Techniques of Bounding the Probability of Decoding Error for Block Coded Modulation Structures

Hanan Herzberg and Gregory Poltyrev

Abstract—Two techniques for upper bounding the average probability of decoding error in coded modulation structures are presented. The first bound, which is applicable to the additive white Gaussian noise (AWGN) channel, is tighter than the well-known union bound and the minimum distance bound, especially for low signal to noise ratio. It is shown that for the Leech lattice this upper bound is very close to a sphere lower bound. For the second upper bound, which is applicable to any memoryless channel (not necessarily AWGN), a method of random coset coding is presented. For the AWGN channel, a tighter upper bound is obtained by employing the method of random coset coding for calculating the average spectrum of distances of the code, which is required for the computation of the first upper bound.

Index Terms—Coded modulation, probability of decoding error, spectrum of distances, cosets of code, memoryless channel, AWGN channel.

I. INTRODUCTION

Coded modulation is an efficient way of combining error correction coding with modulation (e.g., [1]–[6]). The design of a conventional coded modulation scheme is aimed at achieving the largest squared Euclidean distance between the closest transmitted signal sequences. The fact that large squared Euclidean distance leads to asymptotic (i.e., for signal to noise ratio approaching infinity) coding gain, was the motivation for designing many coded modulation schemes (e.g., [1]). For moderate and, especially, very low signal to noise ratios the coding gain is reduced. This reduction is explained mainly by the error coefficient, namely, the number of coded sequences at minimum distance apart of an average transmitted sequence [2]. Note that for low signal to noise ratio, coded sequences beyond the minimum distance may also influence the coding gain. The asymptotic coding gain, as well as the effective coding gain [2], are useful only when the desired probability of error is rather small. However, the coded modulation structure can be implemented as an inner code concatenated with Reed–Solomon outer codes [7], [8]. For the calculation of the coding gain, the desired probability of decoding error of the inner code is in the range 0.1–0.01. Furthermore, the result of this calculation is very sensitive to changes in the probability of decoding error of the inner code. Therefore, for low signal to noise ratio, a tighter upper bound for the probability of decoding error ought to be computed.

An upper bound, which is tighter than the union bound, is presented in [9] for the conditional probability of decoding error. Although the derivation of the sphere upper bound, which is presented in Section II, is guided by a different concept, the two upper bounds coincide for the conditional probability. However, the sphere upper bound presented here is easily calculated and it also allows estimation of the average probability of error for the important case of a finite size constellation. We provide a lower bound for the probability of error. As illustrated in Example 1, the upper and lower bounds are surprisingly close for the Leech lattice.

The general coded modulation structure can be described in the following way. Let \( C \) be a code with length \( n \) and \( M \) codewords over the alphabet \( X = \{0, 1, \ldots, q - 1\} \). Let \( Y = \{y_1, y_2, \ldots, y_M\} \) be a set of channel symbols and let \( \phi(x) \) be a one to one mapping (i.e., modulation) from \( X \) into the signal set \( Y \). Whenever the encoder produces a codeword \( \Phi \in C \), the transmitted sequence is \( \phi(\Phi) \), and the codeword \( \Phi \) is mapped into the corresponding sequence of channel symbols.

The probability of decoding error for transmission of the codeword \( \Phi \in C \), denoted by \( \lambda \), is usually estimated by the following union bound \( \lambda \leq \sum_{i,j} A_{i,j} \), where \( A_{i,j} \) is the probability of decoding error under the assumption that the only code words are \( C_i, \Phi \in C \). For an additive white Gaussian noise (AWGN) channel the probability \( A_{i,j} \) depends only on the Euclidean distance between \( \phi(\Phi) \) and \( \phi(\Phi') \). Therefore, the union...
bound assumes the form $A_{\delta} = \sum_{j=1}^{N} A_{j} \lambda(\delta_{j})$, where $\{\delta_{j}\}$, $j = 1, 2, \ldots, N$ is the set of Euclidean distances between all pairs of transmitted codewords, $\lambda(\delta_{j})$ is the probability $A_{\delta_{j}}$, while the distance between $\varphi(\delta_{j})$ and $\varphi(\delta_{j})$ is $\delta_{j}$, and $A_{\delta_{j}}$ is the number of codewords at distance $\delta_{j}$ from $\varphi_{j} \in \mathbb{C}$ (namely the spectrum of distances of the code). For applying this bound it is necessary to calculate the coefficients $A_{\delta_{j}}$. In many cases this set of coefficients is not the same for all the codewords in $C$, consequently the probability of error is also not the same for all the codewords. Therefore, the average probability of error is bounded by $P_{e} \leq \sum_{r}^{N} A_{\delta} \lambda(\delta_{r})$, where $A_{\delta} = 1/\sum_{j=1}^{N} A_{j} \lambda(\delta_{j})$, $j = 1, 2, \ldots, N$, is the average coefficient, namely, the average spectrum of distances of the code. This spectrum (as well as the set $\{\delta_{j}\}$) depends on both the code $C$ and the modulator $\varphi(x)$. For many structures, especially when the signal constellation has finite size, the calculation of the average coefficients and even of the dominant coefficient (conventionally named the error coefficient) is rather complicated. Furthermore, the union bound is not tight enough for low signal to noise ratios.

An upper bound, named the sphere upper bound, is derived in Section II for the AWGN channel. This upper bound is applicable for the conditional probability of error as well as for the average probability of error in any, finite or infinite, constellation. This bound is tighter than the union bound, especially for low signal to noise ratios, and it is easy to calculate. It is also shown that for the Leech lattice the sphere upper bound is very close to a sphere lower bound. In Section III we will consider, for any given group code $C$, the ensemble of the codes generated by random translates of this code (the ensemble of coset codes). This will enable us to calculate a union bound for the probability of decoding error of an average coset code. The bound, which we derive in Section III, is not limited to a particular channel or coded modulation scheme and is applicable for any memoryless channel (not necessarily an additive noise channel). Furthermore, the method of random coset coding for transmitting the data yields an equal probability of error for each transmitted sequence. The method of calculating the average spectrum of distances of the code is also presented in Section III. This spectrum is used for the computation of the upper bound of Section II. An example exploiting the advantage of the resulting bound is presented in Section IV, where we are concerned with multilevel coding. Conclusions are presented in Section V.

II. SPHERE UPPER AND LOWER BOUNDS ON THE PROBABILITY OF ERROR

A. Sphere Upper Bound

Consider a code with fixed length $n$, $M$ codewords, a set of Euclidean distances $\{\delta_{j}\}$, $j = 1, 2, \ldots, N$, and a set of coefficients $\{A_{j}\}$, $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, N$. Let the additive noise at the input to the decoder be an $n$-dimensional random vector denoted by $z = (z_{1}, z_{2}, \ldots, z_{n})$, and let the event of error at the output of the decoder be denoted by $E$. The probability of decoding error can be expressed as

$$P_{e} = \Pr \left( \|z\| \leq r \right) \Pr \left( \|z\| > r \right) + \Pr \left( \|z\| > r \right) \Pr \left( \|z\| > r \right).$$

where $\|\cdot\|$ stands for the Euclidean norm and $r$ is a real positive parameter (the radius of a sphere) which is left here arbitrary but will be determined later in order to obtain a tight bound.

$$\Pr \left( E \mid \|z\| > r \right) \leq \mathcal{H}.$$  

Hence

$$P_{e} \leq \min \left\{ \sum_{j=1}^{N} A_{\delta_{j}} \lambda(\delta_{r}) \right\} + \Pr \left( \|z\| > r \right).$$

In this section we are concerned with the case of transmitting through an AWGN channel. Since the probability of decoding error is determined by the Euclidean distances between the transmitted codewords, the following union bound is true:

$$\Pr \left( E \mid \|z\| \leq r \right) \leq \sum_{j=1}^{N} A_{\delta_{j}} \Pr \left( E \mid \|z\| \leq r \right),$$

where $E_{j}$ stands for the event of error at the output of the codeword, while the decoded codeword is at a Euclidean distance $\delta_{j}$ from the transmitted codeword, and $A_{\delta_{j}}$ is the average number of codewords at a distance $\delta_{j}$ from the transmitted codeword. Since $\Pr \left( E_{j} \mid \|z\| \leq r \right) = 0$ for $r < \delta_{j}/2$, substitution of the resulting expression into (2) yields

$$P_{e} \leq \min \left\{ \sum_{j=1}^{N} A_{\delta_{j}} \Pr \left( E_{j} \mid \|z\| \leq r \right) + \Pr \left( \|z\| > r \right) \right\}$$

where $N(r)+1$ is the smallest value of $j$ which satisfies $r \leq \delta_{j}/2$. Let

$$P_{e}(r) = \sum_{j=1}^{N(r)} A_{\delta_{j}} \Pr \left( E_{j} \mid \|z\| \leq r \right) + \Pr \left( \|z\| > r \right).$$

It is obvious that $r$ may have a value in the range $\delta_{j}/2 \leq r \leq \infty$, where $\delta_{j}$ is the minimum Euclidean distance between the transmitted codewords. By substituting $r = \delta_{j}/2$ into $P_{e}(r)$, since $N(\delta_{j}/2) = 0$, the well-known minimum distance bound $P_{e}(r = \delta_{j}/2) = \Pr \left( \|z\| \geq \delta_{j}/2 \right)$ (see, e.g., [10]) is obtained. By considering the limit as $r$ approaches infinity, we obtain the well-known union bound (see, e.g., [10]) $P_{e}(r \rightarrow \infty) \leq \sum_{j=1}^{N} A_{\delta_{j}} \Pr \left( E_{j} \right)$. Clearly, the suggested upper bound (4) is tighter than these two bounds.

The geometric location of the intersection $E_{j} \cap \{z : \|z\| \leq r\}$ for a two-dimensional code is the crosshatched region shown in Fig. 1. Let the random variable $y$ be the squared norm $\|z\|^{2} = \sum_{i=1}^{n} z_{i}^{2}$. For an AWGN channel the random variables $z_{i}, i = 1, 2, \ldots, n$, are independent normally distributed with the same variance $\sigma^{2}$ and zero mean. Therefore, the random variable $y$ has the chi-squared distribution (see, e.g., [11]). The joint probability density function of $z_{i}$ and $y$ is given by

$$f(z_{i}, y) = \frac{1}{\sqrt{\pi \sigma^{2}/2} y^{\nu/2} \Gamma(\nu/2)/2} \left( y - z_{i}^{2} \right)^{(\nu-3)/2}$$

where

$$U(y) = \begin{cases} 1, & y \geq 0, \\ 0, & y < 0, \end{cases}$$

and $\Gamma(\nu) = \int_{0}^{\infty} y^{\nu-1} e^{-y} dy$ is the gamma function.

Due to the spherical symmetry of the AWGN we can assume that both the decoded codeword and the transmitted codeword lie on the first coordinate. Thus $\Pr \left( E_{j} \mid \|z\| \leq r \right) = \Pr \left( z_{1} \geq \delta_{j}/2, y \geq r^{2} \right)$. Finally, the probability of decoding error can be upper bounded as follows:

$$P_{e} \leq \min \left\{ \sum_{j=1}^{N(r)} A_{\delta_{j}} \int_{y_{r}/2}^{\infty} f(z_{1}, y) dy + \int_{r^{2}/2}^{\infty} f(y) dy \right\},$$

where $f(y)$ is the density function of the chi-squared distribution.
The optimal value of \( r \), denoted by \( r_o \), is such that the right-hand term in (4) is minimized. Thus it satisfies

\[
\frac{dP_e(r)}{dr} \bigg|_{r=r_o} = 0.
\]

By (7), \( P_e(r) \) can be written as

\[
P_e(r) = \sum_{j=1}^{N(r)} A_j \int_{\delta_j/2}^{\delta_j} \left( y - \frac{z_j^2}{4} \right)^{(n-3)/2} e^{-y/2} \pi^{n/2} \Gamma((n-1)/2) \frac{d\delta_j}{\delta_j} dy + \int_{\delta_j/2}^{\infty} f(y) dy. \tag{8}
\]

Straightforward computation yields \( r_o \), which is the root of the equation

\[
\sum_{j=1}^{N(r)} A_j \int_{\delta_j/2}^{\delta_j} \left( 1 - \frac{x^2}{4} \right)^{(n-3)/2} dx = 1. \tag{9}
\]

From (9) we can deduce that \( r_o \) does not depend on \( \sigma^2 \) and depends only on the sets \( \{\delta_j\} \) and \( \{A_j\} \).

Substitution of \( x = \cos(u) \) and \( \delta_j/2r_o = \cos(\theta_j) \) into (9) yields

\[
\sum_{j=1}^{N(r)} A_j \int_{\theta_j}^{\delta_j/2} \Gamma((n-1)/2) \int_0^{\infty} \sin^{n-2}(u) du = 1. \tag{10}
\]

Let \( \Omega(\theta) \) be the surface area of the unit \( n \)-dimensional sphere, and let \( \Omega(\theta) \) be the surface area in \( n \)-space of a cone of half-angle \( \theta_j \), i.e., the surface area of the unit \( n \)-dimensional sphere cut out by the cone. The following expression for the surface area \( \Omega(\theta) \) can be found in [12]:

\[
\Omega(\theta) = \frac{(n-1)\pi^{(n-1)/2}}{\Gamma((n+1)/2)} \int_0^{\infty} \sin^{n-2}(u) du. \tag{11}
\]

Substitution of (11) into (10) and straightforward computation yield

\[
\sum_{j=1}^{N(r)} A_j \Omega(\theta_j) = \Omega. \tag{12}
\]

The geometric interpretation of (7) and (12) is illustrated in Fig. 2. Given that the transmitted signal is at the center of the sphere, if the received signal is outside the polyhedron, then this event is regarded according to (7) as an error. For lattice structures, \( A_{i,j} = A_j \) for all the codewords. In this case the polyhedron (the decision region for the computation of the bound), which is inscribed inside the sphere of radius \( r_o \), is the worst decision region (also called the Voronoi cell [13]) among all the lattices with the same spectrum of distances (i.e., theta series [13]). It can be shown that \( r_o \) is also a lower bound for the covering radius of any lattice with a given spectrum of distances.

The upper bound presented in ([9], eq. (2.31)), which is developed for the conditional probability of error, can also be interpreted as the probability that the received signal is outside the polyhedron plotted in Fig. 2, given that the transmitted signal is at the center. Therefore, this bound and the upper bound (7) coincide for the case when \( A_{i,j} = A_j \) (e.g., for an infinite lattice structure). The bound presented here is applicable for any infinite size constellation as well as for practical finite size constellations, as illustrated in Section IV.

In order to calculate the upper bound, (7) can be further simplified. The probability \( P_e(E_i, \|z\| \leq r) \) is tightly upper bounded as follows:

\[
P_e(E_i, \|z\| \leq r) \leq P_e(r > z_1 > \delta_j/2, y_1 < r^2 - \delta_j^2/4) \leq P_e(r > z_1 > \delta_j/2, y_1 < r^2 - \delta_j^2/4), \tag{13}
\]

where \( y_j = \sum_{i=1}^{n} z_i^2 \). The geometric location of the intersection \( \{z: r > z_1 > \delta_j/2 \} \cap \{z: y_1 < r^2 - \delta_j^2/4 \} \) for a two-dimensional code is the crosshatched region shown in Fig. 3, which contains the crosshatched region shown in Fig. 1. It is easy to see that the upper bound (13) becomes tight as \( n \) increases. By (4) and (13), the following upper bound is obtained:

\[
P_e \leq \min_r \left\{ \sum_{j=1}^{N(r)} A_j \left[ Q(\delta_j/2\sigma) - Q(r/\sigma) \right] \int_0^{r^2 - \delta_j^2/4} f(y_1) dy_1 \right. \]

\[
\left. + \int_{r^2 - \delta_j^2/4}^{\infty} f(y) dy \right\}, \tag{14}
\]

where \( Q(\cdot) \) is the complementary error function, defined by \( Q(x) = 1 - \sqrt{2\pi} \int_0^x e^{-t^2/2} dt \). For many practical structures the difference between (7) and (14) is negligible.

B. Sphere Lower Bound

Let \( V(\theta) \) be the volume of the decision region of the codeword \( \theta \). The probability of decoding error can be lower bounded by

\[
P_e \geq P_r(\|z\| > r_p), \tag{15}
\]

where \( r_p \) is the radius of an \( n \)-dimensional sphere whose volume
which satisfies
\[ r_o = \sqrt{\frac{\delta_1^2}{2}} \] equals that of the decision region of \( N(r_o) \).

**Fig. 3. Geometric location of the intersection \( \{ z : r > z, \gamma > \delta_1/2 \} \cap \{ z : \gamma < r^2 - \delta_1^2/4 \} \).**

This is a well-known Leech lattice (see, e.g., [3] and [13]), for which the probability of error is close to 1. Thus in most of the interesting range of the probability of decoding error, in which the bound (14) and the lower bound (15), which are very close to each other, are presented versus \( 6, \delta_1/2 \) in Fig. 4. The union bound, which is also presented in Fig. 4, is not tight for high levels of noise.

The parameters of some important lattices (see [2] and [13]) are listed in Table I. The radii \( r_o \) and \( r_p \) are calculated by (9) and (16), respectively. For low levels of noise (i.e., in cases where the union bound is tight enough), the first term in the upper bound (7) is dominant, whereas the second term in this upper bound is dominant for very high levels of noise (i.e., in cases where the probability of error is close to 1). Thus in most of the interesting range of the probability of decoding error, in which the bound (7) is especially important, these two terms are of the same order. The second term in (7) is simply the probability that the noise is outside a sphere of radius \( r_o \), given that the transmitted codeword is at the center of this sphere. On the other hand, the lower bound (15) is determined by the radius \( r_p \). Thus, the values of the ratio \( r_o/r_p \) indicate the tightness of the upper and lower bounds. Therefore, from Table I, one deduces that, since \( r_o \) and \( r_p \) are very close to each other for the listed lattices (and \( r_o \) even slightly exceeds \( r_p \)), the upper and lower bounds for these lattices are very close.

**III. RANDOM COSET CODING**

The set of coefficients \( \{ A_i \} \) is needed for applying the sphere upper bound (7) or (14). Calculation of this set of coefficients is rather complicated for many structures. The dominant coefficient \( A_1 \) of many important structures can be found in the literature only for a signal constellation with an infinite size (see, e.g., [2], [3], and [13]). Since the value of any coefficient \( A_1 \) in a signal constellation with a finite size is less than that of the infinite constellation, a further tightening of the upper bound can be achieved by calculating the set \( \{ A_i \} \) for the practical finite constellation. A random coset coding, which may also serve as a method for evaluating the set \( \{ A_i \} \) and calculating an upper bound for the probability of decoding error for any memoryless channel, is presented in this section.

Let \( X \) be an alphabet which is a group under the operation of addition, denoted by \( \oplus \). Let \( X^n \) be the set of all the sequences of length \( n \) over \( X \). Let \( C \) be a group code, namely, \( C \) is a subgroup of the group sequence \( X^n \). In other words, a group code is a set of sequences that forms a group under a componentwise group operation. Let \( C(x) \) be a coset of the group code \( C \) given by

\[ C(x) = \{ y : y \oplus x, y \in C, x \in X^n \}. \]

**Example 1:** Let the structure used for transmission be the well-known Leech lattice (see, e.g., [3] and [13]), for which \( n = 24 \), the dominant Euclidean distance is \( \delta_1^2 = 4 \), \( A_1 = 196560 \), and the covering radius is \( r_c = \sqrt{2} = 1.414 \). By (9) \( r_o = 1.314 \) is the optimal value of \( r \). By (16) we have \( r_p = 1.2975 \), which satisfies \( r_o > r_p > r_c \). It is easy to see that although \( M(r_o) = 2 \), the effect of \( \delta_1^2/2 \) is negligible. The upper bound (14) and the lower bound (15), which are very close to each other, are presented versus \( 6, \delta_1/2 \) in Fig. 4. The union bound, which is also presented in Fig. 4, is not tight for high levels of noise.

Note that if \( x \in C \), then \( C(x) \) is a permutation of \( C \).

**Theorem 1:** Let \( f(z|y) \) be the transition probability density function of the channel, which is assumed to be memoryless. Let

\[ B_i = - \ln \left( q^{-1} \sum_{x \in X} \int_{-\infty}^{\infty} \sqrt{f(z|\phi(x) \oplus x)} f(z|\phi(x)) \, dz \right), \]

where \( i \) is the \( i \)th element of the group alphabet \( X \) which contains \( q \) elements. The average error probability over the
ensemble of the cosets of the code is upper bounded by

$$P_e \leq \sum_{g \in V} A(g) \exp \left( - \sum_{i=1}^{q-1} \alpha_i B_i \right), \quad (19)$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_q)$, $\alpha_i$ is the number of symbols with the value $i$ in a codeword $g \in C$, $A(g)$ is the number of codewords with the same vector composition $\alpha$, and $V$ is the set of all the possible composition vectors. It is obvious that the sum of the elements of $\alpha$ does not exceed the length of the code, $n$.

The following method of proof is similar to the one used in [17].

**Proof:** Let $g_1 \oplus g_2 \in C(g)$ and $g_2 \oplus x \in C(g)$ be two different codewords. Assuming that these two codewords are the only transmitted sequences and that maximum likelihood decoding is performed, the probability $\lambda_i(g)$ of decoding error for the transmitted sequence $g_1 \oplus g_2 \oplus x$ is given by

$$\lambda_i(g) = \int_{x \in \mathbb{R}^n} \Psi(g_1 \oplus g_2 \oplus x) f(x) dx,$$

where $D(g_1 \oplus g_2 \oplus x)$ is the decision region of $g_1 \oplus g_2 \oplus x$. Let

$$\Psi(g) = \begin{cases} 1, & f(x) > f\left(x \in g_1 \oplus g_2 \oplus x\right), \\ 0, & f(x) < f\left(x \in g_1 \oplus g_2 \oplus x\right). \end{cases}$$

Since $\Psi(g) \leq \left( f(x) / f\left(x \in g_1 \oplus g_2 \oplus x\right) \right)^{1/2}$, by (20)

$$\lambda_i(g) = \int_{x \in \mathbb{R}^n} \left[ f(x) / f\left(x \in g_1 \oplus g_2 \oplus x\right) \right]^{1/2} dx,$$

where $\mathbb{R}^n$ denotes the $n$-dimensional Euclidean space. Substitution of

$$B(g_1, g_2) = \int_{x \in \mathbb{R}^n} \left[ f(x) f\left(x \in g_1 \oplus g_2 \oplus x\right) \right]^{1/2} dx$$

into (21) yields

$$\lambda_i(g) \leq B(g_1, g_2), \quad (22)$$

Let $\lambda(g)$ be the probability of decoding error for the transmission of message $l$ of the coset code $C(g)$ (i.e., the transmitted sequence is $g \oplus \vec{r}_i$). Since any sequence $\vec{x} \in X^n$ is chosen with the same probability $q^{-n}$, the average probability $\lambda_i$ of decoding error when $\vec{r}_i$ is transmitted is given by

$$\lambda_i = q^{-n} \sum_{g \in X^n} \lambda_i(g), \quad (23)$$

and $P_e = \sum_{i=1}^{M} \lambda_i$, where $M = |C|$. Assume that $g_1 = g \oplus C$. By (22), (23), and the union bound we obtain

$$P_e \leq q^{-n} \sum_{j=1}^{M} \sum_{g \in X^n} A(g) \exp \left( - \sum_{i=1}^{q-1} \alpha_i B_i \right).$$

Let $x^{(r)}$ denote the $r$th entry of $x \in X^n$. Since the channel is memoryless,

$$B(x, g_1 \oplus g_2 \oplus x) = \prod_{r=1}^{n} B\left(x^{(r)}, g_1^{(r)} \oplus x^{(r)}\right).$$

Substitution of (25) into (24) yields

$$P_e \leq q^{-n} \sum_{r=1}^{n} \sum_{x \in X^n} \left( \prod_{r=1}^{n} B\left(x^{(r)}, g_1^{(r)} \oplus x^{(r)}\right) \right) = \sum_{g \in X^n} A(g) \exp \left( - \sum_{i=1}^{q-1} \alpha_i B_i \right). \quad (26)$$

The upper bound of Theorem 1 is applicable to most of the known coded modulation schemes, any transmitted signal constellation, and any memoryless channel. Note that the resulting bound is a kind of union bound. However, for the AWGN channel we may use the sphere upper bound (7) (or (14)), which is obviously tighter than the random coset bound (19). Nevertheless, the following corollary we deduce that by utilizing the random coset method, the error coefficients $\{A_k\}$ can be evaluated for a finite constellation. Thus the random coset coding can serve as a method for calculating the average spectrum of distances of the code. By applying these coefficients to the sphere bound, further tightening of the upper bound is achieved (see Example 2).

**Corollary 1:** For the AWGN channel the upper bound (19) of Theorem 1 takes the form

$$P_e \leq q^{-n} \sum_{k=1}^{N} \tilde{A}_k \exp \left( - \delta_k^2 / 2 \sigma^2 \right),$$

where $\tilde{A}_k = q^{-n} \sum_{x \in X^n} A_k(g)$ is the average spectrum of distances over all the cosets of $C$ and $A_k(g)$ is the average spectrum of distances of the code $C(g)$. 

---

**Table 1**

<table>
<thead>
<tr>
<th>$C$</th>
<th>$\delta_1$</th>
<th>$\delta_2$</th>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$V(C_i)$</th>
<th>$r_e$</th>
<th>$r_p$</th>
<th>$r_c$</th>
<th>$N(t_0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_4$</td>
<td>1.414</td>
<td>2</td>
<td>24</td>
<td>24</td>
<td>2</td>
<td>1</td>
<td>0.708</td>
<td>0.854</td>
<td>1</td>
</tr>
<tr>
<td>$E_8$</td>
<td>1.414</td>
<td>2</td>
<td>2</td>
<td>240</td>
<td>2160</td>
<td>1</td>
<td>1</td>
<td>0.839</td>
<td>0.874</td>
</tr>
<tr>
<td>$K_{12}$</td>
<td>1.414</td>
<td>2.449</td>
<td>756</td>
<td>4032</td>
<td>27</td>
<td>1.633</td>
<td>1.285</td>
<td>1.318</td>
<td>2</td>
</tr>
<tr>
<td>$K_{16}$</td>
<td>2</td>
<td>2.449</td>
<td>4220</td>
<td>61440</td>
<td>16</td>
<td>1.732</td>
<td>1.302</td>
<td>1.324</td>
<td>2</td>
</tr>
<tr>
<td>$A_{24}$</td>
<td>2</td>
<td>2.449</td>
<td>195560</td>
<td>16773120</td>
<td>1</td>
<td>1.414</td>
<td>1.297</td>
<td>1.335</td>
<td>2</td>
</tr>
<tr>
<td>$A_{34}$</td>
<td>2</td>
<td>2.449</td>
<td>98256</td>
<td>8384512</td>
<td>2</td>
<td>1.732</td>
<td>1.335</td>
<td>1.344</td>
<td>2</td>
</tr>
<tr>
<td>$P_{480}$</td>
<td>2.449</td>
<td>2.828</td>
<td>5241600</td>
<td>39007332000</td>
<td>1</td>
<td>2</td>
<td>1.766</td>
<td>1.767</td>
<td>4</td>
</tr>
</tbody>
</table>
Proof: By (24) the upper bound on the average probability of decoding error (19) can be rewritten as
\[
P_e \leq q^{-n} \sum_{j=1}^{n} \sum_{g \in X^*} B(\bar{x}_j \oplus g, \bar{x}_j \oplus g),
\]
where for the AWGN channel
\[
B(\bar{x}_j \oplus g, \bar{x}_j \oplus g) = \int_{\bar{g} \in R^n} f(\bar{g} | e(\bar{x}_j \oplus g))
\]
and \(d_2\) is the squared Euclidean distance between \(e(\bar{x}_j \oplus g)\) and \(e(\bar{x}_j \oplus g)\). Therefore,
\[
\sum_{j=1}^{n} \sum_{g \in X^*} B(\bar{x}_j \oplus g, \bar{x}_j \oplus g) = \sum_{k=1}^{N} A(x) \exp(-\delta^2_k/8\sigma^2).
\]
Substitution of (30) into (28) yields (27).

An application of the method of random coset coding and the sphere upper bound is demonstrated in the next section for multilevel coding.

IV. APPLICATION TO MULTILEVEL CODING

The idea of multilevel coding, conceived by Imai and Hirakawa in an early paper [6], is a combination of several error correction codes employing a partition of some signal constellation into subsets. We are concerned with the case of a signal set of \(q = 2^n\) signals and an \(L\)-level partition chain. In this case any signal may be represented by \(m\) bits. Usually, the multilevel codes are based on linear binary codes. Let \(c_1, c_2, \ldots, c_m\) be a set of \(m\) binary linear codes with the same length \(n\); \(m \leq L\). These codes are trivial codes without redundancy (see, e.g., [2]-[6]). The set \(c_i\) constructs the multilevel code \(C\) over the alphabet \(X = \{0, 1, \ldots, 2^n - 1\}\). Let \(b = [b_1, b_2, \ldots, b_m]\) be the binary representation of the symbol \(x \in X\). Note that the operation of addition, \(x_1 \oplus x_2\), in the alphabet \(X\) is defined as an addition modulo 2 of the components in the binary representation of \(x_1\) and \(x_2\).

Here, we consider the case of \(m = 3\), a set of channel symbols \((\pm 3.5, \pm 2.5, \pm 1.5, \pm 0.5)\), and an AWGN channel. The mapping function is chosen to be
\[
\psi(x) = 0.5(-1)^{x_1} + (-1)^{x_2} + 2(-1)^{x_3}.
\]
In this case a two-level, two-way partition based on the signal set of amplitude modulation (AM) constellation of size 8 is employed. However, a two-level, four-way partition based on the signal set of quadrature amplitude modulation (QAM) constellation of size 64 may be employed in a similar fashion (see [4]). In Table I the values of \(B_i\) for 8-AM constellations are given by (18) are presented.

The following lemma is aimed at providing a method for calculating the average error coefficient, \(A_i\), of an important class of truncated lattice codes (e.g., the truncated Leech half lattice and the truncated Barnes–Wall lattices).

Lemma 1: Let a multilevel code consist of an 8-AM or 64-QAM constellation, where the mapping function is given in (31). Let the component code \(c_2\) be the single parity check code \((n, n-1, 2)\), and \(c_1\) be the trivial code \((n, n, 1)\). The dominant coefficient \(A_i\) can be calculated for an average coset of the code by
\[
A_i = A(d_i) \left( \sum_{j=0}^{d_i/2} \sum_{j=0}^{d_i/2} \left( \begin{array}{c} d_i \vspace{0.1cm} \\end{array} \right) \left( \begin{array}{c} d_i/2 \vspace{0.1cm} \end{array} \right) \right) \left( \begin{array}{c} d_i \vspace{0.1cm} \end{array} \right) 0.25U(8 - d_i) + 0.25A(d_i)U(d_i - 8),
\]
where \(U(\cdot)\) is the step function, \(\lfloor x \rfloor\) is the largest integer such that \(\lfloor x \rfloor \leq x\). \(d_j\) is the minimum Hamming distance of the component code \(c_j\), and \(A(d_j)\) is the number of codewords at a Hamming distance of \(d_j\) from a codeword of \(c_j\).

The proof of Lemma 1 is presented in Appendix A. For the calculation of the remaining coefficients, expressions which are similar to (32) can be derived. Mostly, the effect of these coefficients on the error probability is negligible. Conventionally we are interested in the error coefficient of the code, \(A_i\), rather than the average over the error coefficients of all the cosets of the code, \(A_i\). According to the following lemma, the coefficients \(A_i\) and \(A_i\) coincide for structures which satisfy the conditions of Lemma 1.

Lemma 2: If a multilevel code consists of an 8-AM or 64-QAM constellation, the component code \(c_2\) is the single parity check code \((n, n-1, 2)\), and \(c_1\) is the trivial code \((n, n, 1)\), then the error coefficient \(A(x)\) is the same for all cosets of the code. Therefore, \(A_i = A_i\).

The proof of Lemma 2 is given in Appendix B.

Example 2: Let the group code \(C\) be the Leech half-lattice code \(H_{24}\) [3], a sublattice of the Leech lattice \(\Lambda_{24}\). \(H_{24}\) is given [3] by
\[
H_{24} = c_1(24, 12, 8) + c_2(24, 23, 2) + 4Z^{24},
\]
where \(c_1\) is the (24, 12, 8) Golay code, \(c_2\) is a single parity check code of length 24, and the remaining bits are uncoded. The mapping function (31) is employed.

Since the structure of Example 2 satisfies the conditions of Lemma 1, where \(d_1 = 8\), \(d_2 = 2\), \(A(d_1) = 759\), and \(A(d_2) = 276\), \(A_i = 34003\). Due to Lemma 2, \(A_i = A_i\).

The error coefficient of the Leech half lattice, according to [2], is 98256, but the fact that the signal constellation is bounded was not taken into account when computing this figure. Because of the finite size of the constellation, the number of nearest code-words is not the same for all codewords and the error coefficient is less than 98256. In contrast, by the foregoing multilevel coset coding, an equal probability of decoding error is obtained. By (32) and Lemma 2, the error coefficient of the truncated \(H_{24}\) with a signal constellation of 8-AM (or 64-QAM) is \(A_i = 34003\).

The accurate error coefficient, which was calculated for Example 2, can be utilized for tightly upper bounding the probability of decoding error by (7) or (14). The sphere upper bound on the
probability of decoding error is presented in Fig. 5 versus the energy per bit to noise ratio, denoted by $E_b/N_0$. The union bound with an error coefficient of 98256 (see, e.g., [2]) and the union bound with an average error coefficient of 34003 are also presented in Fig. 5. By comparing the suggested sphere upper bound and the union bound with $A_1 = 98256$, we conclude that a significant tightening of the upper bound on the probability of decoding error is achieved for low signal to noise ratio. The improvement in the upper bound is due also to the use of the method of random coset coding for calculating $A_1$. This improvement can be observed by comparing the union bounds with $A_1 = 98256$ and $A_1 = 34003$. The advantage of an exact calculation of the error coefficient for a finite size constellation by the method of random coset coding is also illustrated in Fig. 6, where the sphere upper bound is presented for $A_1 = 98256$ (the lattice structure), $A_1 = 34003$ (truncated Leech half lattice with $q = 8$, i.e., the structure of Example 2), and $A_1 = 10004$ (truncated Leech half lattice with $q = 4$).

V. CONCLUSION

Sphere upper and lower bounds on the probability of decoding error were presented for a coded modulation system over the AWGN channel. A method of random coset coding was also presented. We used this method for deriving an upper bound for any memoryless channel (not necessarily AWGN). We would like to emphasize the following points.

1) The sphere upper bound described here can be interpreted as the probability of error for the code with the worst decision region among all the codes with a given set of spectrum of distances $\{A_i\}$. This worst decision region, defined by (12), is the polyhedron inscribed inside a sphere of radius $r_p$ (see Fig. 2).

2) Our sphere upper bound coincides with the bound of [9] for the conditional probability of error. However, the presented bound can also be employed for the average probability of error if the average spectrum of distances is known.

3) The lower bound we used is essentially the density sphere packing bound. This bound can be calculated if the volume of the Voronoi cell is known.

4) By a comparison between the upper and lower sphere bounds for the Leech lattice, we can deduce that from the viewpoint of the decoding error probability in the AWGN channel with a high level of noise, this lattice is very close to the hypothetical density packing code. The same conclusion applies to the lattices listed in Table I, on the basis of the comparison between the parameters $r_p$ and $r_q$.

5) If the method of random coset coding is employed for transmitting the data, then an equal probability of decoding error is provided for all transmitted sequences.

6) The method of random coset coding enables us to calculate a union bound for the average probability of decoding error (the average is taken over the cosets of the code). This union bound applies to any memoryless channel and any distribution of noise.

7) For the case of the AWGN channel this upper bound, which is based on the random coset coding, is used as a method for calculating the average spectrum of distances $\{A_i\}$ (the average is taken over the cosets of the code). This method is especially useful for structures with finite size constellations.

8) For the case of the truncated Leech half lattice it was shown that the dominant coefficient $A_1$ is equal for all the cosets of the code, thus $A_1 = \tilde{A}_1$. In the same fashion, it can be shown that the dominant components of the average spectrum of distances of many important structures are equal for all the cosets of the code. For structures in which these components of the spectrum are not equal for all the cosets, it would be interesting to find the coset with the lowest probability of decoding error.

APPENDIX A

PROOF OF LEMMA 1

The minimum squared Euclidean distance of a multilevel code with $L$ partition levels is given by $d_i^2 = \text{Min}\{d_1, d_2, \ldots, d_L\}$, where $d_i$ is the minimum squared Euclidean distance in the original signal constellation and $\Delta_i$ is the minimum squared Euclidean distance in the subset which is obtained at the partition level $i$. For the structure of Lemma 1 we have $L = 3$, $\Delta_0 = 1$, $\Delta_1 = 4$, and $d_2^2 = 2$. Therefore,

$$d_i^2 = \begin{cases} d_i, & d_i < 8, \\ 8, & d_i \geq 8. \end{cases} \quad (A1)$$

By Corollary 1, calculation of the upper bound of Theorem 1 provides a sum of coefficients multiplied by an exponential (see (27)), where $A_j$ is the desired error coefficient. Let $d_i < 8$. Then by (A1) $d_i^2 = d_i$, and, since $A_j$ is the coefficient which is multiplied by the exponential $e^{-d_i^2/2\sigma^2}$ in (27), $A_1$ can be obtained...
by computing the sum of (19) only for the composition vectors with the following form (see Table II):

$$g = (\alpha_1, 0, \alpha_2, 0, 0, 0, \alpha_3, 0); \quad \alpha_1 + \alpha_2 + \alpha_3 = d_i,$$

(A2)

Since $d_2 = 2$, $\alpha_1 + \alpha_2$ is even. By Table II we deduce that in this case,

$$\exp \left( - \sum_{i=1}^{q-1} \alpha_i B_i \right) = 0.5 \alpha_i e^{-\alpha_i/2\sigma^2} \left( e^{-1/2\sigma^2} + e^{-9/8\sigma^2} \right) \alpha_i,$$

and

$$A(g) = A(d_2) \left( \frac{d_2}{\alpha_2} \right).$$

(A4)

By (A2)–(A4) we have the error coefficient

$$A_1 = A(d_2) \sum_{i=0}^{1/(2\sigma^2)} \sum_{j=0}^{2/(2\sigma^2)} \left( \frac{2}{2^i} \right) 0.5^{2(1-i)} 0.25^i.$$

(A5)

Now let $d_3 > 8$. Then $\delta_i = 8$ and $A_1$ can be obtained by computing the sum of (19) only for the composition vectors which have the form

$$g = (0, \alpha_2, 0, 0, 0, 0, \alpha_3, 0); \quad \alpha_2 + \alpha_3 = 6.$$

(A6)

From Table II we deduce that in this case,

$$\exp \left( - \sum_{i=1}^{q-1} \alpha_i B_i \right) = 0.5 \alpha_i e^{-\alpha_i/2\sigma^2} \left( e^{-1/2\sigma^2} + e^{-9/8\sigma^2} \right) \alpha_i,$$

and

$$A(g) = A(d_2) \left( \frac{2}{\alpha_2} \right).$$

(A8)

By (A6)–(A8) we have the error coefficient

$$A_1 = A(d_2) \sum_{i=0}^{1/(2\sigma^2)} \sum_{j=0}^{2/(2\sigma^2)} \left( \frac{2}{2^i} \right) 0.5^i = 2.25 A(d_2).$$

(A9)

Clearly, if $d_3 = 8$, then $\delta_i = 8$ and the error-coefficient is determined by the codewords which have a composition vector of the form of (A2) or (A6).

**APPENDIX B**

**Proof of Lemma 2**

It is not difficult to see that the coset leaders of this structure have one of the following forms:

$$x_1 = (7^*, 0^n - 1),$$

(B1)

or

$$x_2 = (1, 7^*, 0^n - 1),$$

(B2)

where $x^*$ is a vector consisting of $n$ entries of $x$. Let $x^* = (0^n) \in C$, and let $x_1 = (x_1^{(1)}, x_2^{(1)}, \ldots, x_1^{(m)}) \in C$. If the coset leader has the form of (B1), then $x_1^{(1)} = (7^*, 0^n - 1) \in C(g)$. Thus, the elements of this coset, $x_1^{(1)}, x_1^{(2)}, \ldots \in C(g)$, have the form

$$x_1^{(1)} = (7 \oplus x_1^{(1)}, 7 \oplus x_1^{(2)}, \ldots, 7 \oplus x_1^{(m)}, x_1^{(r+1)}, \ldots, x_1^{(m)})$$

and

$$x_1^{(2)} = (7 \oplus x_1^{(1)}, 7 \oplus x_1^{(2)}, \ldots, 7 \oplus x_1^{(m)}, x_1^{(r+1)}, \ldots, x_1^{(m)}).$$

For multilevel coding, the group operation of addition is addition modulo 2 between the bits of the binary representation of each symbol. The binary representation of symbols $x \in X$ and their mapping $\varphi(x)$ into channel symbols are presented in Table III. Hence, by Table III, for $x_1, x_2 \in X$ the Euclidean distance between $\varphi(7 \oplus x_1)$ and $\varphi(7 \oplus x_2)$ is equal to the distance between $\varphi(x_1)$ and $\varphi(x_2)$. Therefore, the Euclidean distance between the transmitted codewords $\varphi[\varphi(x)]$ and $\varphi[\varphi(x)]$ is equal to the Euclidean distance between $\varphi[\varphi(x)]$ and $\varphi[\varphi(x)]$.

We now shall turn to the second class of cosets, which have a coset leader with the form of (B2). $u_1^{(1)}(x) = (1, 7^*, 0^n - 1) \in C(g)$. The elements of this coset, $u_1^{(1)}(x), u_1^{(2)}(x) \in C(g)$, have the form

$$u_1^{(1)}(x) = (1 \oplus u_1^{(1)}(x), 7 \oplus u_1^{(2)}(x), \ldots, 7 \oplus u_1^{(r+1)}(x), u_1^{(r+2)}(x), \ldots, u_1^{(m)}(x))$$

and

$$u_1^{(2)}(x) = (1 \oplus u_1^{(1)}(x), 7 \oplus u_1^{(2)}(x), \ldots, 7 \oplus u_1^{(r+1)}(x), u_1^{(r+2)}(x), \ldots, u_1^{(m)}(x)).$$

Since $(4, 0^n - 1) \in C,$

$$u_1^{(1)}(x) = (5 \oplus u_1^{(1)}(x), 7 \oplus u_1^{(2)}(x), \ldots, 7 \oplus u_1^{(r+1)}(x), u_1^{(r+2)}(x), \ldots, u_1^{(m)}(x)$$

$$\in C(g).$$

Clearly, if $u_1^{(1)}(x), u_1^{(2)}(x) \in C$ and the Euclidean distance between $u_1^{(1)}(x)$ and $u_1^{(2)}(x)$ is $\delta_1$, then $\|\varphi(u_1^{(1)}(x)) - \varphi(u_1^{(2)}(x))\| = 0, 1, 2; k = 1, 2, \ldots, n$.

By Table III we deduce the following.

a) If $\|\varphi(x_1) - \varphi(x_2)\| = 2$, then $\|\varphi(1 \oplus x_1) - \varphi(1 \oplus x_2)\| = 2$.

b) Excluding the pairs with $x_1 = 1, x_2 = 2, x_1 = 6, x_2 = 5, x_1 = 3, x_2 = 4$, if $\|\varphi(x_1) - \varphi(x_2)\| = 1$, then $\|\varphi(1 \oplus x_1) - \varphi(1 \oplus x_2)\| = 1$.

c) For these pairs $\|\varphi(5 \oplus x_1) - \varphi(1 \oplus x_2)\| = 1$. Thus, if $\|\varphi(u_1^{(1)}(x)) - \varphi(u_1^{(2)}(x))\| = \delta_1$ or $\|\varphi(u_1^{(1)}(x)) - \varphi(u_1^{(2)}(x))\| = \delta_1$.

Therefore, all the cosets of $C$ contain an equal average number of codewords at a distance $\delta_1$ from the transmitted codeword.

**Acknowledgment**

The authors would like to thank Prof. Jakov Snyders for constructive comments, which were helpful in improving the quality of the presentation.

**References**

The Weight Distribution of Cosets

Toreleiv Kløve

Abstract—Sullivan's inequality between the weight distribution function of a binary linear code and the weight distribution function of a proper subset of the code is generalized to linear codes over arbitrary finite fields.

Index Terms—Weight distribution, coset, inequality.

Consider a subset \( U \) of \( GF(q)^n \), that is, \( U \) is a set of vectors of length \( n \) whose elements are from the finite field of \( q \) elements. The \textit{(Hamming)} weight \( w_H(x) \) of a vector \( x \) is the number of nonzero elements of \( x \). The weight distribution of \( U \) is the sequence

\[
A(U) = A_0(U), A_1(U), \ldots, A_n(U),
\]

where \( A_i(U) \) is the number of vectors in \( U \) of weight \( i \). Moreover,

\[
A_i(z) = \sum_{j=0}^{n} A_j(U) z^i
\]

is known as the weight distribution function of \( U \).

Sullivan [1] proved an inequality which can be reformulated as follows:

\[
\text{Let } C \text{ be binary linear } [n, k] \text{ code and } S \text{ a proper coset of } C. \quad \text{Then}
\]

\[
A_S(z) \leq \frac{(1 + z)^{k+1} - (1 - z)^{k+1}}{(1 + z)^{k+1} + (1 - z)^{k+1}} A_C(z)
\]

for all \( z \in [0,1] \).

One application of Sullivan's theorem was given by Wolf, Michelson, and Levesque [2], who gave a lower bound on the probability of undetected error for a binary linear code: For any binary \([n,k]\) code \( C \) and any \( p \in [0, \frac{1}{2}] \) we have

\[
P_u(C, p) \geq \frac{1 + (1-2p)^{k+1}}{2^{n-k} - (2^{n-k} - 2)(1 - 2p)^{k+1}} - (1-p)^n.
\]

Sullivan's inequality was an important ingredient in the proof of this result.

The increased use of nonbinary codes for practical applications makes it of interest to have a generalization of Sullivan's and Wolf's theorems to linear codes over any finite field; such generalizations are given in this correspondence. The proof of Theorem 1 below is a generalization and simplification of Sullivan's proof for the binary case [1], and the proof of Theorem 2 is a direct generalization of the proof of Wolf, Michelson, and Levesque [2].

Theorem 1: Let \( C \) be an \([n,k]\) code over the field \( GF(q) \) and let \( S \) be a proper coset of \( C \). Then

\[
A_S(z) \leq \frac{(1 + (q-1)z)^{k+1} - (1 - z)^{k+1}}{(1 + (q-1)z)^{k+1} + (q - 1)(1 - z)^{k+1}} A_C(z)
\]

for all \( z \in [0,1] \).

\textbf{Proof:} We will use the following notations in this proof:

\[
F_k(z) = \frac{(1 + (q-1)z)^{k+1} - (1 - z)^{k+1}}{(1 + (q-1)z)^{k+1} + (q - 1)(1 - z)^{k+1}}
\]

\[
T(R, z) = \frac{z + (1 + (q-2)z)R}{1 + (q - 1)zR}.
\]

In this notation, the result of the theorem can be formulated as follows:

\[
A_S(z) \leq F_k(z) A_C(z)
\]

for the following observations (valid for \( z \in [0,1] \) and \( k \geq 0 \):

\[
\begin{align*}
&\text{If } R_1 \geq R_2, \text{ then } T(R_1, z) \geq T(R_2, z), \\
&F_{k+1}(z) = T(F_k(z), z), \\
&z = F_0(z) \leq F_k(z) \leq F_{k+1}(z) \leq 1.
\end{align*}
\]

We introduce the further notation for \( j \in GF(q) \):

\[
(C_j) = \{(x_1, x_2, \ldots, x_{n-1}, j) | (x_1, x_2, \ldots, x_{n-1}) \in C_j\}
\]

where \( j \) is \( i \)-invariant.