldsymbol{Y}$ minimizes the spectrum function is next addressed in section 3. Based on the sufficient condition, the asymptotic minimum covering radius for arbitrary block codes under J-ary quantized channels and constant weight codes under Hamming distance are established in section 4 to display the usage of the spectrum formula.

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

2. Asymptotic minimum covering radius for block codes. Define a sphere centered at y^n with radius r as

$$\mathcal{B}_r(y^n) \stackrel{\triangle}{=} \{x^n \in \mathcal{X}^n : \mu_n(x^n, y^n) \le r\}.$$

DEFINITION 2.1 (minimum α-covering radius under codeword set $\mathcal{S}_n$ and covering distribution P_{X^n}). *Fix $\alpha \in [0,1]$. The minimum α-covering radius under codeword set $\mathcal{S}_n$ and distribution P_{X^n} is given by*

$$\rho_\alpha(M, \mathcal{S}_n \| X^n) \stackrel{\triangle}{=} \inf_{\substack{\mathcal{C}\subset\mathcal{S}_n \\ |\mathcal{C}|=M}} \inf\left\{r \in \Re : P_{X^n}\left(\bigcup_{y^n\in\mathcal{C}} \mathcal{B}_r(y^n)\right) \ge \alpha\right\}.$$

The *α-covering radius* for one specific block code is the smallest sphere radius for which the probability mass of all words, covered by M spheres, is no smaller than α (cf. Figure. 2.1). The probability mass placed on each element x^n in $\mathcal{X}^n$ is assumed to be defined through the covering distribution P_{X^n}. Taking the minimum one among all α-covering radii yields the *minimum α-covering radius*.

It can be verified that the conventional definition of the minimum covering radius $\rho(M, \mathcal{S}_n)$ (cf. (1)) is exactly the 1-*covering radius* under a full-support[2] covering

[2]The support of a distribution is the smallest set with probability mass being equal to 1. Here, "full-support" means the support of P_{X^n} is $\mathcal{X}^n$.

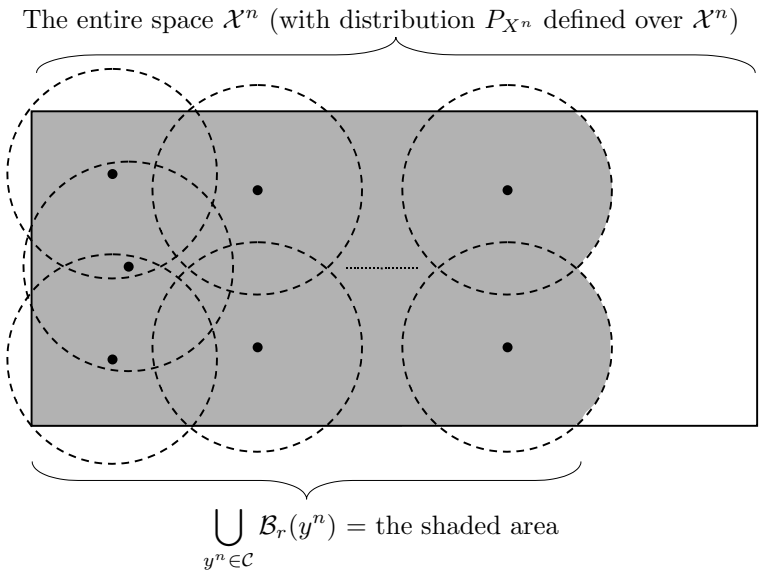

FIG. 2.1. *The α-covering radius for a block code is the smallest radius r such that $P_{X^n}(\cup_{y^n \in \mathcal{C}} \mathcal{B}_r(y^n)) \geq \alpha$.*

distribution P_{X^n}. Specifically, given that the support of P_{X^n} is $\mathcal{X}^n$,

$$\begin{aligned}\rho(M, \mathcal{S}_n) &\stackrel{\triangle}{=} \min_{\substack{\mathcal{C} \subset \mathcal{S}_n \\ |\mathcal{C}|=M}} \max_{x^n \in \mathcal{X}^n} \min_{y^n \in \mathcal{C}} \mu_n(x^n, y^n) \\ &= \inf_{\substack{\mathcal{C} \subset \mathcal{S}_n \\ |\mathcal{C}|=M}} \inf \left\{ r \in \Re : P_{X^n}\left(\bigcup_{y^n \in \mathcal{C}} \mathcal{B}_r(y^n) \right) = 1 \right\} \\ &= \rho_1(M, \mathcal{S}_n \,\|X^n)\,.\end{aligned}$$

Accordingly, the conventional asymptotic covering radius problem is to find the limit, as $n \to \infty$, of the quantity

$$\frac{1}{n}\rho(M, \mathcal{S}_n) = \frac{1}{n}\rho_1(M, \mathcal{S}_n \| X^n)$$

under a full-support distribution P_{X^n} and 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. Now if the targeted quantity becomes $(1/n)\rho_\alpha(M, \mathcal{S}_n \| X^n)$ instead of $(1/n)\rho(M, \mathcal{S}_n)$, then the full-support assumption on covering distribution P_{X^n} can be relaxed. Equipped with the new setting, one can place larger probability mass on those elements that are considered more essential (to cover) than other elements.

The concept of our method is similar to that of the random coding technique employed in the channel reliability function [1]. Each codeword is assumed to be selected independently of all others from $\mathcal{S}_n$ through a generic distribution P_{Y^n} with $P_{Y^n}(\mathcal{S}_n) = 1$. Then the sphere centered at each random codeword with radius r becomes a *random* variable and so does the resultant codebook. For convenience, we use $\mathcal{B}_r(Y^n)$ and $\boldsymbol{\mathcal{C}}$ to denote the *random* sphere and the *random* codebook, respectively.

LEMMA 2.2. *Fix a sequence of codeword sets*

$$\boldsymbol{\mathcal{S}} = \{\mathcal{S}_n\}_{n \geq 1}, \text{ where } \mathcal{S}_n \subset \mathcal{X}^n \text{ and } |\mathcal{S}_n| > 0,$$

and a triangular array of covering distributions $P_{\boldsymbol{X}} = \{P_{X^n}\}_{n=1}^{\infty}$. *For any triangular-array codeword-selecting process*

$$\boldsymbol{Y} = \boldsymbol{Y}(\boldsymbol{S}) = \left\{Y^n = \left(Y_1^{(n)}, Y_2^{(n)}, \ldots, Y_n^{(n)}\right)\right\}_{n=1}^{\infty}$$

satisfying $P_{Y^n}(\mathcal{S}_n) = 1$ *for each* n, *and any* $\alpha \in [0,1)$,

$$\limsup_{n\to\infty} \frac{1}{n}\rho_\alpha(M, \mathcal{S}_n \| X^n) \le \bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R) \tag{2}$$

and

$$\liminf_{n\to\infty} \frac{1}{n}\rho_\alpha(M, \mathcal{S}_n \| X^n) \le \underline{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R), \tag{3}$$

where[3]

$$\bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R) \stackrel{\triangle}{=} \inf\left\{a \in \Re : \limsup_{n\to\infty} E_{X^n}\left[\Pr\left(\frac{1}{n}\mu_n(X^n, Y^n) > a \,\middle|\, X^n\right)^M\right] = 0\right\}$$

and

$$\underline{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R) \stackrel{\triangle}{=} \inf\left\{a \in \Re : \liminf_{n\to\infty} E_{X^n}\left[\Pr\left(\frac{1}{n}\mu_n(X^n, Y^n) > a \,\middle|\, X^n\right)^M\right] = 0\right\}.$$

Proof. We will prove only (2). Inequality (3) can be proved by simply following the same procedure.

Let

$$\lambda \stackrel{\triangle}{=} \inf\left\{a \in \Re \;:\; \limsup_{n\to\infty} E_{X^n}\left[\Pr\left(\frac{1}{n}\mu_n(X^n, Y^n) > a \,\middle|\, X^n\right)^M\right] = 0\right\}.$$

By definition,

$$\limsup_{n\to\infty} E_{X^n}\left[\Pr\left(\frac{1}{n}\mu_n(X^n, Y^n) > \lambda + \varepsilon \,\middle|\, X^n\right)^M\right] = 0 \tag{4}$$

for any $\varepsilon > 0$. Equation (4) then implies that for sufficiently large n,

$$E_{X^n}\left[\Pr\left(\frac{1}{n}\mu_n(X^n, Y^n) > \lambda + \varepsilon \,\middle|\, X^n\right)^M\right] < \frac{1-\alpha}{2}. \tag{5}$$

Now for a given codebook $\mathcal{C}$, define

$$\phi_r(x^n|\,\mathcal{C}) \stackrel{\triangle}{=} \begin{cases} 1 & \text{if } x^n \in \bigcup\limits_{y^n \in \mathcal{C}} \mathcal{B}_r(y^n), \\ 0 & \text{otherwise.} \end{cases}$$

[3]Here, we adopt the notation that

$$E_{X^n}\left[\Pr\left(\frac{1}{n}\mu_n(X^n, Y^n) > a \,\middle|\, X^n\right)^M\right] = \int_{\mathcal{X}^n} \left(P_{Y^n|X^n}\{y^n \in \mathcal{S}_n : \mu_n(x^n, y^n) > na\}\right)^M dP_{X^n}(x^n)$$
$$= \int_{\mathcal{X}^n} \left(P_{Y^n}\{y^n \in \mathcal{S}_n : \mu_n(x^n, y^n) > na\}\right)^M dP_{X^n}(x^n),$$

where the last step follows since X^n and Y^n are implicitly assumed to be independent.

Then

$$E\left[P_{X^n}\left\{\bigcup_{y^n\in\boldsymbol{\mathcal{C}}}\mathcal{B}_r(y^n)\right\}\right] = E\left[\int_{\mathcal{X}^n}\phi_r(x^n|\boldsymbol{\mathcal{C}})dP_{X^n}(x^n)\right]$$
$$= \int_{\mathcal{X}^n} E\left[\phi_r(x^n|\boldsymbol{\mathcal{C}})\right] dP_{X^n}(x^n),$$

where the expectation is taken with respect to the random codebook $\boldsymbol{\mathcal{C}}$ drawn independently from $\mathcal{S}_n$ according to codeword-selecting process Y^n. By definition,

$$\begin{aligned} E\left[\phi_r(x^n|\boldsymbol{\mathcal{C}})\right] &= 1-\Pr\left\{x^n\notin\mathcal{B}_r(Y_1^n) \text{ and } x^n\notin\mathcal{B}_r(Y_2^n) \text{ and } \cdots \text{ and } x^n\notin\mathcal{B}_r(Y_M^n)\right\}\\ &= 1-\left(\Pr\left\{x^n\notin\mathcal{B}_r(Y^n)\right\}\right)^M\\ &= 1-\left(P_{Y^n}\left\{y^n\in\mathcal{S}_n : \mu_n(x^n,y^n)>r\right\}\right)^M. \end{aligned}$$

Therefore, by taking $r=n(\lambda+\varepsilon)$ and for those n satisfying (5), we obtain

$$\begin{aligned} &E\left[P_{X^n}\left\{\bigcup_{y^n\in\boldsymbol{\mathcal{C}}}\mathcal{B}_{n(\lambda+\varepsilon)}(y^n)\right\}\right]\\ &= \int_{\mathcal{X}^n}\left[1-\left(P_{Y^n}\left\{y^n\in\mathcal{S}_n : \mu_n(x^n,y^n)>n(\lambda+\varepsilon)\right\}\right)^M\right]dP_{X^n}(x^n)\\ &= 1-E_{X^n}\left[\left(\Pr\left\{\frac{1}{n}\mu_n(X^n,Y^n)>\lambda+\varepsilon\middle| X^n\right\}\right)^M\right]\\ &> 1-\frac{1-\alpha}{2}>\alpha, \end{aligned}$$

which implies that among all possible selections, there exists one codebook $\mathcal{C}\subset\mathcal{S}_n$ satisfying

$$P_{X^n}\left\{\bigcup_{y^n\in\mathcal{C}}\mathcal{B}_{n(\lambda+\varepsilon)}(y^n)\right\}>\alpha.$$

Consequently, $n(\lambda+\varepsilon)\ge\rho_\alpha(M,\mathcal{S}_n\|X^n)$ or, equivalently,

$$\lambda+\varepsilon\ge\frac{1}{n}\rho_\alpha(M,\mathcal{S}_n\|X^n) \tag{6}$$

for all sufficiently large n. By taking the limsup with respect to n on (6), the proof is completed since ϵ is arbitrary. □

We are now ready to prove the main theorems of the paper.

THEOREM 2.3 (upper bound). *Fix a sequence of codeword set $\boldsymbol{S}=\{\mathcal{S}_n\}_{n\ge1}$. For any covering process $\boldsymbol{X}=\{X^n\}_{n=1}^\infty$,*

$$\sup_{0\le\alpha<1}\limsup_{n\to\infty}\frac{1}{n}\rho_\alpha(M,\mathcal{S}_n\|X^n)\le\inf_{\boldsymbol{Y}(\boldsymbol{S})}\bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R)$$

and

$$\sup_{0\le\alpha<1}\liminf_{n\to\infty}\frac{1}{n}\rho_\alpha(M,\mathcal{S}_n\|X^n)\le\inf_{\boldsymbol{Y}(\boldsymbol{S})}\underline{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R),$$

where the infimum is taken over all processes $\boldsymbol{Y}$ with $P_{Y^n}(\mathcal{S}_n)=1$ (which for convenience is denoted by $\boldsymbol{Y}(\boldsymbol{S})$ in what follows).

Proof. The theorem follows immediately from Lemma 2.2. □

THEOREM 2.4 (lower bound). *Fix a sequence of codeword set* $\boldsymbol{S}=\{\mathcal{S}_n\}_{n\geq 1}$. *For any covering process* $\boldsymbol{X}$,

$$\sup_{0\leq\alpha<1}\limsup_{n\to\infty}\frac{1}{n}\rho_\alpha(M,\mathcal{S}_n\|X^n)\geq\sup_{\gamma>0}\inf_{\boldsymbol{Y}(\boldsymbol{S})}\bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R+\gamma) \tag{7}$$

and

$$\sup_{0\leq\alpha<1}\liminf_{n\to\infty}\frac{1}{n}\rho_\alpha(M,\mathcal{S}_n\|X^n)\geq\sup_{\gamma>0}\inf_{\boldsymbol{Y}(\boldsymbol{S})}\underline{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R+\gamma). \tag{8}$$

Proof. Again, we will prove only (7), since (8) can be proved in a similar fashion. To prove the inequality in (7), it suffices to prove the existence of $\boldsymbol{Y}(\boldsymbol{S})$ such that

$$\sup_{0\leq\alpha<1}\limsup_{n\to\infty}\frac{1}{n}\rho_\alpha(M,\mathcal{S}_n\|X^n)+\varepsilon\geq\bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R+\gamma)$$

for any $\varepsilon>0$. This can be justified as follows.

Fix $\varepsilon>0$ and define

$$\lambda\triangleq\sup_{0\leq\alpha<1}\limsup_{n\to\infty}\frac{1}{n}\rho_\alpha(M,\mathcal{S}_n\|X^n).$$

By definition of infimum (cf. Definition 2.1), for any integer $m>1$, there exists a code $\mathcal{C}_n(m)\subset\mathcal{S}_n$ of size M (for each n) such that

$$\rho_{(m-1)/m}(M,\mathcal{S}_n\|X^n)\geq\inf\left\{r\in\Re:P_{X^n}\left[\bigcup_{y^n\in\mathcal{C}_n(m)}\mathcal{B}_r(y^n)\right]\geq\frac{m-1}{m}\right\}-\varepsilon.$$

Therefore,

$$\begin{aligned}\lambda&\geq\limsup_{n\to\infty}\frac{1}{n}\rho_{(m-1)/m}(M,\mathcal{S}_n\|X^n)\\&\geq\limsup_{n\to\infty}\frac{1}{n}\inf\left\{r\in\Re:P_{X^n}\left[\bigcup_{y^n\in\mathcal{C}_n(m)}\mathcal{B}_r(y^n)\right]\geq\frac{m-1}{m}\right\},\end{aligned}$$

which indicates the existence of N_m such that for all $n\geq N_m$,

$$\frac{1}{n}\inf\left\{r\in\Re:P_{X^n}\left[\bigcup_{y^n\in\mathcal{C}_n(m)}\mathcal{B}_r(y^n)\right]\geq\frac{m-1}{m}\right\}<\lambda+\varepsilon.$$

Hence, for $n\geq N_m$,

$$P_{X^n}\left[\bigcup_{y^n\in\mathcal{C}_n(m)}\mathcal{B}_{n(\lambda+\varepsilon)}(y^n)\right]\geq\frac{m-1}{m}.$$

Now, for $\max_{1<i\leq m}N_m\leq n<\max_{1<i\leq(m+1)}N_m$, choose Y^n to be a uniform distribution over $\mathcal{C}_n(m)$ and let

$$\mathcal{V}_n\triangleq\bigcup_{y^n\in\mathcal{C}_n(m)}\mathcal{B}_{n(\lambda+\varepsilon)}(y^n).$$

By noting that for any $x^n \in \mathcal{V}_n$ there exists $y^n \in \mathcal{C}_n(m)$ satisfying

$$\frac{1}{n}\mu_n(x^n, y^n) \leq \lambda + \varepsilon,$$

we obtain for all $n \geq \max_{1<i\leq m} N_m$,

$$\begin{aligned}
&E_{X^n}\left[\Pr\left(\frac{1}{n}\mu_n(X^n, Y^n) > \lambda + \varepsilon \middle| X^n\right)^{e^{n(R+\gamma)}}\right] \\
&= \int_{\mathcal{V}_n} \Pr\left(\frac{1}{n}\mu_n(x^n, Y^n) > \lambda + \varepsilon\right)^{Me^{n\gamma}} dP_{X^n}(x^n) \\
&\quad + \int_{\mathcal{V}_n^c} \Pr\left(\frac{1}{n}\mu_n(x^n, Y^n) > \lambda + \varepsilon\right)^{Me^{n\gamma}} dP_{X^n}(x^n) \\
&\leq \int_{\mathcal{V}_n} \left(1 - \frac{1}{M}\right)^{Me^{n\gamma}} dP_{X^n}(x^n) + \int_{\mathcal{V}_n^c} 1 \; dP_{X^n}(x^n) \\
&\leq \left(1 - \frac{1}{M}\right)^{Me^{n\gamma}} + \left(1 - \frac{m-1}{m}\right) \\
&= \left(1 - \frac{1}{M}\right)^{Me^{n\gamma}} + \frac{1}{m},
\end{aligned}$$

where the superscript "c" applied on $\mathcal{V}_n$ represents the set complementary operation. This result immediately gives

$$\limsup_{n\to\infty} E_{X^n}\left[\Pr\left(\frac{1}{n}\mu_n(X^n, Y^n) > \lambda + \varepsilon \middle| X^n\right)^{Me^{n\gamma}}\right] \leq \frac{1}{m}.$$

Since we can take arbitrarily large m,

$$\limsup_{n\to\infty} E_{X^n}\left[\Pr\left(\frac{1}{n}\mu_n(X^n, Y^n) > \lambda + \varepsilon \middle| X^n\right)^{Me^{n\gamma}}\right] = 0.$$

Consequently, $\bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R+\gamma) \leq \lambda + \varepsilon$. □

Theorems 2.3 and 2.4 together conclude to

$$\inf_{\boldsymbol{Y}(\boldsymbol{S})} \bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R+\gamma) \leq \sup_{0\leq\alpha<1} \limsup_{n\to\infty} \frac{1}{n}\rho_\alpha(M, \mathcal{S}_n\|X^n) \leq \inf_{\boldsymbol{Y}(\boldsymbol{S})} \bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R)$$

and

$$\inf_{\boldsymbol{Y}(\boldsymbol{S})} \bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R+\gamma) \leq \sup_{0\leq\alpha<1} \liminf_{n\to\infty} \frac{1}{n}\rho_\alpha(M, \mathcal{S}_n\|X^n) \leq \inf_{\boldsymbol{Y}(\boldsymbol{S})} \underline{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R)$$

for every $\gamma > 0$. A direct interpretation on the quantity of

$$\sup_{0\leq\alpha<1} \limsup_{n\to\infty} \frac{1}{n}\rho_\alpha(M, \mathcal{S}_n\|X^n) \quad \left(\text{resp., } \sup_{0\leq\alpha<1} \liminf_{n\to\infty} \frac{1}{n}\rho_\alpha(M, \mathcal{S}_n\|X^n)\right)$$

is as follows. It represents, in asymptotics, the minimum radius with which M spheres centered at some node in $\mathcal{S}_n$ can cover *almost* all the words in the support of P_{X^n}.

In other words, the overall probability mass of those words that are not covered can be made arbitrarily small. This requirement is a little weaker if compared to the conventional definition of covering radius, which dictates (as interpreted probabilistically under full-support covering distribution) the probability of all uncovered words being *zero*.

3. A sufficient condition for the minimization of $\bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R)$ and $\underline{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R)$. The previous section shows that the asymptotic minimum covering radius can be determined by finding

$$\inf_{\boldsymbol{Y}(\boldsymbol{S})} \bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R) \quad \text{and} \quad \inf_{\boldsymbol{Y}(\boldsymbol{S})} \underline{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R). \tag{9}$$

A natural query following this result is "What is the minimizer for (9)?" In our view, there may not exist a universal solution for this query (since in the spectrum formula, there is no restriction on the distance measure and code alphabet, as well as codeword-selection set and covering distribution.) However, when the distance measure $\mu_n(\cdot,\cdot)$ is symmetric, and a full-support uniform covering distribution under finite alphabet is taken, a sufficient condition under which the uniform $\boldsymbol{Y}$ (over the codeword set) is indeed the desired minimizer can be established. We justify this finding as follows.

By rewriting the spectrum formula as

$$\begin{aligned}\bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R) &\stackrel{\triangle}{=} \inf\left\{a \in \Re : \limsup_{n\to\infty} E_{X^n}\left[\Pr\left(\frac{1}{n}\mu_n(X^n,Y^n) > a \middle| X^n\right)^M\right] = 0\right\} \\ &= \inf\left\{a \in \Re : \limsup_{n\to\infty} \sum_{x^n\in\mathcal{X}^n} \frac{1}{q^n} P_{Y^n}[y^n \in \mathcal{S}_n : x^n \notin \mathcal{B}_{na}(y^n)]^M = 0\right\},\end{aligned}$$

where $q \stackrel{\triangle}{=} |\mathcal{X}|$, we note that if for any a,

$$\sum_{x^n\in\mathcal{X}^n} P_{Y^n}\left[y^n \in \mathcal{S}_n : x^n \notin \mathcal{B}_{na}(y^n)\right]^M \tag{10}$$

is minimized by uniform Y^n over $\mathcal{S}_n$, so is $\bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R)$. The next lemma then gives the basis for the validity of (10) being minimized by the Y^n that is uniformly distributed over $\mathcal{S}_n$.

LEMMA 3.1. *The function $(x_1 + x_2 + \cdots + x_k)^M$ is convex*[4] *in $(x_1, x_2, \ldots, x_k)$ over any pregiven convex set for any positive integer M.*

Proof. We prove this lemma by induction.

1. $M = 1$. The function $(x_1+x_2+\cdots+x_k)$ is apparently convex in $(x_1,x_2,\ldots,x_k)$ over the desired convex set.

2. Assume that the above claim holds for $(M-1)$. Then

$$\begin{aligned}&\lambda(x_1+\cdots+x_k)^M + (1-\lambda)(y_1+\cdots+y_k)^M \\ &-[\lambda(x_1+\cdots+x_k) + (1-\lambda)(y_1+\cdots+y_k)]^M \\ \geq\ &\lambda(x_1+\cdots+x_k)^M + (1-\lambda)(y_1+\cdots+y_k)^M \\ &-\left[\lambda(x_1+\cdots+x_k)^{M-1} + (1-\lambda)(y_1+\cdots+y_k)^{M-1}\right]\end{aligned}$$

[4] A subset of real vector space is said to be *convex* if $\boldsymbol{x} \in \mathcal{A}$ and $\boldsymbol{y} \in \mathcal{A}$ imply that $\lambda\boldsymbol{x}+(1-\lambda)\boldsymbol{y} \in \mathcal{A}$ for all $\lambda \in [0,1]$. A real-valued function $f(\boldsymbol{x})$ that is defined over a convex set $\mathcal{A}$ is called a convex function if for all $\lambda \in [0,1]$, and for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathcal{A}$, $f(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1-\lambda)f(\boldsymbol{y})$.

$$
\begin{aligned}
&\times[\lambda(x_1+\cdots+x_k)+(1-\lambda)(y_1+\cdots+y_k)]\\
&=\lambda(1-\lambda)(x_1+\cdots+x_k)^M+\lambda(1-\lambda)(y_1+\cdots+y_k)^M\\
&\quad-\lambda(1-\lambda)(x_1+\cdots+x_k)^{M-1}(y_1+\cdots+y_k)\\
&\quad-\lambda(1-\lambda)(x_1+\cdots+x_k)(y_1+\cdots+y_k)^{M-1}\\
&=\lambda(1-\lambda)[(x_1+\cdots+x_k)-(y_1+\cdots+y_k)]\\
&\quad\times\left[(x_1+\cdots+x_k)^{M-1}-(y_1+\cdots+y_k)^{M-1}\right]\geq 0. \qquad \square
\end{aligned}
$$

LEMMA 3.2. *Under the assumption that* $|\mathcal{X}|<\infty$,

$$\sum_{x^n\in\mathcal{X}^n} P_{Y^n}\left[y^n\in\mathcal{S}_n : x^n\notin\mathcal{B}_{na}(y^n)\right]^M$$

is a convex function in P_{Y^n} *over the convex set*

$$\left\{(P_{Y^n}(y_1^n),\ldots,P_{Y^n}(y_N^n))\in[0,1]^N \;:\; \sum_{i=1}^N P_{Y^n}(y_i^n)=1\right\},$$

where $N=|\mathcal{S}_n|$.

Proof. This can be proved by Lemma 3.1 and the observation that a finite sum of convex functions is convex. $\square$

When the distance measure is symmetric, the quantity (10) can be reformulated as

$$\sum_{x^n\in\mathcal{X}^n} P_{Y^n}\left[\mathcal{S}_n/\mathcal{B}_{na}(x^n)\right]^M, \tag{11}$$

where "/" represents the set subtraction operation. Since it is a convex function defined over a convex set, we can use the Lagrange multiplier technique and the Kuhn–Tucker theorem [1, Thm. A.6] to obtain its global minimizer. To be specific, let

$$f\left(P_{Y^n(\mathcal{S}_n)}\right)\stackrel{\triangle}{=}\sum_{x^n\in\mathcal{X}^n} P_{Y^n}\left[\mathcal{S}_n/\mathcal{B}_{na}(x^n)\right]^M+\lambda\left(\sum_{y^n\in\mathcal{S}_n} P_{Y^n}(y^n)-1\right).$$

Then

$$
\begin{aligned}
\frac{\partial f\left(P_{Y^n(\mathcal{S}_n)}\right)}{\partial P_{Y^n}(y^n)}&=\sum_{x^n\in\mathcal{X}^n} M\cdot P_{Y^n}\left[\mathcal{S}_n/\mathcal{B}_{na}(x^n)\right]^{M-1}\cdot\mathbf{1}\left\{y^n\in\mathcal{S}_n/\mathcal{B}_{na}(x^n)\right\}+\lambda\\
&=M\cdot\sum_{x^n\in\mathcal{X}^n} P_{Y^n}\left[\mathcal{S}_n/\mathcal{B}_{na}(x^n)\right]^{M-1}\cdot\mathbf{1}\left\{x^n\notin\mathcal{B}_{na}(y^n)\right\}+\lambda \;=\; 0, \qquad (12)
\end{aligned}
$$

where (12) follows from the symmetry of the distance measure, and $\mathbf{1}(\cdot)$ is the set indicator function. Taking uniform P_{Y^n} on $\mathcal{S}_n$ into the above equation, we obtain

$$\sum_{x^n\in\mathcal{X}^n}\left(1-\frac{|\mathcal{S}_n\cap\mathcal{B}_{na}(x^n)|}{|\mathcal{S}_n|}\right)^{M-1}\cdot\mathbf{1}\left\{x^n\notin\mathcal{B}_{na}(y^n)\right\}=-\frac{\lambda}{M}. \tag{13}$$

Therefore, if the left-hand side of (13) is independent of $y^n\in\mathcal{S}_n$, then uniform Y^n over $\mathcal{S}_n$ is indeed a solution of (12) (and minimizes $\bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R)$).

We conclude the above discussions in the next corollary.

COROLLARY 3.3. *Assume that* $|\mathcal{X}| < \infty$ *and the (generalized) distance measure* $\mu(\cdot,\cdot)$ *is symmetric. If, for every* n*,*

$$b_r(y^n) = b_r(z^n) \tag{14}$$

holds for every r *and every* y^n, z^n *in* $\mathcal{S}_n$*, then uniform* $\boldsymbol{Y}$ *over* $\boldsymbol{S}$ *minimizes* $\bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R)$*, where*

$$b_r(y^n) \stackrel{\triangle}{=} \sum_{x^n \in \mathcal{X}^n} \left(1 - \frac{|\mathcal{S}_n \cap \mathcal{B}_r(x^n)|}{|\mathcal{S}_n|}\right)^{M-1} \cdot \mathbf{1}\left\{x^n \notin \mathcal{B}_r(y^n)\right\}.$$

The condition in (14) may not hold in general. A quick example is to take the codeword set $\mathcal{S}_3 = \{000, 001, 011, 111\}$ under the symmetric Hamming distance metric and binary code alphabet. In such a case, $b_1(000) = b_1(111) = 2 \neq b_1(001) = b_1(011) = 2.5$ for $M = 2$. As a result, the best codeword-selecting distribution that minimizes (11) is $P_{Y^3}(000) = P_{Y^3}(111) = 1/2$ and $P_{Y^3}(001) = P_{Y^3}(011) = 0$, which is uniformly distributed only over a proper subset of $\mathcal{S}_3$.

Two queries can be further studied: whether it suffices to always take $\boldsymbol{Y}$ to be a uniform distribution over a subset of $\boldsymbol{S}$ as hinted by the previous example and whether sufficient condition (14) is also *necessary*. The proof of Theorem 2.4 indicates that the answer to the first query is affirmative; so to speak, taking $\boldsymbol{Y}$ to be the one chosen in the proof of Theorem 2.4, which is uniformly distributed over $\mathcal{C}_n^{(m)}$ for $\max_{1<i\leq m} N_m \leq n < \max_{1<i\leq(m+1)} N_m$, and following the proof of Lemma 2.2, we obtain

$$\bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R+\gamma) - \varepsilon \leq \sup_{0\leq\alpha<1} \limsup_{n\to\infty} \frac{1}{n}\rho_\alpha(M, \mathcal{S}_n \| X^n) \leq \bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R) + \varepsilon$$

for every $\gamma > 0$ and arbitrary $\varepsilon > 0$. By noting that $\bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R) = \lim_{\gamma\downarrow 0} \bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R+\gamma)$, except for countably many points in R, the first query is answered. We, however, have no answer to the second query. From several example trials, it seems affirmative as well. Nevertheless, equipped with the corollary, we can determine the asymptotic minimum covering radius for the examples in the next section.

4. Asymptotic minimum covering radius for specific block coding schemes. In this section, we demonstrate the usage of the new formula to investigate the asymptotic minimum covering radius in terms of two examples: *arbitrary block codes under* J*-ary quantized channels* and *constant weight codes under Hamming distance*.

4.1. Arbitrary block codes under J-ary quantized channels. We consider a model that is frequently used in practical channels (especially when the soft-decision decoding scheme is performed [4]).

Assume that a binary block code is transmitted over a memoryless channel whose output takes values from $\mathcal{N}_J \stackrel{\triangle}{=} \{0, 1, \ldots, J-1\}$; i.e., the output of the channel is quantized to J levels. The distance measure for quantized channels is defined as

$$\mu_n(x^n, y^n) = \sum_{i=0}^{n-1} |x_i - y_i|,$$

where x^n and y^n are in $\mathcal{N}_J^n$. The codeword set and the entire space are, respectively, $\mathcal{S}_n = \{0, (J-1)\}^n$ and $\mathcal{X}^n = \mathcal{N}_J^n$.

To derive the asymptotic minimum covering radius for this channel, we need to first show that $b_r(y^n)$ is independent of $y^n \in \mathcal{S}_n$ (and, therefore, uniform Y^n over $\mathcal{S}_n$ for each n minimizes $\bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R)$). We justify this claim as follows.

Observe that for any $y^n \in \mathcal{S}_n$,

(i) $x^n \in \mathcal{B}_r(\mathbf{0})$ iff $z^n \in \mathcal{B}_r(y^n)$, where $\mathbf{0}$ represents the all-zero element, and $z_i = |y_i - x_i|$ for $0 \le i \le n-1$;

(ii) furthermore, for any element $x^n \in \mathcal{X}^n$, $u^n \in \mathcal{S}_n \cap \mathcal{B}_r(x^n)$ iff $v^n \in \mathcal{S}_n \cap \mathcal{B}_r(z^n)$, where z^n is defined the same as above, and

$$v_i \stackrel{\triangle}{=} \begin{cases} (J-1) - u_i & \text{if } x_i \neq z_i, \\ u_i & \text{otherwise} \end{cases}$$

for $0 \le i \le n-1$.

Thus,

$$\begin{aligned} b_r(\mathbf{0}) &= \sum_{x^n \in \mathcal{X}^n} \left(1 - \frac{|\mathcal{S}_n \cap \mathcal{B}_r(x^n)|}{|\mathcal{S}_n|}\right)^{M-1} \cdot \mathbf{1}\{x^n \notin \mathcal{B}_r(\mathbf{0})\} \\ &= \sum_{z^n \in \mathcal{X}^n} \left(1 - \frac{|\mathcal{S}_n \cap \mathcal{B}_r(z^n)|}{|\mathcal{S}_n|}\right)^{M-1} \cdot \mathbf{1}\{z^n \notin \mathcal{B}_r(y^n)\} = b_r(y^n) \end{aligned}$$

for all $y^n \in \mathcal{S}_n$.

Now, for uniform Y^n over $\mathcal{S}_n = \{0, J-1\}^n$ (and also uniform X^n over $\mathcal{X}^n = \mathcal{N}_J^n$),

$$\begin{aligned} &\bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R) \\ &= \inf\left\{a \in \Re : \limsup_{n\to\infty} \frac{1}{J^n} \sum_{x^n \in \mathcal{X}^n} P_{Y^n}\left[y^n \in \mathcal{S}_n \ : \ y^n \notin \mathcal{B}_{na}(x^n)\right]^M = 0\right\} \\ &= \inf\left\{a \in \Re : \limsup_{n\to\infty} \sum_{x^n \in \mathcal{X}^n} \frac{1}{J^n}\left[1 - \frac{|\mathcal{S}_n \cap \mathcal{B}_{na}(x^n)|}{2^n}\right]^M = 0\right\} \\ &= \inf\left\{a \in \Re : \limsup_{n\to\infty} \sum_{\substack{n_0+n_1+\cdots \\ +n_{J-1}=n}} \frac{\frac{n!}{n_0!n_1!\cdots n_{J-1}!}}{J^n}\left[1 - \frac{f_{na}(n_0, n_1, \ldots, n_{J-1})}{2^n}\right]^M = 0\right\}, \end{aligned}$$

where n_i is the number of occurrence i's in x^n, and $f_{na}(n_0, n_1, \ldots, n_{J-1})$ is the summation of all $\binom{n_0}{i_0}\binom{n_1}{i_1}\cdots\binom{n_{J-1}}{i_{J-1}}$ satisfying $0 \le i_j \le n_j$ for $0 \le j \le J-1$ and $\sum_{j=0}^{J-1}[i_j j + (n_j - i_j)(J-j-1)] \le na$. By using typical asymptotic approximation for binomial coefficients, we obtain that for $v_0 + v_1 + \cdots + v_{J-1} = 1$,

$$\begin{aligned} &g(v_0, v_1, \ldots, v_{J-1}, a) \\ &\stackrel{\triangle}{=} \lim_{n\to\infty} \frac{1}{n} \log_2 \frac{2^n}{f_{na}(nv_0, nv_1, \ldots, nv_{J-1})} \end{aligned}$$

$$= \begin{cases} \infty & \text{if } a < \sum_{j=0}^{J-1} v_j \cdot \min\{j, (J-1)-j\}, \\ 1 - \max\limits_{(\delta_0,\ldots,\delta_{J-1})\in\mathcal{D}_J} \left[v_0 H(\delta_0) + \cdots + v_{J-1} H(\delta_{J-1})\right] & \\ \qquad\qquad \text{if } \sum_{j=0}^{J-1} v_j \cdot \min\{j, (J-1)-j\} \le a \le \dfrac{J-1}{2}, & \\ 0 & \text{if } a > \dfrac{J-1}{2}, \end{cases}$$

where $H(x) \stackrel{\triangle}{=} -x\log_2(x) - (1-x)\log_2(1-x)$ is the binary entropy function, $\mathcal{D}_J$ consists of all $(\delta_0, \ldots, \delta_{J-1})$ satisfying $0 \le \delta_j \le 1$ for $0 \le j \le J-1$ and $\sum_{j=0}^{J-1}[v_j\delta_j j + v_j(1-\delta_j)(J-j-1)] \le a$, and the result for $a > (J-1)/2$ follows by taking $\delta_1 = \delta_2 = \cdots = \delta_{J-1} = 1/2$. Thus,

$$\sup_{0\le\alpha<1} \lim_{n\to\infty} \frac{1}{n}\rho_\alpha\left(M = e^{nR}, \{0, J-1\}^n\right) = \inf\left\{a \in \Re : \frac{R}{\log(2)} > \max_{v_0+\cdots+v_{J-1}=1} g(v_0, \ldots, v_{J-1}, a)\right\}. \tag{15}$$

We can then derive the asymptotic minimum covering radius for different J values based on (15).

Case A. For J odd,

$$\max_{v_0+\cdots+v_{J-1}=1} g(v_0, \ldots, v_{J-1}, a) = \begin{cases} \infty & \text{if } a < \frac{J-1}{2}, \\ 0 & \text{if } a \ge \frac{J-1}{2}, \end{cases}$$

and, therefore,

$$\sup_{0\le\alpha<1} \lim_{n\to\infty} \frac{1}{n}\rho_\alpha\left(M = e^{nR}, \{0, J-1\}^n\right) = \frac{J-1}{2} \quad \text{for } 0 < R \le \log(2).$$

It can easily be seen that $(J-1)/2$ is the trivial lower bound for the asymptotic minimum covering radius at J odd, since any codeword in $\{0, J-1\}^n$ require radius $(J-1)n/2$ to cover the all-$[(J-1)/2]$ element. Here, in lieu of the new formula, we show that $(J-1)/2$ is actually the asymptotic minimum covering radius at J odd.

Case B. J is even.

Since the case of $J = 2$ reduces to a simple *binary block code* under *Hamming distance*, for which the derivation of its asymptotic minimum covering radius can be easily computed through combinatorial approaches, we will therefore focus on the case of $J \ge 4$.

To derive the asymptotic minimum covering radius, we first establish a general lower bound by using the middle point $x^n_{\text{mid}} \stackrel{\triangle}{=} (J/2-1, J/2-1, \ldots, J/2-1)$ as follows:

$$\bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R) = \inf\left\{a \in \Re : \limsup_{n\to\infty} \frac{1}{J^n} \sum_{\substack{n_0+n_1+\cdots \\ +n_{J-1}=n}} \frac{n!}{n_0! n_1! \cdots n_{J-1}!} \left[1 - \frac{|\mathcal{S}_n \cap \mathcal{B}_{na}(x^n)|}{2^n}\right]^M = 0\right\}$$

$$\geq \inf\left\{a \in \Re : \limsup_{n\to\infty} \frac{1}{J^n}\left[1 - \frac{|\mathcal{S}_n \cap \mathcal{B}_{na}(x^n_{\text{mid}})|}{2^n}\right]^M = 0\right\}$$
$$= \inf\left\{a \in \Re \ : \ \frac{R}{\log(2)} > 1 - H\left(a - \left(\frac{J}{2} - 1\right)\right)\right\} \text{ for } 0 < R \leq \log(2).$$

We then observe that for $v_0 + v_1 + v_2 + v_3 = 1$ and $1 \leq a \leq 3/2$,

$$g(v_0, v_1, v_2, v_3, a) = 1 - \max_{(\delta_0,\ldots,\delta_3)\in\mathcal{D}_4} [v_0 H(\delta_0) + \cdots + v_3 H(\delta_3)] \leq 1 - H(a-1),$$

where the last step follows by taking $1 - \delta_0 = 1 - \delta_1 = \delta_2 = \delta_3 = a - 1$, which is in the range of the maximization operation. Thus, for $0 < R \leq \log(2)$,

$$\begin{aligned}
&\sup_{0\leq\alpha<1} \lim_{n\to\infty} \frac{1}{n}\rho_\alpha\left(M = e^{nR}, \{0,3\}^n\right) \\
&= \inf\left\{a \in \Re : \frac{R}{\log(2)} > \max_{v_0+v_1+v_2+v_3=1} g(v_0, v_1, v_2, v_3, a)\right\} \\
&\leq \inf\left\{a \in \left[1, \frac{3}{2}\right] : \frac{R}{\log(2)} > \max_{v_0+v_1+v_2+v_3=1} g(v_0, v_1, v_2, v_3, a)\right\} \\
&\leq \inf\left\{a \in \left[1, \frac{3}{2}\right] : \frac{R}{\log(2)} > 1 - H(a-1)\right\}.
\end{aligned}$$

As a result, the general lower bound is tight at $J = 4$.

For $J \geq 6$ even, there seems no simple expression for the asymptotic minimum covering radius. However, one can still obtain a numerically plotted curve for the asymptotic minimum covering radius at $J \geq 6$ even, whenever the algorithmic complexity of the optimization operation for (15) is feasible.

4.2. Binary constant weight codes under Hamming distance. Define the codeword set as

$$\mathcal{S}_n(w) \ = \ \{y^n \in \{0,1\}^n \ : \ W(y^n) = w\},$$

where $W(y^n)$ is the number of 1's in y^n. The covering space is assumed to be the entire space $\mathcal{X}^n = \{0,1\}^n$. Let the distance measure $\mu_n(\cdot,\cdot)$ be the n-fold Hamming distance.

In this case, the asymptotic minimum covering radius for codeword set $\mathcal{S}_n(nv)$ is apparently lower bounded by $\max\{v, 1-v\}$, since the code must cover both the all-zero element and the all-one element. Now, in lieu of the new formula, we can show that $\max\{v, 1-v\}$ is indeed the exact asymptotic minimum covering radius for constant weight codes.

Define $f_r(w,\eta) \stackrel{\triangle}{=} |\mathcal{S}_n(w) \cap \mathcal{B}_r(x^n)|$, where $\eta = W(x^n)$. Then, for $y^n \in \mathcal{S}_n(w)$,

$$\begin{aligned}
b_r(y^n) &= \sum_{\eta=0}^{n} \left(1 - \frac{f_r(w,\eta)}{\binom{n}{w}}\right)^{M-1} \left[\sum_{\{x^n \ : \ W(x^n)=\eta\}} \mathbf{1}\left\{x^n \notin \mathcal{B}_r(y^n)\right\}\right] \\
&= \sum_{\eta=0}^{n} \left(1 - \frac{f_r(w,\eta)}{\binom{n}{w}}\right)^{M-1} \left(|\mathcal{S}_n(\eta)| - |\mathcal{S}_n(\eta) \cap \mathcal{B}_r(y^n)|\right) \\
&= \sum_{\eta=0}^{n} \left(1 - \frac{f_r(w,\eta)}{\binom{n}{w}}\right)^{M-1} \left[\binom{n}{\eta} - f_r(\eta, w)\right],
\end{aligned}$$

which is apparently independent of $y^n \in \mathcal{S}_n(w)$ for every r. Hence, uniform Y^n over $\mathcal{S}_n(w)$ for each n minimizes $\bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R)$.

Now, from the observations that for fixed x^n with $W(x^n) = \eta$ the total number of y^n in $\mathcal{S}_n(w)$ satisfying that the weights (1's) of x^n and y^n coincide with each other in exactly d positions is equal to $\binom{\eta}{d}\binom{n-\eta}{w-d}$, and that $W(x^n) + W(y^n) - 2d$ is the Hamming distance between x^n and y^n with d coincidences in their weights, we get[5]

$$\begin{aligned} f_r(w,\eta) &= \sum_{\left\{d\,:\ \substack{0\le d\le \min\{w,\eta\} \\ 0\le w-d\le n-\eta,\ 0\le w+\eta-2d\le r}\right\}} \binom{\eta}{d}\binom{n-\eta}{w-d} \\ &= \sum_{\left\{i\,:\ \substack{0\le (w+\eta-i)/2\le \min\{w,\eta\} \\ 0\le (w-\eta+i)/2\le n-\eta,\ 0\le i\le r}\right\}} \binom{\eta}{\frac{w+\eta-i}{2}}\binom{n-\eta}{\frac{w-\eta+i}{2}} \times \mathbf{1}\left\{(w+\eta-i) \text{ even}\right\} \\ &= \sum_{i=|w-\eta|}^{\min\{r,\ w+\eta,\ 2n-w-\eta\}} \binom{\eta}{\frac{\eta-w+i}{2}}\binom{n-\eta}{\frac{w-\eta+i}{2}} \times \mathbf{1}\left\{(w+\eta-i) \text{ even}\right\}. \end{aligned} \tag{16}$$

Accordingly, for uniform $\boldsymbol{Y}$ over $\boldsymbol{S}(w)$ (and also uniform $\boldsymbol{X}$ over the entire space),

$$\begin{aligned} \bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R) &= \inf\left\{a \in \Re : \limsup_{n\to\infty} \frac{1}{2^n} \sum_{x^n\in\mathcal{X}^n} P_{Y^n}\left[y^n \in \mathcal{S}_n : y^n \notin \mathcal{B}_{na}(x^n)\right]^M = 0\right\} \\ &= \inf\left\{a \in \Re : \limsup_{n\to\infty} \frac{1}{2^n} \sum_{\eta=0}^{n} \binom{n}{\eta}\left[1 - \frac{f_{na}(w,\eta)}{\binom{n}{w}}\right]^M = 0\right\}. \end{aligned}$$

By using typical asymptotic approximation for binomial coefficients, we obtain[6]

$$\begin{aligned} \bar{g}(v,\hat{v},a) &\stackrel{\triangle}{=} \lim_{n\to\infty} \frac{1}{n}\log_2 \frac{\binom{n}{nv}}{f_{na}(nv,n\hat{v})} \\ &= \begin{cases} H(v) - \max\limits_{|v-\hat{v}|\le j\le \min\{a,v+\hat{v},2-(v+\hat{v})\}} \left[\hat{v}H\left(\frac{\hat{v}-v+j}{2\hat{v}}\right) + (1-\hat{v})H\left(\frac{v-\hat{v}+j}{2(1-\hat{v})}\right)\right] \\ \qquad\qquad \text{if } |v-\hat{v}| \le a \le 1, \\ \infty \qquad \text{if } 0 \le a < |v-\hat{v}| \end{cases} \\ &= \begin{cases} H(v) - [\hat{v}\cdot H(1-v) + (1-\hat{v})\cdot H(v)] & \text{if } v+\hat{v}-2v\hat{v} \le a \le 1, \\ H(v) - \left[\hat{v}\cdot H\left(\frac{\hat{v}-v+a}{2\hat{v}}\right) + (1-\hat{v})\cdot H\left(\frac{v-\hat{v}+a}{2(1-\hat{v})}\right)\right] & \\ \qquad\qquad \text{if } |v-\hat{v}| \le a < v+\hat{v}-2v\hat{v}, & \\ \infty \qquad \text{if } 0 \le a < |v-\hat{v}| & \end{cases} \\ &= \begin{cases} 0 & \text{if } v+\hat{v}-2v\hat{v} \le a \le 1, \\ H(v) - \left[\hat{v}\cdot H\left(\frac{\hat{v}-v+a}{2\hat{v}}\right) + (1-\hat{v})\cdot H\left(\frac{v-\hat{v}+a}{2(1-\hat{v})}\right)\right] & \\ \qquad\qquad \text{if } |v-\hat{v}| \le a < v+\hat{v}-2v\hat{v}, & \\ \infty \qquad \text{if } 0 \le a < |v-\hat{v}|. & \end{cases} \end{aligned}$$

[5] By definition, d is the number of coincidences in the weights of x^n and y^n, and hence $0 \le d \le \min\{w,\eta\}$.

[6] The function $\hat{v}\cdot H(\frac{\hat{v}-v+j}{2\hat{v}}) + (1-\hat{v})H(\frac{v-\hat{v}+j}{2(1-\hat{v})})$ is a concave truncated function of j, and is maximized at $j = v+\hat{v}-2v\hat{v}$, if $|v-\hat{v}| \le v+\hat{v}-2v\hat{v} \le \min\{a,\, v+\hat{v},\, 2-(v+\hat{v})\}$. Note that $|v-\hat{v}| \le v+\hat{v}-2v\hat{v} \le \min\{v+\hat{v},\, 2-(v+\hat{v})\}$ is valid for every $0 \le v, \hat{v} \le 1$.

Thus, for $0 \le v \le 1/2$ and $0 < R \le \log(2)$,

$$\begin{aligned}\sup_{0\le\alpha<1} \lim_{n\to\infty} \frac{1}{n}\rho_\alpha\left(M = e^{nR}, \mathcal{S}_n(nv)\right) &= \bar{\Omega}_{\boldsymbol{Y}\|\boldsymbol{X}}(R) \\ &= \inf\left\{a \in \Re \ : \ \frac{R}{\log(2)} > \max_{0\le\hat{v}\le 1} g(v,\hat{v},a)\right\} \\ &= 1 - v,\end{aligned}$$

where the last step follows from[7]

$$\max_{0\le\hat{v}\le 1} g(v,\hat{v},a) = \begin{cases} \infty & \text{if } 0 \le a < \max\{v, 1-v\} = 1-v, \\ 0 & \text{if } \max\{v,1-v\} \le a \le 1. \end{cases}$$

Similarly, for $1/2 < v \le 1$,

$$\sup_{0\le\alpha<1} \lim_{n\to\infty} \frac{1}{n}\rho_\alpha\left(M = e^{nR}, \mathcal{S}_n(nv)\right) = v.$$

Acknowledgment. The authors wish to thank the anonymous reviewers for their valuable suggestions and comments that greatly helped to improve the paper.

REFERENCES

[1] R. E. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley, Reading, MA, 1987.

[2] V. M. Blinovskii, *Lower asymptotic bound on the number of linear code words in a sphere of given radius in f_q^n*, Problemy Peredachi Informatsii, 23 (1987), pp. 50–53.

[3] 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, 46 (2000), pp. 869–885.

[4] G. C. Clark, Jr. and J. B. Cain, *Error-Correction Coding for Digital Communications*, Plenum Press, New York, 1981.

[5] G. D. Cohen, *A nonconstructive upper bound on covering radius*, IEEE Trans. Inform. Theory, 29 (1983), pp. 352–353.

[6] G. D. Cohen and P. Frankl, *Good coverings of Hamming spaces with spheres*, Discrete Math., 56 (1985), pp. 125–131.

[7] G. D. Cohen, I. Honkala, S. Litsyn, and A. Lobstein, *Covering Codes*, North-Holland, Amsterdam, 1997.

[8] G. D. Cohen, M. G. Karpovsky, H. F. Mattson, Jr., and J. R. Schatz, *Covering radius–survey and recent results*, IEEE Trans. Inform. Theory, 31 (1985), pp. 328–343.

[9] G. D. Cohen, S. N. Litsyn, and G. Zémor, *On greedy algorithms in coding theory*, IEEE Trans. Inform. Theory, 42 (1996), pp. 2053–2057.

[10] G. D. Cohen, S. N. Litsyn, A. C. Lobstein, and H. F. Mattson, Jr., *Covering radius* 1985–1994, Appl. Engrg. Comm. Comput., 8 (1997), pp. 173–239.

[11] P. Delsarte and P. Piret, *Do most binary linear codes achieve the Goblick bound on the covering radius?*, IEEE Trans. Inform. Theory, 32 (1986), pp. 826–828.

[12] R. L. Graham and N. J. A. Sloane, *On the covering radius of codes*, IEEE Trans. Inform. Theory, 31 (1985), pp. 385–401.

[13] T. S. Han, *Information-Spectrum Methods in Information Theory*, Baifukan Press, Tokyo, 1998 (in Japanese).

[14] T. Helleseth, T. Kløve, and J. Mykkeltveit, *On the covering radius of binary codes*, IEEE Trans. Inform. Theory, 24 (1978), pp. 627–628.

[7] There exists no $\hat{v} \in [0,1]$ satisfying $|v - \hat{v}| \le a < v + \hat{v} - 2v\hat{v}$, $1 - v \le a \le 1$, and $0 \le v \le 1/2$. This observation can easily be justified by

$$(1-2v) \ge (1-2v)\hat{v} > a - v \ge (1-v) - v = 1 - 2v \quad \text{for } 0 \le \hat{v} \le 1.$$

[15] H. Janwa, *Some new upper bounds on the covering radius of binary linear codes*, IEEE Trans. Inform. Theory, 35 (1989), pp. 110–122.
[16] H. L. Royden, *Real Analysis*, 3rd ed., Macmillan, New York, 1988.
[17] P. Solé, *Asymptotic bounds on the covering radius of binary codes*, IEEE Trans. Inform. Theory, 36 (1990), pp. 1470–1472.
[18] P. Solé, *Packing radius, covering radius, and dual distance*, IEEE Trans. Inform. Theory, 41 (1995), pp. 268–272.
[19] A. A. Tietäväinen, *An asymptotic bound on the covering radii of binary BCH codes*, IEEE Trans. Inform. Theory, 36 (1990), pp. 211–213.
[20] A. A. Tietäväinen, *An upper bound on the covering radius as a function of the dual distance*, IEEE Trans. Inform. Theory, 36 (1990), pp. 1472–1474.
[21] F. Levy-dit-Vehel and S. Litsyn, *More on the covering radius of BCH codes*, IEEE Trans. Inform. Theory, 42 (1996), pp. 1023–1028.