On the Basic Averaging Arguments for Linear Codes

Hans-Andrea Loeliger
ISY / Information Theory
Linköping University
S-58183 Linköping, Sweden

Abstract

Linear codes over F_q are considered for use in detecting and in correcting the additive errors in some subset E of F_q^n. (The most familiar example of such an error set E is the set of all n-tuples of Hamming weight at most t.) In this set-up, the basic averaging arguments for linear codes are reviewed with emphasis on the relation between the combinatorial and the information-theoretic viewpoint. The main theorems are (a correspondingly general version of) the Varshamov-Gilbert bound and a 'random-coding' bound on the probability of an ambiguous syndrome. These bounds are shown to result from applying the same elementary averaging argument to two different packing problems, viz., the combinatorial 'sphere' packing problem and the probabilistic 'Shannon packing'. Some applications of the general bounds are indicated, e.g., hash functions and Euclidean-space codes, and the connection to Justesen-type constructions of asymptotically good codes is outlined.

I Introduction

This paper is a essentially a tutorial review of the basic averaging arguments for linear codes. The main results that are proved are (a version of) the Varshamov-Gilbert bound and a 'random-coding' bound for linear codes.

This will hardly sound exciting — the mentioned venerable bounds belong to the very foundations of coding theory and have been proved and generalized in dozens of ways. What, then, is the purpose of this paper?

One of the origins of this paper is the deep confusion in which I once was put by the following 'paradox'. Consider the binary symmetric channel with crossover probability ε, whose capacity is $C \triangleq 1 - h(\varepsilon)$, where $h(\varepsilon) \triangleq -\varepsilon log_2 \varepsilon - (1-\varepsilon)\log(1-\varepsilon)$ is the binary entropy function. It is well known from basic information theory that, for a fixed rate R less than (but arbitrarily close to) C, almost all binary linear codes of rate R and sufficiently large blocklength n have vanishingly small error probability on this channel. One could thus expect that, for R close to C and large n, most codes have a relative minimum distance d/n of about 2ε. We would thus have $R \approx 1 - h(d/2n)$, which coincides with the asymptotic Hamming upper bound. However, the asymptotic Hamming bound is well known not to be achievable; in fact, asymptotically, almost all binary linear codes are known to have rates close to that of the Varshamov-Gilbert bound, viz., $R = 1 - h(d/n)$.

This 'paradox' exposes a conceptual gap between information theory and combinatorial coding theory of which engineers should be aware, but over which most textbooks pass with silence. (A commendable exception is the classical text by Peterson and Weldon [1, Chapt. 4.3].)

Therefore, the primary purpose of this paper is to illuminate the relation between the information-theoretic and the combinatorial view of coding. We will see that the Varshamov-Gilbert bound and a probabilistic random-coding bound are obtained from applying the same basic averaging argument to two different sphere packing problems, viz., the combinatorial packing of rigid spheres and the probabilistic 'Shannon packing', where the spheres are allowed to overlap slightly.

(It is appropriate here to mention that the intimate relation between the Varshamov-Gilbert bound and random coding à la Shannon was apparently first noticed in the Ph. D. thesis of Jim Massey [2].)

A basic feature of our exposition is that we will consider arbitrary sets $E \subseteq F_q^n$ of additive error patterns, not just the standard case where E is the set of n-tuples of Hamming weight at most t. One motivation for this generality are applications such as burst error correction, multiple access communications, hash functions, constrained codes, and Euclidean-space coding, which will be cursorily reviewed in Section IV. However, the consideration of an arbitrary error set E is believed to be valuable also from a purely pedagogical viewpoint.

The paper is structured as follows. In Section II, the essence of the averaging arguments of this paper is summarized in the form of a definition and three elementary lemmas. The core of the paper is Section III, where (a form of) the Varshamov-Gilbert bound and our probabilistic random-coding bound are derived. The proof of the latter — the only nontrivial proof of this paper — is so simple that it could well be used as an exercise in a first course on algebraic coding theory. These results are then discussed in Section IV, where also some applications are indicated and the connections to Justesen-type constructions of asymptotically good codes are outlined.

To conclude this introduction, I would like to mention that the material of this paper dates back to the time when I was a graduate student of Jim Massey, and he has repeatedly encouraged me to publish it. Here it is — happy birthday!

II Three Averaging Lemmas

Let F_q denote the finite field with q elements. As usual, an (n, k) *q-ary linear code* is a k-dimensional subspace of the vector space F_q^n of n-tuples over F_q, and the *rate* of such a code is the fraction k/n. For any subset E of F_q^n, the set $\{e \in E : e \neq 0\}$ of nonzero elements will be denoted by E^*.

The averaging arguments of this paper hold for every set of codes that is balanced in the following sense.

Definition 1 *A nonempty set $\mathcal{C}$ of linear (n, k) codes over F_q is* balanced *if every nonzero element of F_q^n is contained in the same number, denoted by $N_\mathcal{C}$, of codes from $\mathcal{C}$.*

It is rather obvious that, for fixed n, k, and q, the set of *all* linear (n, k) codes over F_q is balanced. Further examples of balanced sets of codes will be given in Section IV.

The term 'balanced' stems from [3], where, however, it is used with a slightly different meaning, viz., as a property of a set of affine encoders (rather than of linear codes). Very similar sets of codes were also considered in [4]. It was noted in these references that balancedness is a combinatorial version of *pairwise-independence,* which is the key property underlying the usual random coding arguments, cf. [5, Chapt. 6.2]. For sufficiently symmetric channels (i.e., for 'regular' channels in the sense of [3]), the average error probability taken over a balanced set of codes is thus upper bounded by Gallager's celebrated random coding bound [3], [5]. In Section III, we will see that a similar random coding bound can be derived very easily.

The multiplicity $N_{\mathcal{C}}$ of every nonzero q-ary n-tuple in a balanced set $\mathcal{C}$ of (n, k) linear codes is related to the number of codes $|\mathcal{C}|$ by

$$|\mathcal{C}|\,(q^k - 1) = N_{\mathcal{C}}\,(q^n - 1). \tag{1}$$

(This is proved by counting the total number of nonzero codewords of all codes in $\mathcal{C}$, each with its multiplicity.) It will be useful to remember that, for all positive integers n, k, and q such that $q > 1$ and $k \leq n$,

$$\frac{q^n - 1}{q^k - 1} \geq \frac{q^n}{q^k} \tag{2}$$

and the inequality is strict for $k < n$.

Lemma 1 (Basic Averaging Lemma) *Let $f(\cdot)$ be an arbitrary real valued[1] function defined on F_q^n; let $\mathcal{C}$ be a balanced set of linear (n, k) codes over F_q. Then the average, over all codes C in $\mathcal{C}$, of the sum $\sum_{c \in C^*} f(c)$ (over all nonzero codewords) is given by*

$$\frac{1}{|\mathcal{C}|} \sum_{C \in \mathcal{C}} \sum_{c \in C^*} f(c) = \frac{q^k - 1}{q^n - 1} \sum_{v \in (F_q^n)^*} f(v).$$

Proof: It follows from the definition of a balanced set of codes that

$$\sum_{C \in \mathcal{C}} \sum_{c \in C^*} f(c) = N_{\mathcal{C}} \sum_{v \in (F_q^n)^*} f(v),$$

and the lemma follows from (1). □

We will use Lemma 1 primarily in the more special form of the following lemma.

Lemma 2 (Average Intersection Cardinality Lemma) *Let $\mathcal{C}$ be a balanced set of linear (n, k) codes over F_q; let E be an arbitrary subset of F_q^n. Then the average cardinality of $C^* \cap E$ over all codes C in $\mathcal{C}$ is given by*

$$\frac{1}{|\mathcal{C}|} \sum_{C \in \mathcal{C}} |C^* \cap E| = \frac{q^k - 1}{q^n - 1}\,|E^*|. \tag{3}$$

[1] The range of $f(\cdot)$ can actually be more general.

Proof: Define $f : F_q^n \to \{0, 1\}$ as $f(v) = 1$ if $v \in E$ and $f(v) = 0$ otherwise and apply Lemma 1. □

The last lemma of this section shows that, under a slightly stronger balancing condition for $\mathcal{C}$, the cardinality of $C^* \cap E$ is close to the average value (3) for most codes C in $\mathcal{C}$.

Lemma 3 (Intersection Variance Lemma) *Let $\mathcal{C}$ be a set of linear (n, k) codes over F_q that is* doubly balanced, *i.e., it is balanced and every pair of linearly independent elements of F_q^n is contained in the same number of codes from $\mathcal{C}$. Let E be an arbitrary nonempty subset of F_q^n with at least one nonzero element. Then*

$$\frac{1}{|\mathcal{C}|}\sum_{C\in\mathcal{C}} |C^* \cap E|^2 - \left(\frac{1}{|\mathcal{C}|}\sum_{C\in\mathcal{C}} |C^* \cap E|\right)^2 \quad < \quad (q-1)\, q^{k-n}\, |E^*|.$$

It then follows from Chebyshev's inequality that the fraction of codes C in $\mathcal{C}$ for which

$$\left| |C^* \cap E| - \frac{1}{|\mathcal{C}|}\sum_{C\in\mathcal{C}} |C^* \cap E| \right| \geq \gamma\, q^{k-n}|E^*|$$

is at most $(q-1)^2/\gamma^2$. Since Lemma 3 will not be referred to in the sequel, the proof is omitted and the interested reader is referred to [6].

III Error Detection and Correction

Consider the following textbook situation. A transmitter selects a codeword c from an (n, k) linear code C over F_q and sends it over a noisy channel. The channel adds an error pattern $e \in F_q^n$ to the transmitted codeword, and the task of the receiver is to estimate c from $c + e$.

It is clear that, for $k > 0$, c cannot be recovered from $c + e$ for all possible codewords $c \in C$ and all possible error patterns $e \in F_q^n$. Therefore, we restrict our attention to a set $E \subset F_q^n$ of typical error patterns (where 'typical' is meant informally). The most popular choice for E is the discrete ball $S_{n,r} \triangleq \{v \in F_q^n : d(v, 0) \leq r\}$, where $d(\cdot\,, \cdot)$ denotes either Hamming distance or any other suitable metric on F_q^n, but the arguments below hold for arbitrary $E \subseteq F_q^n$.

We will use the notation

$$H_q(E) \triangleq 1/n \log_q |E| \tag{4}$$

for the 'entropy' of the error patterns E or of any other subset of F_q^n. (If a statistical channel model for the noise is available, then, for a reasonable choice of E, the set-theoretic entropy (4) is, of course, closely related to, and asymptotically identical with, the information-theoretic entropy of the noise. In fact, even in a purely combinatorial context, the set-theoretic entropy (4) can sometimes be closely approximated by the information-theoretic entropy of a suitable auxillary probability distribution, cf. [7, Sec. 8], [8].)

At this point, the purely combinatorial viewpoint and the probabilistic (information-theoretic) viewpoint begin to differ. We begin with the former.

We say that the code C *corrects* all errors in E if c (and thus also e) can be recovered from $c + e$ for all codewords $c \in C$ and all error patterns $e \in E$. This is clearly impossible if $q^k|E| > q^n$, which is expressed in the following classical bound.

Proposition 1 (Hamming Bound) *Let E be an arbitrary nonempty subset of F_q^n; let $C \subseteq F_q^n$ be a code (not necessarily linear) with q^k codewords that corrects all errors in E. Then*

$$|E| \leq q^{n-k}$$

or, equivalently, $H_q(E) \leq 1 - k/n$.

It is convenient for the following argument to consider also error detection. A linear code *detects* all errors in a set E of error patterns if and only if $C^* \cap E = \emptyset$.

It is easily seen that a linear code C corrects all errors in E if and only if $C \cap \Delta E = \{0\}$, where $\Delta E \overset{\triangle}{=} \{e - e' : e, e' \in E\}$ is the set of all differences of elements of E. Thus C *corrects* all errors in E if and only if it *detects* all errors in ΔE.

We now use Lemma 2 to obtain an existence proof for error detecting and correcting codes. Let $\mathcal{C}$ be a balanced set of linear (n, k) codes over F_q, and let E be an arbitrary nonempty subset of F_q^n. If the right side of (3) is less than 1, then, on the average over all codes in $\mathcal{C}$, $|C^* \cap E| < 1$, which implies the existence of at least one code C in $\mathcal{C}$ that detects all errors in E; a sufficient condition for this is $|E^*| < q^{n-k}$. We have proved the following theorem, which is one form of the classical Varshamov-Gilbert bound.

Theorem 1 (Varshamov-Gilbert Bound) *Let $\mathcal{C}$ be a balanced set of linear (n, k) codes over F_q and let E be an arbitrary subset of F_q^n that contains the all zero vector. If*

$$|E| \leq q^{n-k} \tag{5}$$

or, equivalently, if $H_q(E) \leq 1 - k/n$, then there exists a code in $\mathcal{C}$ that detects *all errors in E; if*

$$|\Delta E| \leq q^{n-k} \tag{6}$$

or, equivalently, if $H_q(\Delta E) \leq 1 - k/n$, then there exists a code in $\mathcal{C}$ that corrects *all errors in E.*

The following corollary is a bit closer to the usual formulations of the Varshamov-Gilbert bound.

Corollary 1 *Let $d(\cdot\,,\cdot)$ be a metric on F_q^n that satisfies $d(v, v') = d(v - v', 0)$ for all $v, v' \in F_q^n$ (i.e., $d(\cdot\,,\cdot)$ is translation invariant, as is, e.g., Hamming distance). Then there exists a code in $\mathcal{C}$ with minimum distance d such that*

$$|S_{n,d}| > q^{n-k}. \tag{7}$$

Proof: Due to the assumed property for $d(\cdot\,,\cdot)$, it suffices to consider the distances from the all zero codeword. Assume that the largest minimum distance d of any code in $\mathcal{C}$ satisfies $|S_{n,d}| \leq q^{n-k}$. The first part of Theorem 1 then implies the existence of a code $C \in \mathcal{C}$ such that $C^* \cap S_{n,d} = \emptyset$. The minimum distance of C is thus larger than d, a contradiction. □

When applied to Hamming distance, Corollary 1 is slightly weaker than Gilbert's bound [9], which in turn is slightly weaker than Varshamov's bound [10, pp. 33–34]. However, all forms agree asymptotically for $n \to \infty$, which is the case of primary interest. The minor nonasymptotic weakness of Corollary 1 seems to be the price for the generality of Theorem 1.

We now reconsider the situation from an information-theoretic (probabilistic) viewpoint. We assume that both the transmitted codeword $\mathbf{c}$ and the error pattern $\mathbf{e}$ are random variables (indicated by bold type); the former takes values in a q-ary linear code C and the latter takes values in F_q^n. We further assume that $\mathbf{e}$ and $\mathbf{c}$ are independent.

In this probabilistic set-up, Lemma 1 has an immediate interpretation for error *detection.*

Proposition 2 *Let $\mathcal{C}$ be a balanced set of linear (n,k) codes over F_q. The arithmetic average, over all codes C in $\mathcal{C}$, of the probability $P_{ue} \triangleq \sum_{c \in C^*} P(\mathbf{e} = c)$ of an undetectable error is given by*

$$\overline{P_{ue}} = \frac{q^k - 1}{q^n - 1} \left(1 - P(\mathbf{e} = 0)\right),$$

which implies $\overline{P_{ue}} \leq q^{k-n}$.

The published bounds of this type are usually restricted to particular additive channels such as, e.g., the binary symmetric channel [11], [12]. It is therefore remarkable that the upper bound q^{k-n} of Proposition 2 holds independently of the probability distribution of $\mathbf{e}$.

A comparison of Proposition 2 with the first part of Theorem 1 reveals a striking difference between the probabilistic viewpoint of the former and the combinatorial viewpoint of the latter: According to Theorem 1, error detection (with 'average' linear codes) costs $H_q(E)$ in code rate, whereas Proposition 2 makes clear that, for $n \to \infty$, arbitrarily reliable error detection is possible with code rates arbitrarily close to one.

We now consider error *correction.* As before, we restrict our attention to a set $E \subseteq F_q^n$ of typical, i.e., high probability error patterns. (Again, 'typical' is not meant formally — the derivation below holds for arbitrary sets $E \subseteq F_q^n$.)

The event that the received vector $\mathbf{y} \triangleq \mathbf{c} + \mathbf{e}$ can be written in more than one way as $\mathbf{y} = c + e$ with $c \in C$ and $e \in E$ will be called an *ambiguity.* Let $P_{amb|E}$ denote the probability of an ambiguity, conditioned on the event that $\mathbf{e}$ is in E. It is easily seen that further conditioning on the transmitted codeword does not change the probability of an ambiguity, i.e., $P_{amb|E} = P(\text{ambiguity} \mid \mathbf{e} \in E \text{ and } \mathbf{c} = c)$ for all $c \in C$.

Consider a decoder that operates according to the following rule. If $\mathbf{y}$ has a unique decomposition $\mathbf{y} = c + e$ with $c \in C$ and $e \in E$, then $\mathbf{y}$ is decoded to c. We need not specify the decoder action for any other case. For any such decoder, the probability P_e of a decoding error is bounded by

$$P_e \leq P_{amb|E} + P(\mathbf{e} \notin E). \tag{8}$$

In general, it is very difficult to compute $P_{amb|E}$. However, its average, over all codes of a balanced set of codes, is readily bounded as follows.

Theorem 2 (Random Coding Bound) *The arithmetic average, over all codes of a balanced set of q-ary linear (n,k) codes, of $P_{amb|E}$ is bounded by*

$$\overline{P_{amb|E}} \leq q^{k-n} \, |E| \tag{9}$$

or, equivalently, by $\overline{P_{amb|E}} \leq q^{-n[1-k/n-H_q(E)]}$.

Proof: Since $P_{amb|E}$ is independent of the transmitted codeword, we can assume that the all-zero codeword is transmitted. Any given received y has a unique decomposition $y = c + e$, with $c = 0$ and $e \in E$, if and only if $(y - E) \cap C = \{0\}$. Therefore, $P_{amb|E}$ is bounded by

$$P_{amb|E} \leq \sum_{y \in F_q^n} P\left(\mathbf{y} = y \mid \mathbf{c} = 0 \text{ and } \mathbf{e} \in E\right) \cdot \left|C^* \cap (y - E)\right|,$$

which is the critical step of the proof. The rest is straightforward: averaging over all codes in $\mathcal{C}$ and applying Lemma 2 yields

$$\begin{aligned}
\overline{P_{amb|E}} &= \frac{1}{|\mathcal{C}|} \sum_{C \in \mathcal{C}} P_{amb|E} \\
&\leq \sum_{y \in F_q^n} P\left(\mathbf{y} = y \mid \mathbf{c} = 0 \text{ and } \mathbf{e} \in E\right) \frac{1}{|\mathcal{C}|} \sum_{C \in \mathcal{C}} |C^* \cap (y - E)| \\
&= \sum_{y \in F_q^n} P\left(\mathbf{y} = y \mid \mathbf{c} = 0 \text{ and } \mathbf{e} \in E\right) \frac{q^k - 1}{q^n - 1} |(y - E)^*| \\
&\leq q^{k-n} |E|.
\end{aligned}$$

□

Note that Theorem 2 implies Shannon's channel coding theorem for all ergodic q-ary additive channels: if E are the typical error patterns (where, this time, 'typical' is meant more formally, cf. [13]), it is clear that both right-hand terms of (8) vanish for $n \to \infty$ and fixed code rate k/n, provided only that the entropy $H_q(E)$ of the errors tends to a limit below $1 - k/n$. Note, however, that Theorem 2 makes sense even without assuming ergodic errors and is in this sense more general than the usual information-theoretic random coding bounds.

IV Discussion

The striking formal similarity among the results of Section III makes it easy to compare the combinatorial 'sphere' packing problem with the information-theoretic 'Shannon packing', as was promised in the introduction. Thereafter, we will sketch how the generality of the two theorems of Section III makes them useful for a variety of applications. Finally, some balanced classes of linear codes are briefly mentioned and the connection to the asymptotically 'good' codes of Justesen [14] and of Delsarte and Piret [3] is outlined.

IV-1 Shannon Packing vs. Rigid-Sphere Packing

How many Hamming spheres — or, more generally, arbitrary 'objects' $E \subseteq F_q^n$ — can be packed into F_q^n such that they do not overlap? An obvious upper bound is the Hamming bound (Proposition 1), and the Varshamov-Gilbert bound (Theorem 1) gives a lower bound: if the spheres are centered on the codewords of a linear code, Theorem 1 gives the code rate where, on the average over all codes of a balanced set of codes, the intersection of the spheres drops below one. The asymptotically best possible packing rate (for spheres) is not known in general — its determination is considered by mathematicians as one of the main goals of coding theory — but it is known to be closer to the Varshamov-Gilbert bound than to the Hamming bound.

A different packing problem was (implicitly) introduced — and solved — by Shannon: the requirement that the spheres do not intersect is relaxed to the condition that the volume of the intersection of any sphere with its neighbors is at most a fraction ε of the volume of the sphere. A very sharp and general solution to this 'Shannon packing' problem is Theorem 2, applied to the case that $\mathbf{e}$ is uniform over E: no matter how the objects E are shaped and for any positive ε, the asymptotically achievable packing rate coincides with that for cubes of the same volume, i.e., with the Hamming bound! Moreover, the same answer still holds if the interior of E is weighted by an arbitrary probability distribution.

It can not be overemphasized that the latter type of packing (i.e., the Shannon packing) is the more important one for most engineering applications. In particular, the restriction of most algebraic decoding algorithms to *spherical* decoding regions (i.e., bounded-distance decoding, cf. [15]) is one of the reasons for the limited practical usefulness of much of algebraic coding theory. (The other main problem of algebraic decoding is the well-known difficulty to use 'soft-decision' reliability information.) In fact, it is evident that, in all successful applications of coding to channels with moderate to high noise level, the decoding regions are far from being spherical; good examples are convolutional codes and concatenated codes of all kinds.

IV-2 Applications

The generality of Theorems 1 and 2 makes them useful in a large variety of applications. For the correction of burst errors or in similar situations, all that needs to be done is to specify the set E of error patterns and to evaluate its cardinality.

An interesting application are multiple-access systems. Any of a collection of users can lump the activity of the other users together with the channel noise into a set E of possible interference patterns. It is clear from Theorem 2 that the full sum capacity of one q-ary symbol per channel use is achievable if each user independently and randomly selects a code from a balanced set of codes.

The syndrome mapping $F_q^n \rightarrow F_q^r$ of a linear $(n, k = n - r)$ code can be viewed as a linear source encoder for the error patterns. In this way, any result on linear codes has an interpretation in source coding and vice versa, cf. [16].

The primary application area of linear source coding is the technique of hashing in computer science: a set $E \subseteq F_2^n$ of 'keys' is coded into F_2^r by means of a hash function $F_2^n \rightarrow F_2^r$, where $r < n$ is so small (typically in the range 8–16) that decoding can be done by table lookup. (Large keys are used, however, for cryptographic purposes.) In fact, results essentially equivalent to those of Section III have been published in the literature on hash functions, and our notion of a balanced set of linear codes is closely related to (the specialization to linear hash functions of) the notion of a universal class of hash functions, cf. [17].

Another application area is shaping of linear codes; i.e., we use a linear code C for error correction but allow only the codewords $C \cap E_s$ that satisfy certain constraints, which are specified by a set $E_s \subset F_q^n$ of allowed words. E.g., it is easily established from the arguments of this paper that constrained subcodes of linear codes can achieve any rate less than $H(E_s) - H(E_e)$ on any q-ary channel with additive errors of entropy $H(E_e)$, cf. [6]. (The specialization of this result to the binary symmetric channel and runlength-limited codes has been reported in [18].)

The averaging arguments of this paper can also be applied to Euclidean-space codes and lattices. E.g., it is clear that Theorem 1, as well as Corollary 1, applies to M-PSK (M-ary phase-shift keying) when M is a prime and linear codes over F_M are used. In fact, it was shown in [19] that the Minkowski-Hlawka theorem — the basic asymptotic existence theorem for lattices — can be derived from Lemma 1 (and essentially the same proof is outlined in [20, pp. 534–535]). Further applications to Euclidean-space codes and lattices are given in [6].

IV-3 Balanced Classes of Codes and Explicit Constructions of Asymptotically Good Codes

An interesting method to construct balanced sets of linear codes over F_q is due to Delsarte and Piret [3]. The space F_q^n of q-ary n-tuples is identified with the field F_{q^n}. Let F_{q^k} be a subfield of F_{q^n}. For any nonzero $v \in F_{q^n}$, let $\overline{v} \overset{\triangle}{=} \{av : a \in F_{q^k}\}$ be the set of all F_{q^k} multiples of v, which is clearly a linear (n, k) code over F_q. But the set $\mathcal{C} \overset{\triangle}{=} \{\overline{v} : v \in F_{q^n}^*\}$ of these codes is a partition of $F_{q^n}^*$ and therefore balanced with $N_{\mathcal{C}} = 1$.

This set $\mathcal{C}$ was used in [3] for a Justesen-type concatenated code construction where the inner code varies over all codes in $\mathcal{C}$ and the outer code is a Reed-Solomon code; it was shown that, for regular channels, such code constructions allow the derivation of an upper bound on error probability of the form $P_e < q^{-NE(R)}$, where N and R are the blocklength and the rate, respectively, of the concatenated code, and where the exponent $E(R)$ is positive for all rates R less than channel capacity. (Since this construction relies on the goodness of the inner codes $\mathcal{C}$, the exponent $E(R)$ is smaller than Gallager's exponent.)

We conclude the discussion of balanced sets of codes by pointing out that weaker versions of Theorems 1 and 2 can often be proved for sets of codes that are only 'almost balanced'. In particular, an asymptotic version of Definition 1 suffices to prove the asymptotic Varshamov-Gilbert bound and information-theoretic random coding bounds. Apart from the celebrated codes from algebraic geometry [21] [20], all classes of codes that have so far been shown to satisfy the asymptotic Varshamov-Gilbert bound are of this type, e.g., alternant codes, (classical) Goppa codes, quasi-cyclic codes, self-dual codes [10], and shortened cyclic codes (cf. [22, Appendix II]). The last member in this list (i.e., the shortened cyclic codes) is seldom mentioned in textbooks but (for error detection) very popular in applications.

V Conclusions

We have seen that the elementary averaging arguments for linear codes are amazingly versatile. They can be used both in a combinatorial way (which leads to Varshamov-Gilbert-type bounds) and in a probabilistic way (which leads to Shannon-type random coding theorems) and thereby illustrate the discrepancy between these two approaches to coding. We have also seen that these averaging arguments, despite their simplicity, are able to yield nontrivial insights in a surprising variety of application areas.

References

[1] W. W. Peterson and E. J. Weldon, Jr., *Error-Correcting Codes*, 2nd ed., Cambridge: MIT Press, 1972.

[2] J. L. Massey, *Threshold Decoding.* Cambridge, Mass.: MIT Press, 1963.

[3] Ph. Delsarte and Ph. Piret, 'Algebraic constructions of Shannon codes for regular channels', *IEEE Trans. Inform. Theory*, vol. 28, pp. 593–599, July 1982.

[4] G. Séguin, 'Linear ensembles of codes', *IEEE Trans. Inform. Theory*, vol. 25, pp. 477–480, July 1979.

[5] R. G. Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.

[6] H.-A. Loeliger, 'On the information-theoretic limits of lattices and related codes', in preparation.

[7] C. E. Shannon, 'A mathematical theory of communication', *Bell Syst. Techn. J.,* vol. 27, pp. 379–423, July 1948, and pp. 623–656, Oct. 1948. Reprinted in *Key Papers in the Development of Information Theory*, New York: IEEE Press, 1974.

[8] H.-A. Loeliger, 'An upper bound on the volume of discrete spheres', submitted to *IEEE Trans. Inform. Theory.*

[9] E. N. Gilbert, 'A comparison of signalling alphabets', *Bell Syst. Techn. J.,* vol. 31, pp. 504–522, May 1952. Reprinted in *Key Papers in the Development of Information Theory,* New York: IEEE Press, 1974.

[10] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes,* Amsterdam: North-Holland, 1977.

[11] J. L. Massey, 'Coding techniques for digital data networks', in *Proc. Int. Conf. Inform. Theory and Syst.,* NTG-Fachberichte, vol. 65, Berlin, Germany, Sept. 18–20, 1978.

[12] J. K. Wolf, A. M. Michelson, and A. H. Levesque, 'On the probability of undetected error for linear block codes', *IEEE Trans. Comm.*, vol. 30, pp. 317–324, Feb. 1982.

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

[14] J. Justesen, 'A class of constructive asymptotically good algebraic codes', *IEEE Trans. Inform. Theory,* vol. 18, pp. 652–656, Sept. 1972. Reprinted in *Key Papers in the Development of Coding Theory*, New York: IEEE Press, 1974.

[15] A. D. Wyner, 'Capabilities of bounded discrepancy decoding', *Bell Syst. Tech. J.,* vol. 54, pp. 1061–1122, 1965.

[16] T. C. Ancheta, Jr., ‘Syndrome-source-coding and its universal generalization’, *IEEE Trans. Inform. Theory*, vol. 22, pp. 432–436, July 1976.

[17] J. L. Carter and M. N. Wegmann, ‘Universal classes of hash functions’, *Journal of Computer and System Sciences*, vol. 18, pp. 143–154, 1979.

[18] A. Patapoutian and P. V. Kumar, ‘The (d, k) subcode of a linear block code’, *IEEE Trans. Inform. Theory*, vol. 38, pp. 1375–1382, July 1992.

[19] H.-A. Loeliger, ‘On existence proofs for asymptotically good Euclidean-space group codes’, Proc. of Joint DIMACS/IEEE Workshop on Coding and Quantization, Piscataway, NJ, USA, Oct. 19-21, 1992, to appear.

[20] M. A. Tsfasman and S. G. Vlăduţ, *Algebraic-Geometric Codes,* Kluwer, 1991.

[21] M. A. Tsfasman, S. G. Vlădut, and Th. Zink, ‘Modular curves, Shimura curves, and Goppa codes, better than Varshamov-Gilbert bound’, *Math. Nachr.,* vol. 109, pp. 21–28, 1982.

[22] T. Kasami, ‘An upper bound on k/n for affine-invariant codes with fixed d/n’, *IEEE Trans. Inform. Theory*, vol. 15, pp. 174–176, Jan. 1969.