EXTREMAL PROBLEMS OF CODING THEORY

Alexander Barg

Bell Labs, Lucent Technologies, 600 Mountain Ave., Rm. 2C-375
Murray Hill, NJ 07974, USA
and IPPI RAN, Moscow, Russia
E-mail: abarg@research.bell-labs.com

This article is concerned with properties of codes as packings of metric spaces. We present a selection of results on extremal problems of geometric coding theory.

CONTENTS

2   *A. Barg*

Notation

$A_w(\mathcal{C}), A_w$	element of the distance distribution of a code $\mathcal{C}$
$A_{\mathcal{C}}(x,y), A(x,y)$	distance enumerator of a code $\mathcal{C}$
$a_\omega(\mathcal{C}), a_\omega$	distance profile
$\sharp(A), \lvert A \rvert$	number of elements of a finite set A
$\mathcal{B}_w(X,x), \mathcal{B}_w(x)$	ball of radius w in X with center at x
B_w	$\mathrm{vol}(\mathcal{B}_w)$
$\mathcal{C}$	code
$\mathcal{C}'$	dual code of a linear code $\mathcal{C}$
$D(a\Vert b)$	information distance between two binomial distributions, Section 2.1
$d(\mathcal{C})$	distance of the code $\mathcal{C}$
$\mathrm{d}(x,y)$	distance between x and y
$\delta_{\mathrm{GV}}(R)$	relative Gilbert-Varshamov distance, Definition 4.3
$\mathcal{H}_q^n$	Hamming space, the n-fold direct product of a q-set
$h_q(x)$	entropy function, Section 2.1
$\mathcal{J}_q^{n,w}$	Johnson space, sphere of radius w about zero in $\mathcal{H}_q^n$
$\mathcal{J}^{n,w}$	binary Johnson space
$K_k(x)$	Krawtchouk polynomial of degree k
$M(X;d)$	maximum size of a d-code in X
$P_{\mathrm{de}}(\mathcal{C})\ [P_{\mathrm{ue}}(\mathcal{C})]$	probability of decoding [undetected] error of the code $\mathcal{C}$
$p_{i,j}^k(\mathcal{H}_q^n)$	intersection number of $\mathcal{H}_q^n$, see Section 2
$\mathcal{S}^n$	unit sphere in $\mathbb{R}^n$
$\mathcal{S}_w(X,x), \mathcal{S}_w(x)$	sphere of radius w in X with center at x
S_w	$\mathrm{vol}(\mathcal{S}_w)$
$\lvert x \rvert$	Hamming norm of x
$\lVert x \rVert$	ℓ_2-norm of x
$\chi(x)$	indicator function of a subset in a set X
$f(n) \cong g(n)$	exponential equivalence of functions, Section 2.1

1. Introduction

This article is devoted to results of coding theory that are centered around the concept of a code as a packing of the corresponding metric space. We derive a few results in several rather diverse directions such as combinato-

rial bounds on codes and their invariants, properties of linear codes, error exponents, and applications of the polynomial method. A common goal of the problems considered is to establish bounds on natural combinatorial parameters of a code. The primary aim of this article is to explain the basic ideas that drive this part of coding theory. In particular, we do not strive to explain the best known result for each problem that we discuss. Our motivation is that, as in each living mathematical discipline, the current best results are often of an *ad hoc* nature and do not add to our understanding of the central ideas. Pointers to the literature that develops the subjects of this article are supposed to compensate for this.

The title of this article refers to estimates of code parameters for codes of large length. This approach helps us to highlight fundamental properties of codes. There is also a vast literature devoted to beating current records for parameters of short codes in various metric spaces; this will not be discussed below (see [41]).

Let X be a metric space. $\mathcal{B}_w(c) = \mathcal{B}_w(X, c)$ denotes the ball and $\mathcal{S}_w(c) = \mathcal{S}_w(X, c)$ the sphere of radius w with center at the point $c \in X$. The volume of a subset $Y \subseteq X$ will be denoted by vol Y (we only deal with the counting measure for finite spaces and the Lebesgue measure in $\mathbb{R}^n$).

2. Metric Spaces

In this section we list the main examples of metric spaces occurring in coding theory.

A. *Hamming space $\mathcal{H}_q^n$.* Let Q be a finite set of size q.

$$\mathcal{H}_q^n = \{(x_1, \ldots, x_n), x_i \in Q\}.$$

Another definition: $\mathcal{H}_q^n = (K_q)^n$, where K_q is a complete graph on q vertices; in this context $\mathcal{H}_q^n$ is also called the Hamming graph.

We denote the elements of Q by $0, 1, \ldots, q - 1$. If q is a power of a prime, then we assume that Q is the finite field $\mathbb{F}_q$, and then $\mathcal{H}_q^n$ is an n-dimensional linear space over Q. We call elements of $\mathcal{H}_q^n$ words, or points, or vectors.

The *Hamming norm* or the Hamming weight of a point $x = (x_1, \ldots, x_n) \in \mathcal{H}_q^n$ is defined as

$$|x| = \sharp\{i : x_i \neq 0\}.$$

The distance induced by this norm is called the *Hamming distance*.

4   *A. Barg*

A ball $\mathcal{B}_w(\mathcal{H}_q^n, c)$ of radius w with center at any point $c \in \mathcal{H}_q^n$ has the volume

$$B_w = \mathrm{vol}(\mathcal{B}_w) = \sum_{i=0}^{w} \binom{n}{i} (q-1)^i.$$

Intersection numbers $p_{i,j}^k$ of the space by definition are

$$p_{i,j}^k(\mathcal{H}_q^n) = \sharp\{z \in \mathcal{H}_q^n : \mathrm{d}(z,x) = i, \mathrm{d}(z,y) = j; \mathrm{d}(x,y) = k\}.$$

Thus, $p_{i,j}^k$ is the number of triangles with fixed vertices x and y, distance k apart, and a floating vertex z that obeys the distance conditions. Explicitly,

$$p_{i,j}^k(\mathcal{H}_q^n) = \sum_{\alpha=0}^{\lfloor \frac{i+j-k}{2} \rfloor} \binom{k}{j-\alpha} \binom{j-\alpha}{k+\alpha-i} \binom{n-k}{\alpha} (q-1)^\alpha (q-2)^{i+j-k-2\alpha}.$$

In particular,

$$p_{ij}^k(\mathcal{H}_2^n) = \binom{k}{(1/2)(j-i+k)} \binom{n-k}{(1/2)(j+i-k)} \chi(j-i+k \in 2\mathbb{Z}).$$

For any two vectors x, y define their *support* as

$$\mathrm{supp}(x, y) = \{i : x_i \neq y_i\}.$$

If $y = 0$, we write $\mathrm{supp}(x)$ instead of $\mathrm{supp}(x, 0)$. If $A \subseteq \mathcal{H}_q^n$, then

$$\mathrm{supp}(A) = \bigcup_{a, a' \in A} \mathrm{supp}(a, a').$$

We note that the Hamming metric is not the only interesting distance on $\mathcal{H}_q^n$. Generally the set Q can support various algebraic structures such as groups and rings. Even for $q = 4$ this already leads to nonequivalent metrics on $\mathcal{H}_q^n$: the Hamming distance and the Lee distance. This diversity increases for larger q; eventually the taxonomy of norms and norm-like functions itself becomes a subject of study.

B. *Johnson space.*

$$\mathcal{J}_q^{n,w} = \{x \in \mathcal{H}_q^n : |x| = w\}.$$

The metric in $\mathcal{J}_q^w$ is the Hamming distance.

If $q = 2$, then we write $\mathcal{J}^{n,w}$ and omit the lower index. In this case it is sometimes more convenient to use the Johnson metric $\mathrm{d}_J = (1/2)\,\mathrm{d}_H$, where d_H is the Hamming distance.

A ball $\mathcal{B}_r(\mathscr{J}^{n,w})$ has the volume

$$B_r = \sum_{i=0}^{\lfloor r/2 \rfloor} \binom{w}{i}\binom{n-w}{i}.$$

C. *Unit sphere in* $\mathbb{R}^n$.

$$\mathscr{S}^n = \{x \in \mathbb{R}^n : \|x\| = 1\}.$$

The distance in $\mathscr{S}^n$ is defined by the angle between the vectors:

$$\theta(x,y) = \arccos(x,y).$$

It is often convenient to use the inner product $t = (x,y)$ instead of θ.

D. *Projective spaces and beyond.* Coding theory is mostly concerned with $X = P^{n-1}\mathbb{R}$, $P^{n-1}\mathbb{C}$, and $P^{n-1}\mathbb{H}$. The distance between $x,y \in X$ is measured by the angle $\theta = \arccos|(x,y)|$ or by the absolute value of the inner product $t = |(x,y)|$.

One generalization of these projective spaces has recently gained attention in coding theory. Let $X = G_{n,k}(L)$ be a Grassmann space, i.e., the manifold of k-planes ($k < n/2$) through the origin in the n-space over L (here $L = \mathbb{R}$ or $\mathbb{C}$). To define the distance between two planes we need to introduce principal angles. Let p and q be two planes in X. The absolute value of the inner product $|(x,y)|, x \in p, y \in q$, as a function of $2k$ variables has k stationary points $\rho_1, \ldots, \rho_k$. Define the principal angles $\theta_1, \ldots, \theta_k$ between p and q by their cosines: $\theta_i = \arccos \rho_i$. There are several justifiable ways of defining the distance between p and q. So far the following definition received most attention in coding theory:

$$\mathrm{d}(p,q) = \sqrt{\sin^2 \theta_1 + \cdots + \sin^2 \theta_k}.$$

2.1. *Asymptotics*

We next turn to asymptotic formulas for sphere volume in metric spaces of interest to coding theorists. They will be used to derive volume bounds on codes in Section 6.

Let $f(n)$ and $g(n)$ be two functions of $n \in \mathbb{N}$. We write $f \cong g$ if $\lim_{n\to\infty} \frac{1}{n} \log \frac{f(n)}{g(n)} = 1$. The base of the logarithms and exponents is 2 throughout.

6   *A. Barg*

2.1.1. *Hamming space.* Let $X = \mathscr{H}_q^n$, where q is fixed, $n \to \infty$ and let $w = \omega n$. We have, for $\omega, p \in (0, (q-1)/q)$,

$$\sum_{i=0}^{w} \binom{n}{i}(q-1)^i \left(\frac{p}{q-1}\right)^i (1-p)^{n-i} \cong \exp[-nD(\omega\|p)],$$

where the information divergence between two binomial distributions, $D(\omega\|p)$, equals

$$D(\omega\|p) = \omega \log \frac{\omega}{p} + (1-\omega) \log \frac{1-\omega}{1-p}.$$

In particular, with $p = (q-1)/q$ we obtain an asymptotic formula for the volume of the ball:

$$B_{\omega n} \cong \exp[nh_q(\omega)],$$

where $h_q(y)$ is the entropy function defined by

$$h_q(y) = -y \log \frac{y}{q-1} - (1-y) \log(1-y)$$

for $y \in (0,1)$ and extended by continuity to $y = 0, y = 1$.

2.1.2. *Binary Johnson space* $\mathscr{J}^{n,w}$. Let $w = \omega n$. The volume of the ball is given by

$$B_{\rho n} \cong \exp\left[n(\omega h_2\left(\frac{\rho}{2\omega}\right) + (1-\omega)h_2\left(\frac{\rho}{2(1-\omega)}\right)\right].$$

2.1.3. *The sphere* $\mathscr{S}^n$. A ball in $X = \mathscr{S}^n$,

$$\mathcal{B}_\theta(X;x) = \{y \in X : \angle(x,y) \le \theta\},$$

is the spherical cap cut on the surface of X by the circular cone $\mathrm{Con}\,(x,\theta)$ with apex at the origin and axis along x. Let $\Omega(\theta) = \mathrm{vol}(\mathcal{B}_\theta(X;x))$. We have

$$n^{-1} \log \Omega(\theta) = \frac{1}{2} \log \frac{2e\pi \sin^2\theta}{n}(1 + o(1)) \quad (0 \le \theta \le \pi/2).$$

2.1.4. *Grassmann space* $G_{n,k}(L)$. Let $\mathcal{B}_\delta(G_{n,k})$ be a ball in the Grassmann manifold of radius δ with respect to the distance $\mathrm{d}(p,q)$. The volume of the ball of radius δ is given by

$$B_\delta = \left(\frac{\delta}{\sqrt{k}}\right)^{\beta nk + o(n)} \quad (\beta = 1 \text{ if } L = \mathbb{R};\ \beta = 2 \text{ if } L = \mathbb{C}).$$

All the results of Section 2.1 except the last one are standard. The volume of the ball in $G_{n,k}$ is computed in [10] (see [17] for a discussion of sphere packings in $G_{n,k}$).

3. Codes

Let X be a finite or compact infinite metric space. A code $\mathcal{C}$ is a finite subset of X. The *distance* of the code is defined as $d(\mathcal{C}) = \min_{x,y \in \mathcal{C}; x \neq y} d(x,y)$.

Let $M = |\mathcal{C}|$ be the size of (i.e., the number of points in) the code. The rate of the code, measured in bits, is

$$R(\mathcal{C}) = n^{-1} \log M,$$

where n is the dimension of X, clear from the context. The *relative distance* of $\mathcal{C}$ is defined as

$$\delta(\mathcal{C}) = \frac{d(\mathcal{C})}{n}.$$

The argument $\mathcal{C}$ will often be omitted.

A code $\mathcal{C} \subseteq \mathscr{H}_q^n$ of size M and distance d is denoted by $\mathcal{C}(n, M, d)$. If the distance of a code $\mathcal{C}$ is d, we sometimes call it a d-code.

The distance between a point $x \in X$ and a subset $Y \subseteq X$ is defined as

$$\mathrm{d}(x, Y) = \min_{y \in Y} \mathrm{d}(x, y).$$

3.1. *Distance Distribution*

Let X be a finite space of diameter D.

Definition 3.1: The *distance distribution* of a code $\mathcal{C} \subseteq X$ is the vector $A = (A_0, A_1, \ldots, A_D)$, where

$$A_i = |\mathcal{C}|^{-1} \sharp \{(x,y) \in \mathcal{C} \times \mathcal{C} : \mathrm{d}(x,y) = i\}.$$

Thus $A_0 = 1$.

Let $X = \mathscr{H}_q^n$. If $\mathcal{C}$ is a linear code in $\mathscr{H}_q^n$ (where q is a prime power), then its distance distribution is equal to the weight distribution $(A_0, A_d, \ldots, A_n)$, where

$$A_i = \sharp \{x \in \mathcal{C} : |x| = i\}.$$

Let $w = \omega n$ and let $\alpha_\omega(\mathcal{C}) = n^{-1} \log A_{\omega n}$. The $(n+1)$-vector $[\alpha_\omega(\mathcal{C}), \omega = n^{-1}(0, 1, \ldots, n)]$ is called the *distance (weight)* profile of $\mathcal{C}$. The main use

8   *A. Barg*

of the distance profile is in asymptotic problems, where it is represented by
a real (usually, continuous) function.

The *distance enumerator* of a code $\mathcal{C} \subseteq X$ is the polynomial

$$A_{\mathcal{C}}(x, y) = \sum_{i=0}^{n} A_i x^{n-i} y^i.$$

For codes in infinite spaces we use a slightly more convenient defini-
tion of the distance distribution. For instance, let $\mathcal{C} \subseteq \mathscr{S}^n$. The distance
distribution of $\mathcal{C}$ is given by

$$b(s, t) = |\mathcal{C}|^{-1} \, \sharp\{(x, y) \in \mathcal{C} \times \mathcal{C} : s \le (x, y) < t\}.$$

We have $|\mathcal{C}| = \int_{-1}^{1} db(x)$, where $db(x)$ is a discrete measure defined from
$b(s, t)$ in a standard way.

The main applications of the distance distribution are in combinatorial
properties of codes, bounds on error probability of decoding (Sections 9,
10), and other extremal problems of coding theory (*e.g.*, Section 16).

Definition 3.2: The *covering radius* of a code $\mathcal{C} \in X$ is defined as

$$r(\mathcal{C}) = \max_{x \in X} \mathrm{d}(x, \mathcal{C}).$$

3.2. *Asymptotic Parameters*

Let X be one of the metric spaces introduced above. Let

$$M(X; d) = \max_{\mathcal{C} \in X, d(\mathcal{C}) = d} |\mathcal{C}|;$$

$$R(\delta) = R(X; \delta) = \lim_{n \to \infty} [n^{-1} \log M(X; d)].$$

In the cases of interest to coding theory, this limit is not known to exist
Therefore, it is usually replaced by $\limsup$ and $\liminf$ as appropriate, and
the corresponding quantities are denoted by $\overline{R}(\delta), \underline{R}(\delta)$. We write X in the
notation only if the underlying space is not clear by the context.

A notational convention: $R(\mathscr{H}_q^n; \delta)$, for instance, means the highest
achievable rate of a sequence of codes of relative distance δ in the Hamming
space; here $\mathscr{H}_q^n$ is used as a notation symbol in which n has no particular
value. This agreement is used throughout the text.

Analogously,

$$\delta(R) = \delta(X; R) = \lim_{n \to \infty} \max_{\mathcal{C} \in X; R(\mathcal{C}) \ge R} d(\mathcal{C})/n.$$

4. Average Properties of Codes

This section is concerned with estimates of codes' parameters obtained by various averaging arguments. In many cases, the existence bounds thus obtained are the best known for large code length. We will establish the average rate, distance, and the distance distribution of unrestricted codes and linear codes in the Hamming space.

Theorem 4.1: *Let $X = \mathscr{H}_q^n$. Let M be such that*

$$M(M-1) < \frac{q^n}{B_{d-1}}.$$

Then there exists an (n, M, d) code.

Proof: Let $\mathcal{C} = \{x_1, \ldots, x_M\}$ be an ordered collection of points. Call $\mathcal{C}$ bad if $d(\mathcal{C}) \leq d-1$ and call a point $x_i \in \mathcal{C}$ bad if it has neighbors in $\mathcal{C}$ at distance $\leq d-1$. If the points $x_2, \ldots, x_M$ are fixed, then x_1 is bad in at most $(M-1)B_{d-1}$ codes. The points $x_2, \ldots, x_M$ can be chosen in $q^{n(M-1)}$ ways, so there are no more than $(M-1)B_{d-1}q^{n(M-1)}$ codes in which point x_1 is bad. This is true for any point $x_i, 1 \leq i \leq M$; thus, there are no more than $M(M-1)B_{d-1}q^{n(M-1)}$ bad codes. If this number is less than the total number of codes q^{nM}, i.e., if

$$M(M-1)B_{d-1} < q^n,$$

then there exists a good code.   □

This result can be improved by the so-called Gilbert procedure (Section 6). However, for large n, Theorem 4.1 accurately describes the parameters of typical codes in $\mathscr{H}_q^n$. More formally, we have the following result.

Theorem 4.2: *Let $X = \mathscr{H}_2^n$ and $n \to \infty$. For all codes in X of rate R except for a fraction of codes that decreases exponentially with n, the relative distance approaches the bound*

$$2R = 1 - h_2(\delta).$$

Proof: Consider the Shannon ensemble $\mathcal{A}$ of 2^{nM} binary codes, $M = 2^{Rn}$, where every code has probability 2^{-nM}. Or, what is the same, consider a random code formed of M independently chosen vectors, where all the

coordinates of every vector are i.i.d. Bernoulli r.v.'s with $P(0) = P(1) = 1/2$.

Let us assume that δ is chosen to satisfy $2R = 1 - h_2(\delta) + \varepsilon$, where $\varepsilon > 0$. We will prove that with probability approaching 1 a random (n, M) code $\mathcal{C} \in \mathcal{A}$ contains a pair of vectors at distance δn or less. Let $x_1, x_2, \ldots, x_M$ be an ordered (multi)set of independent random vectors such that $\Pr[x_i = y] = 2^{-n}$ for any $y \in \{0,1\}^n$. Let $\nu_{i,j}, 1 < j < i < M$, be the indicator random variable of the event $\mathrm{d}(x_i, x_j) = \delta n$. The $\nu_{i,j}$ are pairwise-independent random variables, each with mean

$$\mathsf{E}\,\nu_{i,j} = \Pr[\nu_{i,j} = 1]$$

and variance

$$\mathrm{Var}[\nu_{i,j}] = \mathsf{E}\,\nu_{i,j}^2 - (\mathsf{E}\,\nu_{i,j})^2 = \mathsf{E}\,\nu_{i,j} - (\mathsf{E}\,\nu_{i,j})^2 < \mathsf{E}\,\nu_{i,j}.$$

Consider the number $N_{\mathcal{C}}(d) = \sum_{j<i} \nu_{i,j}$ of unordered pairs of codewords (x_i, x_j) with $i \neq j$ in $\mathcal{C}$ at distance $d = \delta n$ apart. We have

$$\mathsf{E}\,N_{\mathcal{C}}(d) = \binom{M}{2} \mathsf{E}\,\nu_{i,j} \cong 2^{n(2R-1+h_2(\delta))}$$

$$\mathrm{Var}[N_{\mathcal{C}}(d)] = \binom{M}{2} \mathrm{Var}[\nu_{i,j}] < \mathsf{E}\,N_{\mathcal{C}}(d).$$

For any $\alpha > 0$ by the Chebyshev inequality we have

$$\Pr[|N_{\mathcal{C}}(d) - \mathsf{E}\,N_{\mathcal{C}}(d)| \geq \mathsf{E}N_{\mathcal{C}}(d)^{(1+\alpha)/2}] \leq (\mathsf{E}\,N_{\mathcal{C}}(d))^\alpha$$
$$\cong 2^{\alpha n(1-2R-h_2(\delta))} = 2^{-\alpha n \varepsilon} \to 0.$$

Thus, in particular, with probability tending to 1 we have $N_{\mathcal{C}}(d) > 0$, or, in other words, a random code contains a pair of vectors at distance $d = \delta n$. Since $\varepsilon > 0$ can be taken arbitrarily small, this proves an upper bound $\delta \leq h_2^{-1}(1 - 2R)$ on the relative distance of almost all codes in $\mathcal{A}$.

On the other hand, for any δ such that $2R = 1 - h_2(\delta) - \varepsilon$ the average number of codeword pairs with relative distance $d = \delta n$ decreases exponentially with n. Then

$$\Pr[N_{\mathcal{C}}(d) > 1] \leq \mathsf{E}\,N_{\mathcal{C}}(d) \to 0;$$

hence with probability tending to 1 a random code $\mathcal{C}$ has distance $\geq \delta n$. $\square$

This theorem implies that for $R > 1/2$ the relative distance of almost all long codes converges to zero. Thus, unrestricted codes on the average are much worse than linear codes (cf. Theorem 4.4 below).

Definition 4.3: The *relative Gilbert-Varshamov distance* $\delta_{\mathrm{GV}}(R)$ is defined by the equation

$$R = \log q - h_q(\delta) \qquad (0 \le \delta \le 1 - 1/q).$$

Theorem 4.4: *Let $X = \mathscr{H}_q^n$ and $n \to \infty$. For all linear codes in X of rate R except for a fraction of codes that decreases exponentially with n, the relative distance approaches $\delta_{\mathrm{GV}}(R)$.*

Proof: (outline) Consider the ensemble $\mathcal{L}$ of random $[n, k = Rn]$ linear binary codes defined by $(n - k) \times n$ parity-check matrices whose elements are chosen independently with $P(0) = P(1) = 1/2$. If N_w is the random variable equal to the number of vectors of weight $w > 0$ in a code $\mathcal{C} \in \mathcal{L}$, then $\mathsf{E}\, N_w = \binom{n}{w}(q - 1)^w / q^{n-k}$ and $\operatorname{Var} N_w \le \mathsf{E}\, N_w$. Thus, $\mathsf{E}\, N_w$ grows exponentially in n for $\omega := \frac{w}{n} > \delta_{\mathrm{GV}}(R)$. Thus, the relative distance of a random linear code approaches $\delta_{\mathrm{GV}}(R)$ as n grows, and the fraction of codes whose relative distance deviates from δ_{GV} by ε tends to 0 exponentially in n for any $\varepsilon > 0$. $\qquad\square$

Theorem 4.5: *There exists a linear $[n, k]$ code $\mathcal{C}$ with $A_0 = 1$,*

$$A_w(\mathcal{C}) \le \begin{cases} n^2 q^{k-n} S_w, & w \text{ such that } \log S_w \ge (n - k)\log q - 2\log n, \\ 0, & w : \log S_w < (n - k)\log q - 2\log n. \end{cases}$$

Proof: Consider linear codes defined in the same way as in the proof of Theorem 4.4. A vector of weight $w > 0$ lies in the kernel of $q^{(n-1)(n-k)}$ matrices. All the $S_w = \binom{n}{w}(q - 1)^w$ vectors of weight w are annihilated by at most $S_w q^{(n-1)(n-k)}$ matrices. Thus, on the average the number of vectors of weight w in the code does not exceed $S_w q^{-(n-k)}$ and the fraction of matrices for which this number is $\ge n^2 S_w q^{-(n-k)}$ (call them bad) is at most n^{-2}. Even if the sets of bad matrices for different $w = 1, 2, \ldots, n$ are disjoint, this leaves us with a fraction of $1 - n^{-1}$ of good matrices; any good matrix defines a code $\mathcal{C}$ of dimension $\dim \mathcal{C} \ge k$ over $\mathbb{F}_q$ with

$$A_w(\mathcal{C}) \le n^2 q^{k-n} S_w, \quad 1 \le w \le n.$$

Writing the right-hand side as $\exp[(k - n)\log q + \log S_w + 2\log n]$, we see that once w is such that the exponent becomes negative, we obtain $A_w < 1$. Since $\mathcal{C}$ is linear, this implies that $A_w = 0$ for these values of w. $\qquad\square$

Corollary 4.6: *For any $R < \log q - h_q(\delta)$ there exists a sequence of linear codes of growing length n with weight profile α_0, where $\alpha_{0,0} = 0$,*

$$\alpha_{0,\omega} \le R - \log q + h_q(\omega) \quad (\delta_{GV}(R) < \omega < 1 - \delta_{GV}(R)),$$
$$\alpha_{0,\omega} = -\infty \quad\quad\quad\quad\quad (0 < \omega < \delta_{GV}(R)).$$

Linear codes that satisfy Theorem 4.5 or Corollary 4.6 will be called *random*.

Theorem 4.7: (*Average value of the distance*) *Let $\mathcal{C}$ be an (n, M) code. Then, provided in each case that the denominator is positive,*

$$M \le \frac{d}{d - \frac{q-1}{q}n} \quad\quad\quad\quad \mathcal{C} \subseteq \mathcal{H}_q^n,\ d = d(\mathcal{C}),$$

$$M \le \frac{nd}{nd - 2wn + \frac{q}{q-1}w^2} \quad\quad \mathcal{C} \subseteq \mathcal{J}_q^{n,w},\ d = d(\mathcal{C}),$$

$$M \le \frac{1-t}{-t} \quad\quad\quad\quad\quad\quad \mathcal{C} \subseteq \mathcal{S}^n,\ t = t(\mathcal{C}).$$

Here $t(\mathcal{C})$ is the maximal inner product of the code $\mathcal{C} \subseteq \mathcal{S}^n$.

This result is called the *Plotkin* bound for $\mathcal{H}_q^n$, the *Johnson* bound for $\mathcal{J}_q^{w,n}$, and the *Rankin* bound for $\mathcal{S}^n$. It is proved by computing the average distance between pairs of points in $\mathcal{C}$ [18], [38].

Thus, for large values of the code distance $d(\mathcal{C})$ the value of M cannot grow exponentially with n, and so for any family of codes $\mathcal{C}$ the rate $R \to 0$. For instance, for $X = \mathcal{H}_q^n$ the code size M is at most $O(n)$ if $\delta = d(\mathcal{C})/n > \frac{q-1}{q}$. Below in asymptotic problems we always assume the reverse inequality.

Already for general unrestricted codes the technique presented in this section produces weak results. In more complicated problems of coding theory one resorts to more refined averaging methods, such as averaging over the choice of subsets rather than individual vectors, etc. [11].

5. Averaging over Translations

This section presents another averaging technique which is useful for deriving upper bounds on code parameters and linking the Hamming and Johnson spaces.

Lemma 5.1: *Let x, y be two vectors in $\mathcal{H}_q^n$ with $\mathrm{d}(x,y) = u$. The number of vectors $z \in \mathcal{H}_q^n$ such that $x - z \in \mathcal{J}_q^{n,w}$ and $y - z \in \mathcal{J}_q^{n,w}$ equals $p_{ww}^u(\mathcal{H}_q^n)$.*

Proof: Obvious.   □

Lemma 5.2: [34] *Let* $\mathcal{C} \subseteq X \subseteq \mathscr{H}_q^n$ *be a code and* $Y, Z \subseteq \mathscr{H}_q^n$ *be arbitrary subsets. Then*

$$\sum_{c \in \mathcal{C}} |(Y - c) \cap Z| = \sum_{z \in Z} |(\mathcal{C} + z) \cap Y|.$$

Proof:

$$\sum_{c \in \mathcal{C}} |(Y - c) \cap Z| = \sum_{c \in \mathcal{C}} \sum_{y \in Y} \sum_{z \in Z} \chi\{y - c = z\} = \sum_{z \in Z} \sum_{y \in Y} \sum_{c \in \mathcal{C}} \chi\{c + z = y\}$$
$$= \sum_{z \in Z} |(\mathcal{C} + z) \cap Y|.$$
                                                                                                       □

Corollary 5.3: *Let* $\mathcal{C} \subseteq \mathscr{J}_q^{\,n,v}$ *be a* d*-code. Then*

$$|\mathcal{C}|p_{u,w}^v \leq S_u M(\mathscr{J}_q^{\,n,w}; d).$$

Proof: In Lemma 5.2 take $X = \mathcal{S}_v(0)$, $Y = \mathcal{S}_w(0)$, $Z = \mathcal{S}_u(0)$. Let $y \in Y, c \in \mathcal{C}$, then $y - c \in Z$ if and only if $\mathrm{d}(y, c) = u$. The number of $y \in Y$ with this property for a fixed c equals $p_{u,w}^v$. On the right-hand side we observe that the set $(\mathcal{C} + z) \cap Y$ is a d-code in Y.   □

Corollary 5.4: *Let* $\mathcal{C}, Y$ *be subsets of* $\mathscr{H}_q^n$*. Then*

$$|Y||\mathcal{C}| = \sum_{z \in \mathscr{H}_q^n} |(\mathcal{C} + z) \cap Y|.$$

Proof: Follows by putting $X = Z = \mathscr{H}_q^n$ in Lemma 5.2.   □

Lemma 5.5: [36] *Let* $\mathcal{C}$ *be code in* $\mathscr{H}_q^n$*. Then*

$$|\mathcal{C}|A_i(\mathcal{C})p_{w,w}^i = \sum_{x \in \mathscr{H}_q^n} |\mathcal{C}(x,w)|A_i(\mathcal{C}(x,w)),$$

where $p_{w,w}^i$ *is the intersection number and* $\mathcal{C}(x,w) = (\mathcal{C} + x) \cap \mathscr{J}_q^{\,n,w}$*.*

Proof: Count in two ways the total number of pairs of codewords in $\mathcal{C}(x,w)$ distance i apart for all $x \in \mathscr{H}_q^n$. By definition, this is the right-hand side of the claimed identity. On the other hand, every pair of codewords in $\mathcal{C}$ at a distance i falls in $\mathscr{J}_q^{\,n,w}$ in $p_{w,w}^i$ shifts of $\mathcal{C}$ (Lemma 5.1).   □

14   *A. Barg*

Any of last two results implies the well-known *Bassalygo-Elias inequality*:

$$M(\mathscr{H}_q^n; d)\binom{n}{w}(q-1)^w \leq M(\mathscr{J}_q^{n,w}; d)q^n \qquad (5.1)$$

(take $i = 0$ in Lemma 5.5 or take $Y = \mathscr{J}_q^{n,w}$ in Corollary 5.4).

Theorem 5.6: (*Elias-type bounds*)

$$R(\mathscr{H}_q^n; \delta) \leq \log q - h_q\big(\lambda(1 - \sqrt{1 - \delta/\lambda}\,)\big) \quad (\lambda = 1 - q^{-1}),$$
$$R(\mathscr{J}^{n,\omega n}; \delta) \leq h_2(\omega) - h_2\Big(\frac{1}{2} - \frac{1}{2}\sqrt{1 - 2\delta}\Big),$$
$$R(\mathscr{S}^n; \theta) \leq -\log(\sqrt{2}\sin(\theta/2)).$$

Proof: Consider the Hamming case. From Theorem 4.7 for the Johnson space we see that when $w = \omega n$ approaches the (smaller) root of the denominator,

$$\omega_0 = \lambda - \sqrt{\lambda(\lambda - \delta)},$$

the quantity $n^{-1}\log M(\mathscr{J}_q^{n,w}; d, w) \to 0$. Substituting ω_0 into inequality (5.1) and computing logarithms completes the proof. The other two cases follow by some modifications of this argument.   □

Solving the inequality of the theorem for Hamming space with respect to δ, we obtain the bound

$$\delta(R) \leq \delta_{\mathrm{E}}(R),$$

where

$$\delta_{\mathrm{E}}(R) := 2\delta_{\mathrm{GV}}(R)(1 - \delta_{\mathrm{GV}}(R)/2\lambda)$$

is sometimes called the *Elias radius*. The bound itself is called the Bassalygo-Elias bound (the Hamming case) and the Coxeter-Rankin bound (the spherical case). For the Johnson space the result can be proved analogously to the Hamming case; see [33], where a nonasymptotic version of this bound is also derived. In Section 6.2 we give a proof based on a volume argument.

6. Volume Bounds

This section is devoted to the standard bounds proved via a volume, or packing argument. We begin with the standard Gilbert-Varshamov and Hamming bounds, which basically say that if the spheres of radius d are disjoint, then their centers form a d-code, and that the size of a d-code is bounded above by the number of spheres of radius $d/2$ that can be packed into X. The second part of Section 6.1 deals with an improvement of the Hamming bound for the Johnson space. The ideas developed there will also be central in the derivation of error exponents in Section 9.

6.1. *Basic Volume Bounds*

Let X be one of the metric spaces discussed above with distance d and volume form vol. Let $B_d = \mathrm{vol}(\mathcal{B}_d)$ be the volume of the ball of radius d in X.

Theorem 6.1: (*Gilbert bound*) *If M is any number such that $MB_d < \mathrm{vol}(X)$, then X contains a code of size $M+1$ and distance d. If the distance d takes only natural values, then d can be replaced with $d-1$.*

For $\mathscr{H}_q^n$ this bound was given in Theorem 4.4 and Corollary 4.6.

Theorem 6.2: (*Shannon bound* [45]) *For any*

$$R < -\log \sin \theta$$

there exists a number n_0 such that for every $n \geq n_0$ the sphere $\mathscr{S}^n$ contains a code of size 2^{Rn} and distance θ.

Proof: Follows on substituting the volume of the spherical cap from Section 2.1.3 into the Gilbert bound. $\qquad\qquad\Box$

Theorem 6.3: (*Hamming bound*) *Let $\mathcal{C}$ be a code of size M in X. Then $M \leq \mathrm{vol}(X)/B_{d/2}$.*

Concrete expressions of this bound for various metric spaces can be obtained using the formulas of Section 2.1. The Hamming bound is usually good for small values of d and weak otherwise. It can be improved for $\mathscr{J}^{n,w}$ by making use of the embedding $\mathscr{J}^{n,w} \subseteq \mathscr{H}_2^n$. The idea is to notice [12] that for some v, a ball of radius $d/2$ with center on $\mathcal{S}_w(\mathscr{H}_2^n, 0) = \mathscr{J}^{n,w}$

A. Barg

intersects the sphere $\mathcal{S}_v(\mathcal{H}_2^n, 0)$ by a larger subset than it intersects the sphere $\mathcal{S}_w$ itself.
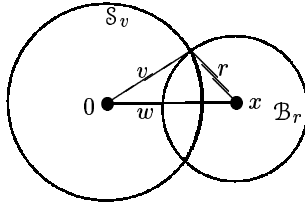
$\mathcal{B}_r(x)$ intersects $\mathcal{S}_v(0)$ by a large subset

Lemma 6.4: *Let $w = \omega n \le n/2, r = \rho n \le v$. Then*

$$\sum_{x \in \mathcal{S}_w(0)} |\mathcal{B}_r(x) \cap \mathcal{S}_v(0)| = \binom{n}{w} \sum_{i=w-v}^{r} p_{i,v}^w \le r \binom{n}{r} \binom{n}{w - j_0},$$

where $j_0/n \to \gamma_0 = \rho \frac{1-2\omega}{1-2\rho}$. This inequality is asymptotically tight, i.e., for $w - v \sim j_0$ it turns into an equality in the $\cong$ sense.

Corollary 6.5: *Let*

$$N(w,r) = \sum_{x \in \mathcal{S}_w(0)} |\mathcal{B}_r(x) \cap \mathcal{S}_r(0)|.$$

Then $N(w,r) \le r \binom{n}{r}^2 (1 + o(1))$ with equality $N(w,r) \cong \binom{n}{r}^2$ only for $w/n \sim 2\rho(1 - \rho)$.

Proof: (of Lemma 6.4 and Corollary 6.5) Let $V = \sum_{\substack{x \in \mathcal{H}_2^n \\ |x|=w}} |\mathcal{B}_r(x) \cap \mathcal{S}_v(0)|$.

Let $j = w - v$. The equality

$$V = \binom{n}{w} \sum_{i=j}^{r} p_{i,v}^w$$

follows by definition. To prove the inequality, observe that

$$p_{r,v}^w \le \sum_{i=j}^{r} p_{i,v}^w = \sum_{i=j}^{r} \binom{w}{\frac{1}{2}(i+j)} \binom{n-w}{\frac{1}{2}(i-j)}.$$

As is easily verified, the summation term in the last sum grows on i. Substituting $i = r$, we obtain

$$\sum_{i=j}^{r} p_{i,v}^{w} \leq r\binom{w}{\frac{1}{2}(r+j)}\binom{n-w}{\frac{1}{2}(r-j)} = rp_{r,v}^{w}.$$

Therefore, $\binom{n}{w}p_{r,v} \leq V \leq r\binom{n}{w}p_{r,v}^{w}$.

Further,

$$\binom{n}{w}p_{r,v}^{w} = \binom{n}{r}p_{v,w}^{r} = \binom{n}{r}\binom{r}{\frac{1}{2}(r+j)}\binom{n-r}{w-\frac{1}{2}(r+j)}.$$

In the last expression, for a fixed vector y of weight r, we are counting vectors x of weight w with $d(x,y) = n(\omega - \gamma)$, where $\gamma = j/n$. Their number is maximized if x is a typical vector obtained after n independent drawings from the binomial probability distribution given by $P(1) = \omega - \gamma, P(0) = 1 - \omega + \gamma$. We obtain the following condition on the maximizing value of γ:

$$\rho - \frac{1}{2}(\rho + \gamma) = \rho(\omega - \gamma).$$

In other words, the maximizing value of j is attained for $j/n \to \gamma_0$. Note that, at least for large n, $j_0 = \gamma_0 n$ satisfies the condition on $j = w - v$ implied by the restriction on v in the Lemma, so the choice $j = j_0$ is consistent. Thus,

$$\max_{j:j\leq r}\binom{r}{\frac{1}{2}(r+j)}\binom{n-r}{w-\frac{1}{2}(r+j)}$$

$$\cong \exp\left[n\left(\rho h_2\left(\frac{\rho + \gamma_0}{2\rho}\right) + (1-\rho)h_2\left(\frac{2\omega - \rho - \gamma_0}{1-\rho}\right)\right)\right]$$

$$\cong \exp[h_2(\omega - \gamma_0)] \cong \binom{n}{w-j_0},$$

where the last line follows by substituting the value of γ_0 and simplifying.

The corollary follows on substituting $v = r$ into the lemma. $\qquad\square$

Theorem 6.6: *Let* $n \to \infty, w/n \to \omega, 0 < \omega < 1/2, \omega > \delta > 0$. *Then*

$$M(\mathscr{J}^{n,w}; \delta n) \leq a(n)\frac{\mathrm{vol}(\mathcal{S}_w)}{\mathrm{vol}(\mathcal{B}_{\delta n/2})}, \qquad (6.1)$$

where $n^{-1}\log a(n) \to 0$ *as* $n \to \infty$. *Therefore,*

$$R(\mathscr{J}^{n,\omega n}; \delta) \leq h_2(\omega) - h_2(\delta/2).$$

Proof: Let $\mathcal{C} \subseteq \mathscr{J}^{n,\omega n}$ be an (n, M, d) code and $r = \lfloor (d-1)/2 \rfloor$. We have

$$|\mathcal{B}_r(c) \cap \mathcal{S}_v(0)| = \sum_{i=w-v}^{r} p_{i,v}^w.$$

Also $\mathcal{B}_r(c) \cap \mathcal{B}_r(c') = \emptyset$ for two different codewords c and c'. Therefore for any $v, w - r \le v \le w + r$,

$$M \le \frac{\mathrm{vol}(\mathcal{S}_v)}{\displaystyle\sum_{i=w-v}^{r} p_{i,v}^w}.$$

Take $v = \frac{\omega - \delta/2}{1-\delta} n$. Then using Lemma 6.4, we obtain

$$M \le \frac{\binom{n}{w} S_v}{\binom{n}{w} \displaystyle\sum_{i=w-v}^{r} p_{i,v}^w} \cong \frac{\binom{n}{w} S_v}{\binom{n}{r}\binom{n}{v}} = \frac{\binom{n}{w}}{\binom{n}{r}}.$$

$\square$

The denominator in the estimate (6.1) is an exponentially greater quantity than the volume computed in 2.1.2; hence the estimate itself is asymptotically better than the Hamming bound. In particular, by Theorem 6.6, $R(\mathscr{J}^{n,\omega n}; \delta) = 0$ for $\delta \ge 2\omega$ while the Hamming bound implies this conclusion only for $\delta \ge 4\omega(1-\omega)$. The actual answer is

$$R(\mathscr{J}^{n,\omega n}; \delta) \begin{cases} > 0 & 0 \le \delta < 2\omega(1-\omega) \\ = 0 & \delta \ge 2\omega(1-\omega) \end{cases}$$

by combining the Gilbert-Varshamov and Elias bounds.

6.2. *Elias-Type Bounds as Sphere-Packing Bounds*

The Hamming bound is not the best bound obtainable by the volume argument for large code length. Namely, suppose that the balls of radius $r > d/2$ around the codewords intersect, but the intersection volume grows as a polynomial function $p(n)$ of the code length. Then

$$|\mathcal{C}| B_r \le \mathrm{vol}(X) p(n).$$

Letting $n \to \infty$, we obtain a bound better than the Hamming bound. For instance, let $\mathcal{C}$ be an $(n, M = q^{Rn}, \delta n)$ code in $\mathscr{H}_q^n$ and let $\mathcal{B}_{\omega n}$ be a ball of radius ωn. By a slight modification of Theorem 4.7 we conclude that for $\omega < \omega_{\mathrm{crit}} := \lambda - \sqrt{\lambda(\lambda - \delta)}$ the number of codewords inside the ball grows

at most polynomially in n. Therefore, for any codeword c a point $x \in \mathscr{H}_q^n$ with $\mathrm{d}(x, c) < n\omega_{\mathrm{crit}}$ can be distance $n\omega_{\mathrm{crit}}$ or less away from at most $p(n)$ codewords, where $p(n)$ is some polynomial. We then have

$$\frac{1}{n} \log(MB_{\omega n}) = R - \log q + h_q(\omega_{\mathrm{crit}}) + o(1) < (\log p(n))/n,$$

which again proves Theorem 5.6 for $X = \mathscr{H}_q^n$. Other parts of this theorem can be proved by a similar argument.

7. Linear Codes

This section deals with combinatorial and linear-algebraic properties of codes. The technique used here is based on an interplay of the rank distributions and weight distributions of linear codes. Readers familiar with matroids will immediately notice a connection with representable matroids and their invariants.

Let q be a prime power. A linear $[n, k, d]$ code $\mathcal{C}$ is a subspace of $\mathbb{F}_q^n$ of dimension k and distance d. A matrix $\mathbf{G}$ whose rows form a basis of $\mathcal{C}$ is called the *generator matrix* of the code. The linear $[n, n - k, d']$ code $\mathcal{C}' = \{x : \forall_{c \in \mathcal{C}}(c, x) = 0\}$ is called the dual code of $\mathcal{C}$. The generator matrix $\mathbf{H}$ of $\mathcal{C}'$ is called the parity-check matrix of $\mathcal{C}$. For any matrix $\mathbf{G}$ with n columns and a subset $E \subseteq \{1, 2, \ldots, n\}$ we denote by $\mathbf{G}(E)$ the subset of columns of $\mathbf{G}$ indexed by the elements of E.

The goal of this section is to derive in a simple way some combinatorial identities related to the famous MacWilliams theorem.

Definition 7.1: *Puncturings and shortenings.* Let $\mathcal{C}$ be an $[n, k, d]$ code. Puncturing it results in an $[n - 1, k, d - 1]$ code. More generally, let $E \subseteq \{1, 2, \ldots, n\}, |E| = n - t$. The projection $\mathcal{C}_E = \mathrm{proj}_E \mathcal{C}$ is a linear subcode of $\mathcal{C}$ of length $n - t$ and dimension equal to $\mathrm{rk}\, \mathbf{G}(E)$.

A 1-shortening of $\mathcal{C}$ on coordinate i is formed of $|\mathcal{C}|/q$ codewords $c \in \mathcal{C}$ such that $c_i = 0$; this is a linear $[n, k - 1, d]$ subcode. Successively applying this operation, we obtain a t-shortening of $\mathcal{C}$ on the coordinates in $\{1, 2, \ldots, n\} \setminus E$, where E is some t-subset. This is a linear subcode $\mathcal{C}^E \subseteq \mathcal{C}$ such that $\mathrm{supp}\, \mathcal{C}^E \subseteq E$.

7.1. *Shortening of Unrestricted Codes*

Before proceeding further, we give an application of shortenings to properties of unrestricted (i.e., linear or not) codes. In general, for an (n, M, d) code $C \subseteq \mathscr{H}_q^n$, shortening is defined as follows: out of the M codewords at least M/q coincide in a given coordinate i. Consider the code formed of these codewords with the ith coordinate deleted. This gives an $(n - 1, \geq M/q, d)$ code. Iterating, we get

Lemma 7.2: *Let $\mathcal{C}$ be an (n, M, d) code. For any $t \leq n - d$, we have*
$$M \leq q^t M(\mathscr{H}_q^{n-t}, d).$$

Proof: Let $E \subseteq \{1, 2, \ldots, n\}, |E| = n - t \geq d$. Shortening of $\mathcal{C}$ on the coordinates outside E gives a code $\mathcal{C}^E (n - t, \geq q^{-t} M, d)$.   □

Theorem 7.3: $\overline{R}(\delta)$ *is continuous.*

Proof: From the previous lemma we obtain
$$\overline{R}(\delta) \leq \tau + (1 - \tau)\overline{R}\Big(\frac{\delta}{1 - \tau}\Big).$$
Assume that $\tau < 1 - 2\delta$. Let $\eta = \delta/(1 - \tau)$, then $0 < \delta < \eta < 1/2$. We have
$$0 \leq \overline{R}(\delta) - (1 - \tau)\overline{R}(\eta) \leq \tau.$$
Letting $\tau \to 0$ proves the claim.   □

The same claim is also valid for $\overline{R}(\delta)$ for codes on S^{n-1}.

7.2. *Supports, Ranks, Subcodes*

Theorem 7.4: (i) *If $t \leq d(\mathcal{C}) - 1$, then $\mathcal{C}_E$ is an $[n - t, k, \geq d(\mathcal{C}) - t]$ code.*
 (ii) $\dim \mathcal{C}^E = |E| - \mathrm{rk}\, \mathbf{H}(E)$.
 (iii) $\mathcal{C}_E \cong \mathcal{C}/\mathcal{C}^{\bar{E}}$.
 (iv) $(\mathcal{C}_E)' = (\mathcal{C}')^E$; $(\mathcal{C}^E)' = (\mathcal{C}')_E$.

Proof: (i) will follow by Lemma 7.5. (ii)-(iii) are obvious. Let us prove the first part of (iv). Let $a \in (\mathcal{C}')^E$, then $\mathbf{G}(E)a^T = 0$, so $a \in (\mathcal{C}_E)'$. Further, by (ii)
$$\begin{aligned}\dim(\mathcal{C}')^E &= |E| - \mathrm{rk}(\mathbf{G}(E)) \\ &= |E| - \dim \mathcal{C}_E = \dim(\mathcal{C}_E)'.\end{aligned}$$

The second part of (iv) is analogous. $\qquad\qquad\square$

Lemma 7.5: $|E| - \mathrm{rk}(\mathbf{H}(E)) = k - \mathrm{rk}(\mathbf{G}(\bar{E}))$.

Proof: Let $\mathcal{C}_E = \mathrm{proj}_E\,\mathcal{C}$ be the projection of $\mathcal{C}$ on the coordinates in E. Clearly, $\dim \mathcal{C}_E = \mathrm{rk}(\mathbf{G}(E))$. On the other hand, $\mathcal{C}_E \cong \mathcal{C}/\mathcal{C}^{\bar{E}}$ by Theorem 7.4(iii); hence by (ii)

$$\dim \mathcal{C}_E = k - \dim \mathcal{C}^{\bar{E}} = k - |E| + \mathrm{rk}(\mathbf{H}(\bar{E})).$$
$$\square$$

Lemma 7.6: (The MacWilliams identities)

$$\sum_{i=0}^{n-u} A_i' \binom{n-i}{u} = |\mathcal{C}'| q^{-u} \sum_{i=0}^{u} A_i \binom{n-i}{n-u}. \tag{7.1}$$

Proof: We have the following chain of equalities:

$$\sum_{i=0}^{n-u} A_i' \binom{n-i}{u} = \sum_{|E|=n-u} |(\mathcal{C}')^E| \tag{7.2}$$

$$= \sum_{|E|=n-u} q^{n-u-\mathrm{rk}(\mathbf{G}(E))}$$

$$= q^{n-k-u} \sum_{|E|=n-u} q^{u-\mathrm{rk}(\mathbf{H}(\bar{E}))}$$

$$= q^{n-k-u} \sum_{i=0}^{u} A_i \binom{n-i}{n-u}.$$

Here the first equality follows by counting in two ways the size of the set

$$\{(E,c) : \sharp E = n - u \text{ and } c \in (\mathcal{C}')^E, |c| \le n - u\},$$

the second one is straightforward, the third one (the central step in the proof) is implied by Lemma 7.5, and the final step follows by the same argument as the first one. $\qquad\qquad\square$

Theorem 7.7:

$$|\mathcal{C}| A_j' = \sum_{i=0}^{n} A_i K_j(i), \tag{7.3}$$

where

$$K_j(i) = \sum_{\ell=0}^{i} (-1)^\ell \binom{i}{\ell} \binom{n-i}{j-\ell} (q-1)^{j-\ell}$$

is the Krawtchouk number.

Proof: Multiply both sides of (7.1) by $(-1)^u \binom{u}{\ell}$, sum over u from 0 to n, and use the fact that $\sum_{u \in \mathbb{Z}} (-1)^u \binom{n-i}{u} \binom{u}{s} = (-1)^{n-i} \delta_{n-i,s}$. This gives

$$A_j = |\mathcal{C}|^{-1} \sum_{i=0}^{n} A_i' \sum_{u=i}^{n} (-1)^{u-n+j} q^{n-u} \binom{u}{n-j} \binom{n-i}{n-u}.$$

The sum on u in the last expression is just another form for $K_j(i)$. $\square$

7.3. *Rank Distribution of a Linear Code*

Rewrite (7.2) by collecting on the right-hand side subsets of one and the same rank. Namely, let

$$U_u^v = |\{E \subseteq \{1, 2, \ldots, n\} \mid |E| = u, \mathrm{rk}(\mathbf{G}(E)) = v\}|.$$

Then by (7.2) and Theorem 7.4(ii) we have

$$\sum_{i=0}^{w} \binom{n-i}{n-w} A_i = \sum_{v=0}^{n-k} q^{w-v} (U')_w^v, \tag{7.4}$$

where the numbers U' are the rank coefficients of $\mathcal{C}'$. Further, Lemma 7.5 implies that

$$(U')_u^{u-k+v} = U_{n-u}^v.$$

The last two equations relate the weight enumerator of $\mathcal{C}$ and its *rank distribution* $(U_u^v, 0 \le u \le n, 0 \le v \le k)$.

Example 7.8: (MDS codes) An (n, M, d) code is called *maximum distance separable* (MDS) if $M = q^{n-d+1}$. Let $\mathcal{C}$ be an $[n, k, n-k+1]$ q-ary linear MDS code with a parity-check matrix $\mathbf{H}$. Then $\rho^* E = \mathrm{rk}(\mathbf{H}(E)) = \min\{|E|, n-k\}$ for all $E \subseteq S$, $0 \le |E| \le n-k$. Therefore

$$(U')_u^v = \begin{cases} \binom{n}{u} & \text{if } (0 \le v = u \le n-k) \text{ or } (u \ge n-k+1, \, v = n-k); \\ 0 & \text{otherwise.} \end{cases}$$

This enables us to compute the weight spectrum of $\mathcal{C}$. Substituting the values of $(U')_u^v$ into (7.4), we obtain $A_0 = 1$, $A_i = 0$ for $1 \leq i \leq n - k$, and

$$A_{n-k+\ell} = \binom{n}{k-\ell} \sum_{j=0}^{\ell-1} (-1)^j \binom{n-k+\ell}{j} (q^{\ell-j} - 1) \quad (1 \leq \ell \leq k). \quad (7.5)$$

Clearly, the dual code $\mathcal{C}'$ is also MDS of dimension $n - k$.

Definition 7.9: The *rank polynomial* of a linear code $\mathcal{C}$ is

$$U(x,y) = \sum_{u=0}^{n} \sum_{v=0}^{k} U_u^v x^u y^v.$$

Relations of the rank polynomial of $\mathcal{C}$, its dual code $\mathcal{C}'$, and the weight polynomial of $\mathcal{C}$ are given by the following results.

Theorem 7.10:

$$U'(x,y) = x^n y^{\dim \mathcal{C}'} U\left(\frac{1}{xy}, y\right).$$

Theorem 7.11:

$$A(x,y) = y^n |\mathcal{C}| \, U\left(\frac{x-y}{y}, \frac{1}{q}\right) = (x-y)^n U'\left(\frac{qy}{x-y}, \frac{1}{q}\right).$$

7.4. *Higher Weights and Weight Enumerators*

Let $\mathcal{C}$ be an $[n, k, d]$ q-ary linear code. Define the r-th support weight distribution of $\mathcal{C}$, $0 \leq r \leq k$, as the vector $(A_i^r, 0 \leq i \leq n)$, where

$$A_i^r = \sharp\{\mathcal{D} : \mathcal{D} \text{ is a linear subcode of } \mathcal{C}, \dim \mathcal{D} = r, |\operatorname{supp}(\mathcal{D})| = i\}.$$

Theorem 7.12: (*Generalized MacWilliams identities* [46])

$$\sum_{i=0}^{w} \binom{n-i}{n-w} A_i^r = \sum_{v=0}^{n-k} \begin{bmatrix} w-v \\ r \end{bmatrix} (U')_w^v \quad (0 \leq w \leq n, \ 0 \leq r \leq k).$$

This theorem is proved similarly to (7.4).

Theorem 7.13: [30] *Let*

$$D_{\mathcal{C}}^r(x,y) = \sum_{i=0}^{n} \left(\sum_{m=0}^{r} [r]_m A_i^m \right) x^{n-i} y^i,$$

24   *A. Barg*

then

$$D_{\mathcal{C}}^r(x,y) = q^{-r(n-k)}(y + (q^r - 1)x)^n D_{\mathcal{C}'}^r \left(\frac{y-x}{y + (q^r - 1)x} \right), \quad r \geq 0.$$

Here $[r]_m = \prod_{u=0}^{m-1}(q^m - q^i)$.

A simple way to prove this theorem is to realize that $D_{\mathcal{C}}^r(x,y)$ is the Hamming weight enumerator of the code $\mathcal{C}^{(r)} = \mathbb{F}_{q^r} \otimes_{\mathbb{F}_q} \mathcal{C}$.

Concluding remarks. The ideas of this section can be developed in several directions. First, it is possible to consider different versions of rank polynomials and of support weight distributions such as, for instance,

$$A_i^{(r)} = \sum_{E \subseteq \{1,\ldots,n\}, |E|=i} \sharp\big\{\{c_1, c_2, \ldots, c_r\} \subseteq \mathcal{C}, \mathrm{supp}(c_1, c_2, \ldots, c_r) = E\big\}$$

$$(i = 0, 1, \ldots, n).$$

The corresponding generating functions, as a rule, satisfy MacWilliams-type identities. This line of thought leads to matroid invariants that we mentioned in the beginning of this section. See [7], [15] for more on this subject.

Another avenue is to study alphabets with some algebraic structure such as abelian groups, finite rings, and modules over them. This enables one to define various norm-like functions on the alphabet and study weight enumerators of codes with respect to these functions [27]. When duality is appropriately defined, these enumerators typically also satisfy MacWilliams-type identities.

8. Decoding

Definition 8.1: Let $\mathcal{C} \subseteq X$ be a code in a metric space X. A (partial) mapping $\psi_t : X \to \mathcal{C}$ is called *decoding* if for any y such that $\mathrm{d}(y, \mathcal{C}) \leq t$,

$$\psi(y) = \arg\min_{c \in \mathcal{C}} \mathrm{d}(c, y).$$

For $y \notin \cup_{c \in \mathcal{C}} \mathcal{B}_t(c)$ the value of $\psi_t(y)$ is undefined.

We will only consider the two extremes: *complete decoding* and *error detection*. Under complete decoding, $t = r(\mathcal{C})$ (see Definition 3.2), *i.e.*, $X \subseteq \cup_{c \in \mathcal{C}} B_t(c)$. Under error detection, $t = 0$. Error detection will be briefly considered in the beginning of Section 9; otherwise we will focus on complete decoding, denoted by ψ hereafter.

To justify the term "decoding", assume that $\mathcal{C} \subseteq \mathscr{H}_q^n$ is used for transmission over a q-ary symmetric channel (qSC) given by a random mapping $Q \to Q$ such that

$$P(b|a) = (1-p)\delta_{a,b} + \frac{p}{q-1}(1-\delta_{a,b}),$$

where p is called the crossover probability of the channel, $p < (q-1)/q$. Suppose that the transmitted codeword c is received as $y \in \mathscr{H}_q^n$. The event $\psi(y) \neq c$ is called a decoding error. Let $P_{\mathrm{de}}(c)$ be its probability. As it turns out, the complete decoder ψ is a good choice for minimizing $P_{\mathrm{de}}(c)$.

Definition 8.2: Let $\mathcal{C}$ be a code. The *Voronoi domain* of a codeword c with respect to $\mathcal{C}$ is the set

$$D(c, \mathcal{C}) = \{x \in X : \forall_{c' \in \mathcal{C}} \; \mathrm{d}(x, c) \leq \mathrm{d}(x, c')\}.$$

Lemma 8.3: *Let $\mathcal{C}$ be a linear code and x a vector in $\mathscr{H}_q^n$. The complete decoding of x can be defined as follows:*

$$\psi(x) = x - \ell(\mathcal{C} - x),$$

where $\ell(\mathcal{C} - x)$ is a vector of lowest weight in the coset $\mathcal{C} - x$.

9. Error Probability

In this and the next section we are concerned with upper bounds on the error probability of complete decoding of the best possible codes used on a binary symmetric channel (BSC). The analysis performed for this channel offers a simple model for results on error exponents for arbitrary memoryless channels. For the BSC it is possible to derive the error exponent bound starting with a transparent geometric description of error events. The ideas developed below are central to (single-user) information theory. Although they are several decades old, they continue to attract attention of researchers to this date. In particular, the main problem in this area, that of the exact error exponent, is still unsolved.

Let $X = \mathscr{H}_q^n$. Let $\mathcal{C}$ be used for transmission over a qSC with crossover probability p. If c is the transmitted codeword, then the channel defines a probability distribution on X given by

$$P(y|c) = \left(\frac{p}{q-1}\right)^{\mathrm{d}(y,c)}(1-p)^{n-\mathrm{d}(y,c)}.$$

The error probability of decoding for a given vector $c \in \mathcal{C}$ is defined as

$$P_{\mathrm{de}}(c) = \Pr\{X \setminus D(c, \mathcal{C})\} = \sum_{y \in X \setminus D(c, \mathcal{C})} P(y|c).$$

The average error probability for the code $\mathcal{C}$ is

$$P_{\mathrm{de}}(\mathcal{C}) = |\mathcal{C}|^{-1} \sum_{c \in \mathcal{C}} P_{\mathrm{de}}(c).$$

The probability of undetected error $P_{\mathrm{ue}}(\mathcal{C})$ for the code $\mathcal{C}$ is defined analogously.

Theorem 9.1:

$$P_{\mathrm{ue}}(\mathcal{C}) = A\left(1 - p, \frac{p}{q-1}\right) - (1-p)^n.$$

Proof: Let $\pi(i) = \left(\frac{p}{q-1}\right)^i (1-p)^{n-i}$ and let $A_i(c) = \sharp\{c' \in \mathcal{C} : \mathrm{d}(c, c') = i\}$. We calculate

$$P_{ue}(\mathcal{C}) = |\mathcal{C}|^{-1} \sum_{c \in \mathcal{C}} \sum_{c' \in \mathcal{C} \setminus \{c\}} \pi(\mathrm{d}(c, c')) = |\mathcal{C}|^{-1} \sum_{c \in \mathcal{C}} \sum_{i=1}^{n} A_i(c) \pi(i)$$

$$= \sum_{i=1}^{n} A_i(\mathcal{C}) \pi(i)$$

$$= A\left(1 - p, \frac{p}{q-1}\right) - (1-p)^n. \qquad \square$$

Definition 9.2: The *error exponent* for the Hamming space (known also as the *reliability function* of the q-ary symmetric channel) is defined as follows:

$$E(R, p, n) = -n^{-1} \log\left(\min_{\mathcal{C}: R(\mathcal{C}) \geq R} P_{\mathrm{de}}(\mathcal{C})\right),$$

$$E(R, p) = \lim_{n \to \infty} E(R, p, n).$$

Conventions made after the formula for $R(\delta)$ in Section 3.2 apply to this definition as well.

Analogously to this definition one defines the exponent $E_{\mathrm{ue}}(R, p)$ of the probability of undetected error. It is easy to derive a lower bound on this exponent.

Theorem 9.3:

$$E_{\mathrm{ue}}(R,p) \geq D(\delta_{\mathrm{GV}}(R)||p) + \log q - R \qquad 0 \leq R \leq \delta_{\mathrm{GV}}(p),$$

$$E_{\mathrm{ue}}(R,p) \geq \log q - R \qquad\qquad\qquad \delta_{\mathrm{GV}}(p) \leq R \leq 1.$$

Proof: Follows by combining Theorem 9.1 and Corollary 4.6. □

9.1. A Bound on $P_{de}(\mathcal{C})$

We proceed with bounds on $P_{\mathrm{de}}(\mathcal{C})$.

$$P_{\mathrm{de}}(\mathcal{C}) \leq |\mathcal{C}|^{-1} \sum_{c \in \mathcal{C}} \sum_{c' \in \mathcal{C} \backslash \{c\}} P(c \to c'),$$

where

$$P(c \to c') := \sum_{y \in X:\, \mathrm{d}(y,c') \leq \mathrm{d}(y,c)} P(y|c)$$

is the probability, under the binomial distribution, of the half-space cut out by the median hyperplane between c and c'. Note that $P(y|c)$ depends only on the Hamming weight of the error vector $x = y - c$.

Lemma 9.4: *Let $P_{de}(\mathcal{C}, x \in U)$ be the joint probability of decoding error and the event $x \in U \subseteq X$. Then for any $r = 0, 1, \ldots, n$,*

$$P_{de}(\mathcal{C}, p) \leq P(\mathcal{C}, x \in \mathcal{B}_r(0)) + P(x \notin \mathcal{B}_r(0)).$$

Let us specialize this result using the distance distribution of the code.

Lemma 9.5: *Let $\mathcal{C}$ be a d-code with distance distribution $(A_0, A_d, \ldots, A_n)$. Then for any $r = 0, 1, \ldots, n$,*

$$P_{de}(\mathcal{C}, p) \leq P_1 + P_2, \tag{9.1}$$
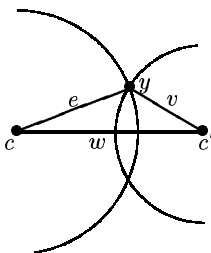
where

$$P_1 = \sum_{w=d}^{2r} A_w \sum_{e=\lceil w/2 \rceil}^{r} |\mathcal{B}_e(c) \cap \mathcal{S}_e(0)| p^e (1-p)^{n-e} \tag{9.2}$$

$$= \sum_{w=d}^{2r} A_w \sum_{i=\lceil w/2 \rceil}^{r} \sum_{\ell=0}^{r-i} \binom{w}{i} \binom{n-w}{\ell} p^{i+\ell} (1-p)^{n-i-\ell}, \tag{9.3}$$

where c is a code vector with $|c| = w$, and

$$P_2 = \sum_{e=r+1}^{n} \binom{n}{e} p^e (1-p)^{n-e}. \tag{9.4}$$

Proof: Let c be the transmitted codeword, let x be the error vector in the channel, $|x| = e$, and let $y = c + x$ be the received vector. A decoding error occurs if $v = \mathrm{d}(y, c') \leq \mathrm{d}(y, c) = e$.

Let $\mathrm{d}(c, c') = w$ and suppose that $A_w(c)$ is the number of codewords in $\mathcal{C}$ at distance w from c. Since all the error vectors of one and the same weight are equiprobable, we have

$$P_{\mathrm{de}}(c, x \in \mathcal{B}_r(0)) \leq A_w(c) \sum_{e=\lceil w/2 \rceil}^{r} |\mathcal{B}_e(c') \cap \mathcal{S}_e(0)| p^e (1-p)^{n-e}.$$

Computing the average value of $P_{\mathrm{de}}(c, x \in \mathcal{B}_r(0))$ over the code, we obtain (9.2). To obtain (9.3), observe that, as in Lemma 6.4,

$$P_1 = \sum_{w=d}^{2r} A_w \sum_{e=\lceil w/2 \rceil}^{r} p^e (1-p)^{n-e} \sum_{v=0}^{r} p_{e,v}^w. \tag{9.5}$$

Let $i = |\operatorname{supp}(x) \cap \operatorname{supp}(c, c')|$ and $\ell = e - i$. Now (9.3) follows by substituting the definition of $p_{e,v}^w$ and renaming the summation indexes.

The expression for P_2 is immediate.   □

The choice of $U = \mathcal{B}_r(0)$ in the last two lemmas is justified by the fact that the noise is spherically symmetric. It makes sense to choose the radius r so that the bound on P_{de} is minimized. An interesting fact is that for random codes of rate R this minimum is attained for $r/n \to \delta_{\mathrm{GV}}(R)$.

Lemma 9.6: *The minimum of the bound (9.1) is attained for $r = r_0$, where*

$$r_0 = \min \Big\{ r : \sum_{w=d}^{2r} A_w \sum_{e=\lceil w/2 \rceil}^{r} \big| \mathcal{B}_e(c) \cap \mathcal{S}_e(0) \big| \geq S_r \Big\},$$

where $c \in \mathcal{C}, |c| = w$. Further, if $(\mathcal{C}_i, i = 1, 2, \ldots)$ is a sequence of random linear codes of growing length, then $\lim_{n \to \infty} \frac{r_0}{n} = \delta_{GV}(R)$.

Proof: The first part of the claim is obvious since P_1 is a growing and P_2 a falling function of r. Let us prove the second part. Let A_w be an element of the weight distribution of a random linear code of length n. We then have (see Theorem 4.5)

$$A_w \leq n^2 \binom{n}{w} 2^{Rn-n}.$$

Using Corollary 6.5, we now obtain

$$n^2 2^{Rn-n} \sum_{w=d}^{2r} \binom{n}{w} \sum_{e=\lceil w/2 \rceil}^{r} \big| \mathcal{B}_e(c) \cap \mathcal{S}_e(0) \big| \cong 2^{Rn-n} \binom{n}{r}^2.$$

To find r_0 for large n, we equate this to $S_r = \binom{n}{r}$ to obtain

$$2^{Rn-n} \binom{n}{r_0} \cong 1,$$

i.e., $r_0/n \to \delta_{\mathrm{GV}}(R)$, as claimed. $\qquad\square$

10. Upper Bound on the Error Probability

The error probability of decoding of good codes is known to fall exponentially when the code length increases, for all code rates less than the capacity $\mathscr{C}$ of the channel. The following theorem gives a bound on this exponent.

Theorem 10.1: *$E(R, p) \geq E_0(R, p)$, where for $0 \leq R \leq R_x$*

$$E_0(R, p) = E_x(R, p) = -\delta_{\mathrm{GV}}(R) \log 2 \sqrt{p(1-p)}, \qquad (10.1)$$

for $R_x \leq R \leq R_{\mathrm{crit}}$

$$E_0(R, p) = E_r(R, p) = D(\rho_0 \| p) + R_{\mathrm{crit}} - R, \qquad (10.2)$$

A. Barg

and for $R_{\mathrm{crit}} \leq R \leq \mathscr{C} = 1 - h_2(p)$

$$E_0(R,p) = E_{\mathrm{sp}}(R,p) = D(\delta_{\mathrm{GV}}(R)\|p); \qquad (10.3)$$

here

$$R_x = 1 - h_2(\omega_0), \qquad (10.4)$$

$$R_{\mathrm{crit}} = 1 - h_2(\rho_0), \qquad (10.5)$$

$$\rho_0 = \frac{\sqrt{p}}{\sqrt{p} + \sqrt{1-p}}, \qquad \omega_0 := 2\rho_0(1-\rho_0) = \frac{2\sqrt{p(1-p)}}{1 + 2\sqrt{p(1-p)}}.$$

For $\mathscr{C} \leq R \leq 1$, $E(R,p) = 0$.

Proof: Suppose we transmit with random $[n, Rn, d]$ linear codes, $d/n \to \delta_{\mathrm{GV}}(R)$. We use (9.1), (9.2), and (9.4) to bound the error probability

$$P_1 \leq n^2 2^{Rn-n} \sum_{w=d}^{2r} \binom{n}{w} \sum_{e=\lceil w/2 \rceil}^{r} |\mathcal{B}_e(c) \cap \mathcal{S}_e(0)| p^e (1-p)^{n-e}. \qquad (10.6)$$

Letting $e = \rho n$, we obtain from Corollary 6.5 the estimate

$$P_1 \cong \max_{\delta_{\mathrm{GV}}(R)/2 \leq \rho \leq \delta_{\mathrm{GV}}(R)} \exp[-n(D(\rho\|p) + (1-R) - h_2(\rho))].$$

The unconstrained maximum on ρ on the right-hand side (the minimum of the exponent) is attained for $\rho = \rho_0$, and, again by Corollary 6.5, the unconstrained maximum on $\omega = w/n$ in (10.6) is attained for $\omega = \omega_0$. The three cases (10.1)-(10.3) are realized depending on how ω_0 and ρ_0 are located with respect to the optimization limits.

The case (10.2) corresponds to ω_0, ρ_0 within the limits: $\delta_{\mathrm{GV}}(R)/2 \leq \rho_0 \leq \delta_{\mathrm{GV}}(R), \omega_0 \geq \delta_{\mathrm{GV}}(R)$. Then the exponent of P_1 is

$$D(\rho_0\|p) + h_2(\delta_{\mathrm{GV}}(R)) - h_2(\rho_0), \qquad (10.7)$$

i.e., the random coding exponent of (10.2). We need to compare the exponent of P_1 with the exponent $D(\delta_{\mathrm{GV}}(R)\|p)$ of P_2. Under the assumption $\rho_0 < \delta_{\mathrm{GV}}(R)$ their difference is

$$[D(\rho_0\|p) - h_2(\rho_0)] - [D(\delta_{\mathrm{GV}}(R)\|p) - h_2(\delta_{\mathrm{GV}}(R))] < 0$$

since $D(x\|p) - h_2(x)$ is an increasing function of x for $x > \rho_0$. This proves that the dominating exponent for $R \leq R_{\mathrm{crit}}$ is given by (10.7).

Suppose now that $\omega_0 \geq \delta_{\mathrm{GV}}(R)$ and $\rho_0 \geq \delta_{\mathrm{GV}}(R)$, i.e., $R \geq R_{\mathrm{crit}}$. In this case the exponent of P_1 is dominated by the term with $\rho = \delta_{\mathrm{GV}}(R)$. Then we conclude that the exponents of P_1 and P_2 are both equal to the sphere-packing exponent of (10.3).

If $\omega_0 \leq \delta_{\mathrm{GV}}$, i.e., $R \leq R_x$, the maximum on ω is attained for $\omega = \delta_{\mathrm{GV}}$, and we get for the right-hand side of (10.6) the following expression:

$$\sum_{e \geq d/2} \binom{n\delta_{\mathrm{GV}}}{n\delta_{\mathrm{GV}}/2} \binom{n(1 - \delta_{\mathrm{GV}})}{e - (^{1}/_{2})n\delta_{\mathrm{GV}}} p^e (1 - p)^{n-e}.$$

This is maximized when $e - (1/2)n\delta_{\mathrm{GV}} = n(1 - \delta_{\mathrm{GV}})p$, i.e., for

$$\rho = \frac{e}{n} = (1 - \delta_{\mathrm{GV}})p + \frac{\delta_{\mathrm{GV}}}{2}.$$

Substituting, we obtain the "expurgation exponent" $E_x(R, p)$ of (10.1). To finish off this case we need to show that the exponent $D(\delta_{\mathrm{GV}} \| p)$ of the term $P[\mathrm{w}(y) \geq d]$ is greater for $\omega_0 \leq \delta_{\mathrm{GV}} \leq 1/2$ than $E_x(R, p)$. This is confirmed by a straightforward calculation.

The proof of the equality $E(R, 0) = 0$ for $R \geq \mathscr{C}$ (the "converse coding theorem") will be omitted. □

In the next two subsections we study a geometric interpretation of this theorem and provide background and intuition behind it.

10.1. *Geometric View of Theorem 10.1*

A close examination of the proof reveals the intuition behind decoding error events for random codes. The capacity region of the BSC is given on the (R, p)-plane by $(0 \leq R, 0 \leq p \leq 1/2, R + h_2(p) \leq 1)$. According to the three cases in the theorem, this region can be partitioned naturally into the regions of low noise A, moderate noise B, and high noise C, where

$$\begin{aligned}
(10.1) \qquad & A = \{(R, p) : R \leq 1 - h_2(\omega_0)\}, \\
(10.2) \qquad & B = \{(R, p) : 1 - h_2(\omega_0) \leq R \leq 1 - h_2(\rho_0)\}, \\
(10.2) \qquad & C = \{(R, p) : 1 - h_2(\rho_0) \leq R \leq 1 - h_2(p)\}.
\end{aligned}$$

As n increases, within each region the error events are dominated by a particular (relative) weight ω_{typ} of incorrectly decoded codewords. Moreover, the relative weight ρ_{typ} of error vectors that form the main contribution to

the error rate also converges to a particular value. We have, for the regions A, B, and C, respectively,

$$\omega_0 < \delta_{\mathrm{GV}}, \qquad\qquad \rho_{\mathrm{typ}} = (1 - \delta_{\mathrm{GV}})p + \tfrac{1}{2}\delta_{\mathrm{GV}}, \quad \omega_{\mathrm{typ}} = \delta_{\mathrm{GV}},$$
$$\omega_0 \geq \delta_{\mathrm{GV}}, \ \rho_0 < \delta_{\mathrm{GV}}, \quad \rho_{\mathrm{typ}} = \rho_0, \qquad\qquad\qquad \omega_{\mathrm{typ}} = 2\rho_0(1 - \rho_0),$$
$$\omega_0 \geq \delta_{\mathrm{GV}}, \ \rho_0 \geq \delta_{\mathrm{GV}}, \quad \rho_{\mathrm{typ}} = \delta_{\mathrm{GV}}, \qquad\qquad\quad \omega_{\mathrm{typ}} = 2\delta_{\mathrm{GV}}(1 - \delta_{\mathrm{GV}}).$$

When the code is used in the low-noise region, the typical relative weight of incorrectly decoded codewords is $\delta_{\mathrm{GV}}(R)n$, i.e., it does not depend on the channel. In the moderate-noise region, the typical weight of incorrect codewords is ρ_0 and in the high-noise region it is $\delta_{\mathrm{GV}}(R)$. We observe therefore that for $R > R_x$ the error probability does not depend on the minimum distance of the code.

The geometry of decoding for $R < R_{\mathrm{crit}}$ and for $R > R_{\mathrm{crit}}$ is of very different nature. Consider an error event that corresponds to the moderate-noise region. Its probability is dominated by errors y of relative weight ρ_0. From the proof of the theorem and Corollary 6.5 it can be seen that the number of points of the sphere $\mathcal{S}_{\rho_0 n}$ that are decoded incorrectly behaves exponentially as

$$\frac{\binom{n}{\rho_0 n}}{\binom{n}{\delta_{\mathrm{GV}} n}} \binom{n}{\rho_0 n};$$

hence their fraction has the same exponent as $\binom{n}{\rho_0 n} / \binom{n}{\delta_{\mathrm{GV}} n}$. We see that for $\rho_0 < \delta_{\mathrm{GV}}$ an exponentially small fraction of error vectors y of weight $\rho_0 n$ leads to an incorrect codeword c'. We have $|c'|/n \to \omega_0, \mathrm{d}(y, c')/n \to \rho_0$.

Moreover, with some additional arguments it is possible to show that an error vector y on the sphere of radius $\rho_0 n$ around the transmitted codeword typically falls in at most one ball $\mathcal{B}_{\rho_0 n}(c')$ around an incorrect codeword c'. Hence, the union bound of $(9.2), (10.6)$ *for random linear codes* is exponentially tight. Error vectors y that are incorrectly decoded to a codeword c' occupy one and the same fraction

$$\cong \frac{\binom{n}{\rho_0 n}}{\binom{n}{\delta_{\mathrm{GV}} n}} \frac{\binom{n}{\delta_{\mathrm{GV}} n}}{\binom{n}{w}} = \frac{\binom{n}{\rho_0 n}}{\binom{n}{w}}$$

of the sphere $\mathcal{S}_{\rho_0 n}$ for almost every incorrect codeword $c', |c'| = \omega_0 n$.

On the other hand, once R exceeds R_{crit} or $\rho_0 > \delta_{\mathrm{GV}}$, almost every point y on the sphere $\mathcal{S}_{\rho_0 n}(c)$ leads to a decoding error; the only relief comes from the fact that such points are received from the channel in an

exponentially small fraction of transmissions. For $R > R_{\mathrm{crit}}$ almost every incorrectly decoded error vector y will have exponentially many codewords that are same distance or nearer to it as the transmitted word.

10.2. *Explication of the Random Coding Bounds*

Random coding bounds were a central topic in information theory in the first few decades of its development. They can be derived by a variety of approaches. Following the methods developed in this chapter, we single out two related, though not identical ways to random coding bounds, both with long lineage. The first one is writing an estimate for the error probability of complete decoding of a code and then averaging this expression over an ensemble of random codes. The strongest results on this way were obtained by Gallager whose treatise [26] also gives the most comprehensive account of this method. A recent informal discussion of it by one of the main contributors is found in [13].

The second approach suggests first to establish properties of a good code in the ensemble such as distance distribution and then to estimate the error probability for this particular code. This idea was also suggested by Gallager [25]. This is what we did when we first proved Theorem 4.5 and then used the code whose existence is proved in it, in Theorem 10.1.

There are two reasons for which the second approach may prevail. First, under it, Theorem 10.1 generalizes without difficulty to *arbitrary discrete memoryless channels.* Of course, in this case the Hamming weight and distances do not describe the effect of noise adequately; their role is played by the composition ("type") codewords and information distance between types, respectively [20]. Remarkably, it is possible to extend some of the geometric intuition of Section 10.1 to this very general case [19].

Apart from this generalization, in engineering applications it is more important to be able to bound the error probability for a specific code than for an ensemble of codes. Constructing good bounds on this probability received much attention in the last decade (see [42], [43]). This problem has also revived interest in deriving random coding bounds via estimating the error probability for a specific code.

There is also a minor technical problem with computing the average error probability for some code ensembles: for low rates typical codes are often poor (cf. Theorem 4.2). To deal with this, one usually employs "expurgation" of the code, i.e., removing from it codewords for which the error

probability $P_{\mathrm{de}}(C)$ is large. This issue is discussed in more detail in [26], [8].

Other methods in this area include hypothesis testing [14] and, lately, applications of statistical mechanics.

10.3. *Spherical Codes*

Similar results can be derived for codes on the sphere $\mathscr{S}^n$ in $\mathbb{R}^n$ used over a Gaussian channel with additive white noise. In particular, an analog of Theorem 10.1 was proved by Shannon in [45]. As it is seen now [6], the idea of Shannon's proof is qualitatively similar to the proof of Theorem 10.1, though the analytic part is somewhat more involved.

11. Polynomial Method

This section is concerned with one application of harmonic analysis to extremal problems of coding theory. The ideas are primarily due to MacWilliams [37] and Delsarte [21]. We will provide details for codes in the Hamming space; however, it should be kept in mind that a similar theory can be developed for a large class of finite and infinite metric spaces [22], [28].

Theorem 11.1: [21] *Let $x \in \mathscr{H}_q^n, |x| = i$, be a vector. Then the Krawtchouk number $K_k(i)$ equals*

$$K_k(i) = \sum_{y \in \mathscr{H}_q^n, \, |y|=k} \varphi_x(y),$$

where $\varphi_x(y) = \exp(2\pi i(x, y)/q)$ is a character of the additive group $(\mathbb{Z}_q)^n$.

Definition 11.2: Let $\mathbf{K} = \|K_k(i)\|$ be the $(n+1) \times (n+1)$ Krawtchouk matrix with rows numbered by i and columns by k, and let $A = (A_0, A_1, \ldots, A_n)$ be the distance distribution of an (n, M) code $\mathcal{C}$ in $\mathscr{H}_q^n$. The vector

$$(A'_0, A'_1, \ldots, A'_n) = M^{-1} A \mathbf{K}$$

is called the *dual distance distribution* of $\mathcal{C}$.

The number $d' = \min(i \geq 1 : A_i > 0)$ is called the *dual distance* of the code $\mathcal{C}$. If $\mathcal{C}$ is linear, then d' is the minimum distance of the dual code $\mathcal{C}'$.

The code $\mathcal{C}$ is called a *design* of strength t if $d'(\mathcal{C}) = t + 1$.

Theorem 11.3: [21] *The components of the dual distance distribution of any code are nonnegative, and $A_0' = 1$, $\sum A_i' = q^n/M$.*

The main application of this result to extremal problems of coding theory is given in the following theorem.

Theorem 11.4: *Let $\mathcal{C}$ be code with distance d and dual distance d'. To bound*

$$F(g) = \sum_{i=d}^{n} g(i)A_i(\mathcal{C})$$

below, choose

$$f(x) = \sum_{k=0}^{n} f_k K_k(x), \ f_k \geq 0, d' \leq k \leq n,$$

so that $f(i) \leq g(i), d \leq i \leq n$. Then

$$F(g) \geq |\mathcal{C}|f_0 - f(0).$$

To bound $F(g)$ above, choose

$$h(x) = \sum_{k=0}^{n} h_k K_k(x), \ h_k \leq 0, d' \leq k \leq n,$$

so that $h(i) \geq g(i), d \leq i \leq n$. Then

$$F(g) \leq |\mathcal{C}|h_0 - h(0).$$

Proof: For instance, let us prove the second part. We have

$$\sum_{i=0}^{n} h(i)A_i = \sum_{i=0}^{n} A_i \sum_{j=0}^{n} h_j K_j(i) = \sum_{j=0}^{n} h_j |\mathcal{C}|A_j' \leq h_0 |\mathcal{C}|.$$

Here the second step is by definition of A_j' and the final step is justified by the Delsarte inequalities. Hence

$$F(g) = \sum_{i=d}^{n} g(i)A_i \leq \sum_{i=d}^{n} h(i)A_i \leq h_0|\mathcal{C}| - h(0).$$

The first part of the theorem is established analogously.   □

A. Barg

Corollary 11.5: *Let*

$$f(x) = \sum_{k=0}^{n} f_k K_k(x), \ f_0 > 0, \ f_k \geq 0, k = 1, 2, \ldots, n,$$

be a polynomial such that $f(i) \leq 0, i = d, d+1, \ldots, n$. Then for any (n, M, d) code we have

$$M \leq \frac{f(0)}{f_0}.$$

Proof: Take $g \equiv 0$ in the first statement of Theorem 11.4.   □

Theorem 11.4 is essentially due to Delsarte [21] who proved it in the form given in the last corollary. It was realized not long ago [3] that the same method can be used for bounding code parameters other than its size. This approach implies numerous results in extremal problems in coding theory, both "finite" and asymptotic.

Corollary 11.6: *Let $g(i) = p^i (1-p)^{n-i}$ and let $f(x)$ be a polynomial that satisfies the conditions of Theorem 11.4. Then for any code $\mathcal{C}$,*

$$P_{ue}(\mathcal{C}) \geq |\mathcal{C}| f_0 - f(0).$$

Corollary 11.7: *Let $r \in \{1, 2, \ldots, n\}$. For any code $\mathcal{C}$ there exists a number $i, 1 \leq i \leq r$, such that*

$$A_i(\mathcal{C}) \geq \frac{|\mathcal{C}| f_0 - f(0)}{f(i)}.$$

Results similar to those established here for $\mathcal{H}_q^n$ can be proved for the Johnson space $\mathcal{J}^{n,w}$. In the Johnson space the role of Krawtchouk polynomials is played by one class of Hahn polynomials $H_k(x)$.

Spherical codes. Let $\mathcal{C} \subseteq X = \mathcal{S}^n$ and $b(s, t)$ be its distance (inner product) distribution, see Section 3.1. The analog of Delsarte inequalities has the following form.

Theorem 11.8: [23], [28]

$$\int_{-1}^{1} P_k^{\alpha, \alpha}(x) db(x) \geq 0, \quad k = 0, 1, \ldots,$$

where $\alpha = (n-3)/2$ and $P_k^{\alpha, \alpha}(x)$ is a Gegenbauer polynomial of degree k.

Corollary 11.9: *Let $\mathcal{C}$ be a code with distance distribution $b(s,t)$. To bound*

$$F(g) = \int_{-1}^{t} g(x)db(x),$$

choose

$$f(x) = \sum_{k=0}^{l} f_k P_k^{\alpha,\alpha}(x), \quad f_k \geq 0, \ 1 \leq k \leq l, \ \ l = 1, 2, \ldots,$$

so that $f(x) \leq g(x), -1 \leq x \leq t$. Then

$$F(g) \geq |\mathcal{C}|f_0 - f(1).$$

12. Krawtchouk Polynomials

Definition 12.1: Krawtchouk polynomials are real polynomials orthogonal on $\{0, 1, \ldots, n\}$ with weight $\mu(i) = q^{-n}\binom{n}{i}(q-1)^i$:

$$\langle K_i, K_j \rangle = \binom{n}{i}(q-1)^i \delta_{i,j},$$

where

$$\langle K_i, K_j \rangle = \sum_{s=0}^{n} K_i(s)K_j(s)\mu(s).$$

Explicitly,

$$K_k(x) = \sum_{\sigma=0}^{k} (-1)^\sigma \binom{x}{\sigma}\binom{n-x}{k-\sigma}(q-1)^{k-j}.$$

Properties.

$$\|K_i\|^2 = \binom{n}{i}(q-1)^i; \ K_i(0) = \binom{n}{i}(q-1)^i.$$

$$K_i(s) \leq \sqrt{\frac{\binom{n}{i}(q-1)^i}{\mu(s)}}.$$

Any polynomial $f(x)$ of degree $\leq n$ can be expanded into the Krawtchouk basis: $f = \sum f_k K_k$, where

$$f_k = \frac{\langle f, K_k \rangle}{\langle K_k, K_k \rangle}. \tag{12.1}$$

In particular, let $f(x) = K_i(x)K_j(x)$, then

$$f(a) = \sum_{k=0}^{n} p_{i,j}^k K_k(a), \ a = 0, 1, \ldots, n,$$

where $p_{i,j}^k$ is the intersection number of $\mathscr{H}_2^n$.

Proof: We treat the case $q = 2$, the general case being analogous.

$$K_i(s)K_j(s) = \sum_{|\mathbf{y}|=i} (-1)^{(\mathbf{x},\mathbf{y})} \sum_{|\mathbf{y}'|=j} (-1)^{(\mathbf{x},\mathbf{y}')} = \sum_{\substack{\mathbf{y},\mathbf{y}' \\ |\mathbf{y}|=i, |\mathbf{y}'|=j}} (-1)^{(\mathbf{x},\mathbf{y}-\mathbf{y}')}.$$ $\square$

From this we can derive another expression for $p_{i,j}^k$:

$$\left\langle \sum_{k=0}^{n} p_{i,j}^k K_k, K_l \right\rangle = p_{i,j}^l \binom{n}{l} = \langle K_i K_j, K_l \rangle = \sum_{\sigma=0}^{n} K_i(\sigma)K_j(\sigma)K_l(\sigma)\mu(\sigma).$$

The polynomial $K_k(x)$ has k simple zeros $x_{1,k} < x_{2,k} < \cdots < x_{k,k}$ located between 0 and n. Zeros of $K_k(x)$ and $K_{k+1}(x)$ possess the "interlacing" property:

$$0 < x_{1,k+1} < x_{1,k} < x_{2,k+1} < \cdots < x_{k,k} < x_{k+1,k+1} < n.$$

Most of these and many other properties of Krawtchouk polynomials were proved by Delsarte [22]. Given the importance of Krawtchouk polynomials for coding theory, they have received much attention. Comprehensive surveys of their properties are provided in [35], [32].

In the Johnson space $\mathscr{J}^{n,w}$ the role of Krawtchouk polynomials is played by Eberlein polynomials $E_k(x)$ and some Hahn polynomials $H_k(x)$ (they are the p- and q-polynomials of the Johnson association scheme, respectively). The polynomials $H_k(x)$ are orthogonal on the set $\{0, 1, \ldots, w\}$ with weight $\mu(i) = \binom{w}{i}\binom{n-w}{i}/\binom{n}{w}$. Their properties were established in [22], see also [39].

Asymptotics. Asymptotic behavior of extremal zeros and of the polynomials themselves plays a major role in coding-theoretic applications of the polynomial method.

Let $K_{\tau n}(\xi n)$ be a binary Krawtchouk polynomial, $\tau < 1/2$. We are concerned with the asymptotics of the first zero $x_{1,\tau n}$ and of the exponent $k(\tau, \xi) = \lim_{n\to\infty} n^{-1} \log_2 K_t(x)$. Let $\phi(u) = (1/2) - \sqrt{u(1-u)}$.

The zeros of $K_t(x)$ are located inside the segment $[n\phi(\tau), n(1 - \phi(\tau))]$ and for the minimum zero we have

$$n\phi(\tau) \leq x_{1,\tau n} \leq n\phi(\tau) + t^{1/6}\sqrt{n - t}. \qquad (12.2)$$

The following is known about $k(\tau, \xi)$.

$$k(\tau, \xi) \leq k_1(\tau, \xi) = \frac{1}{2}(h_2(\tau) - h_2(\xi) + 1) \quad (0 \leq \tau \leq 1),$$

$$k(\tau, \xi) \sim k_2(\tau, \xi) = h_2(\tau) + I(\tau, \xi) \quad (0 \leq \xi < \phi(\tau)),$$

$$k(\tau, \xi) < k_3(\tau, \xi) = \frac{1}{2}\Big[\xi \log_2 \frac{\phi(\tau)}{1 - \phi(\tau)} + \log_2(1 - \phi(\tau)) + h_2(\tau) + 1\Big]$$

$$(0 \leq \xi \leq \phi(\tau)).$$

Here

$$I(\tau, \xi) = \int_0^\xi \log \frac{s + \sqrt{s^2 - 4y(1 - y)}}{2 - 2y} dy,$$

where $s = 1 - 2\tau$. While these asymptotic relations cover coding theory needs, much more accurate results on the asymptotic behavior of $K_t(x)$ in the region $x < (1/2)(n-1) - \sqrt{t(n-t)}$ (i.e., outside the oscillatory region) are given in [31].

We note that $k_1(\tau, \xi) \geq k_3(\tau, \xi) \geq k_2(\tau, \xi)$ for $0 \leq \xi \leq \phi(\tau)$, with equality if and only if $\xi = \phi(\tau)$. Moreover, for this ξ also

$$(k_1(\tau, \xi))'_\xi = (k_2(\tau, \xi))'_\xi = (k_3(\tau, \xi))'_\xi.$$

It is also possible to prove a lower bound on Krawtchouk polynomials.

Theorem 12.2: *Let $x \in [x_{1,t}, n - x_{1,t}]$ be an integer. Then*

$$(K_t(x) + K_{t+1}(x))^2 \geq O(1)\binom{n}{t}/\mu(x).$$

With the exception of the last theorem, similar results are known for the Hahn [36] and Gegenbauer [2] polynomials.

12.1. *Jacobi Polynomials*

The role of Gegenbauer polynomials for projective spaces over $\mathbb{R}, \mathbb{C}$, and $\mathbb{H}$ is played by various Jacobi polynomials. Jacobi polynomials $P_k^{\alpha,\beta}(x)$ are

A. Barg

orthogonal on the segment $[-1, 1]$ with respect to the measure $d\mu(x) = (1 - x)^\alpha (1 + x)^\beta dx$. We have

$$P_k^{\alpha,\beta}(x) = L_k \prod_{j=1}^{k} (x - t_{j,k}),$$

where

$$L_k = 2^{-k} \sum_{\nu=0}^{k} \binom{k + \alpha}{k - \nu} \binom{k + \beta}{\nu}.$$

The zeros $t_{j,k}$ of P_k are located between -1 and 1; we assume numbering from right to left: $t_{k,k} < t_{k-1,k} < \cdots < t_{1,k}$. We have

$$P_k^{\alpha,\beta}(x) = (-1)^k P_k^{\beta,\alpha}(-x).$$

Zeros of $P_k^{\alpha,\alpha}(x)$ are symmetric with respect to 0, and $t_{k,k} = -t_{1,k}$.

For $k \to \infty, \alpha = ak, \beta = bk$, we have [28]

$$t_{1,k} \to q(a, b) := 4\sqrt{(a + b + 1)(a + 1)(b + 1)} - \frac{a^2 + b^2}{(a + b + 2)^2},$$

$$t_{k,k} \to -q(b, a);$$

in particular, for $\alpha = \beta$ this implies

$$t_{1,k} \to q(a, a) = \frac{\sqrt{1 + 2a}}{1 + a}.$$

The asymptotic exponent of $P_k^{\alpha,\beta}$ has the same qualitative behavior as that of the Krawtchouk and Hahn polynomials, see [2], [31].

13. Christoffel-Darboux Kernel and Bounds on Codes

This section deals with a famous application of the polynomial method to extremal problems. Following [39], consider the polynomial $f(x) = (K_t(a))^2 W_t(x)$, where

$$W_t(x) = \frac{(K_{t+1}(x) + K_t(x))^2}{a - x}.$$

Here a is the smallest root of $K_{t+1}(x) + K_t(x)$ and t is a parameter.

Theorem 13.1: *The polynomial $f(x)$ has the following properties:*
(i) *in the expansion $f(x) = \sum_{k=0}^{2t+1} f_k K_k(x)$ the coefficients f_k are nonnegative and*

$$f_0 = \frac{2}{t+1}(K_t(a))^2 \binom{n}{t};$$

(ii)

$$f(0) = \frac{(K_{t+1}(a))^2}{a} \binom{n}{t}^2 \frac{(n+1)^2}{(t+1)^2};$$

(iii) $f_k \cong \log_2(K_t(a))^2 p_{t,t}^k.$

Proof: We rewrite $f(x)$ as

$$f(x) = \frac{(K_t(a)K_{t+1}(x) - K_t(x)K_{t+1}(a))^2}{a-x}$$

and use the Christoffel-Darboux formula

$$\frac{K_t(y)K_{t+1}(x) - K_t(x)K_{t+1}(y)}{y-x} = \frac{2\binom{n}{t}}{t+1} \sum_{j=0}^{t} \frac{K_j(x)K_j(y)}{\binom{n}{j}}.$$

Then by (12.1)

$$\binom{n}{k} f_k = \left\langle K_t(a)(K_{t+1}(x) + K_t(x)) \frac{2\binom{n}{t}}{t+1} \sum_{j=0}^{t} \frac{K_j(x)K_j(a)}{\binom{n}{j}}, K_k \right\rangle$$

$$= \left\langle \frac{2K_t(a)\binom{n}{t}}{t+1} \sum_{j=0}^{t} \frac{K_j(a)}{\binom{n}{j}} \sum_{i=0}^{n} (p_{t,j}^i + p_{t+1,j}^i) K_i, K_k \right\rangle$$

$$= \frac{2K_t(a)\binom{n}{t}}{t+1} \binom{n}{k} \sum_{j=0}^{t} \frac{K_j(a)}{\binom{n}{j}} (p_{t,j}^k + p_{t+1,j}^k) \geq 0.$$

From the last line we also find f_0; $f(0)$ is found from the definition of $f(x)$. This proves parts (i)-(ii).

From the last line of the above calculation we also find

$$f_k \geq \frac{2K_t(a)}{t+1} p_{t,t}^k.$$

To prove part (iii), we need a matching asymptotic upper bound on f_k. This bound indeed holds true (see [1]), though its proof is somewhat more complicated. $\qquad\square$

A. Barg

Theorem 13.2: [39] *Let $\mathcal{C}$ be an (n, M, d) binary code. Then*

$$M \leq \frac{(n+1)^2}{2(t+1)a} \binom{n}{t},$$

where t satisfies $x_{1,t+1} < d < x_{1,t}$.

Proof: The bound follows on verifying that the polynomial $f(x)$ from the previous section satisfies the conditions of Corollary 11.5. □

This result can be improved for all finite n, d by using in Corollary 11.5 another polynomial related to but different from $f_t(x)$ [35].

Theorem 13.3:

$$R(\mathcal{H}^n; \delta) \leq h_2(\phi(\delta)), \tag{13.1}$$

$$R(\mathcal{J}^{n,\omega n}; \delta) \leq h_2\Big(\frac{1}{2}\Big(1 - \sqrt{1 - (\sqrt{4\omega(1-\omega) - \delta(2-\delta)} - \delta)^2}\Big)\Big), \tag{13.2}$$

$$R(\mathcal{S}^n; \theta) \leq \frac{1 + \sin\theta}{2\sin\theta} h_2\Big(\frac{1 - \sin\theta}{1 + \sin\theta}\Big). \tag{13.3}$$

The bound on $R(\mathcal{H}^n; \delta)$ (the so-called first MRRW bound [39]) follows easily from the previous theorem and (12.2). Bounds (13.2) [39] and (13.3) [28] are proved by repeating the above argument for Hahn and Gegenbauer polynomials, respectively.

Bound (13.1) can be improved for small δ using (13.2) in the Bassalygo-Elias inequality (5.1). The resulting estimate is called the "second MRRW bound"; it is the best upper bound for the Hamming space known to-date. Bound (13.2) is the best known for large δ; however, for small δ it is not as good as the result of Theorem 5.6. Another improvement of (13.2) is given in [44].

Although it is not clear if the second MRRW bound can be improved within the frame of the polynomial method, it is shown in [44] that relying on this method it is not possible to prove tightness of the Gilbert-Varshamov bound. After more than two decades of holding the record, it is believed by many that the second MRRW bound is asymptotically the best obtainable by the polynomial method. Experimental evidence confirming this conjecture is provided in [9].

Asymptotics of the coefficients of the polynomial $f(x)$ in Theorem 13.1 calls for geometric interpretation. It is tempting to conjecture the existence

of some packing argument which would lead to the same asymptotic result as Theorem 13.2 and link it to the results of Section 16 on the covering radius, to bounds on constant weight codes and to the union bound on the error probability of decoding. At the time of writing this, any argument of this kind is missing.

14. Bounding the Distance Distribution

14.1. *Upper Bounds*

Let $\mathcal{C} \subseteq \mathcal{H}_2^n$ be a code with distance d, dual distance d' and distance distribution $(1, A_d, \ldots, A_n)$. If either d or d' are unknown, below we assume them to be zero.

The nature of the polynomial approach to bounding the distance distribution leads to asymptotic upper bounds that depend either on d or on d'. Lower bounds depend on the code rate R.

Upper bounds in terms of d' also involve R and bound the distance distribution of designs. Since a design in a finite space approximates a uniformly distributed set of points, one expects its distance distribution to approach that of the random code. Therefore in this direction one obtains bounds on the deviation of the distance profile of codes from the function $\alpha_0(R)$ (Section 3.1).

A straightforward way to bound the component A_w above is by saying

$$A_w \leq M(\mathcal{J}^{n,w}; d). \tag{14.1}$$

Better results in many cases are obtained by Theorem 11.4.

Theorem 14.1: [3] *Let $\mathcal{C}$ be a code with distance d and dual distance d':*

$$F(d, d') = \sum_{i=d}^{n} g_w(i) A_i(\mathcal{C}).$$

Let $g_w(i) = (K_w(i))^2$,

$$c = \begin{cases} \dfrac{t+1}{2} \dfrac{\binom{n-d'}{w-d'/2}}{\binom{n-d'}{t-d'/2}}, & \text{if } d'/2 \leq w \leq t, \\ 0, & \text{if } 0 \leq w \leq d'/2. \end{cases}$$

Then for sufficiently large n and $0 \leq w < t \leq \frac{n}{2}$,

$$F_w(d, d') \ \leq \ |\mathcal{C}| \left[\binom{n}{w} - c \binom{n}{t} \right] \ - \ \binom{n}{w}^2 + \frac{c}{a} \left[\binom{n}{t+1} + \binom{n}{t} \right]^2.$$

The proof is accomplished by choosing in Theorem 11.4

$$f(x) = (K_w(i))^2 - \frac{c}{a-i}(K_{t+1}(i) + K_t(i))^2 \quad (t = \frac{n}{2} - \sqrt{d(n-d)}).$$

Theorem 14.2: [3], [1] *Let $\mathcal{C}$ be a code of distance δn. Its distance profile is bounded above as follows: $A_{\xi n} \leq \exp(n(\alpha_\xi + o(1)))$, where*

$$\alpha_\xi \leq \begin{cases} h_2(\xi) + h_2(\phi(\delta)) - 1 & \delta \leq \xi \leq 1 - \delta, \\ -2I(\phi(\delta), \xi) & 1 - \delta \leq \xi \leq 1. \end{cases}$$

This bound for large δ and ξ is better than (14.1).

Let us summarize the bounds as follows: *There exist sequences of codes of rate R with distance profile $\alpha_{0,\omega} = h_2(\omega) - h_2(\delta_{GV}(R))$; for no sequence of codes of relative distance δ the weight profile can exceed the bound of Theorem 14.2.*

Theorem 14.3: *Let $\mathcal{C}$ be a code with rate R and relative dual distance δ'. Let $\xi_1 = (1/2)(1 - \sqrt{\delta'(2-\delta')})$ and ξ_2 be the root of the equation*

$$R = (1-\delta')h_2\Big(\frac{\phi(\xi) - \delta'/2}{1-\delta'}\Big) + 1 + \xi - h_2(\phi(\xi)),$$

or 0, whichever is greater. For any $\xi \in [\min(\xi_1, \xi_2), 1/2]$ and sufficiently large n the distance profile of the code $\mathcal{C}$ approaches $\alpha_{0,\xi}$.

Remark. Similar results can be obtained for $\mathscr{J}^{n,w}$ and $\mathscr{H}_q^n$; for $\mathscr{S}^n$ the dual distance of codes is not well defined.

14.2. *Lower Bounds*

We give an example of results in this direction, proved by Corollary 11.7.

Theorem 14.4: [36] *Let $\mathcal{C} \subseteq \mathscr{H}_2^n$ be a code of rate R and let $0 \leq \beta \leq h_2^{-1}(R)$. For sufficiently large n there exists a value of $\xi \in [0, \phi(\beta)]$ such that the distance profile of $\mathcal{C}$ satisfies*

$$\alpha_\xi \geq R - h_2(\beta) - I(\beta, \xi).$$

The interpretation of this theorem is as follows: For a code of rate R and any $s \geq \phi(h_2^{-1}(R))$ there exists a value ξ of the relative distance such that the average number of neighbors of a codeword is $\exp[n(R - h_2(\beta) - I(\beta, \xi))]$ or greater. Note that $\phi(h_2^{-1}(R))$ is the value of the distance from Theorem 13.3. Further results in this direction are found in [36], [2].

15. Linear Codes with Many Light Vectors

The results of the previous section do not resolve the following question: do there exist sequences of linear codes whose number of codewords of minimum weight grows exponentially in n ? It was conjectured in [29] that the answer is negative. This conjecture was disproved in [4], where it is shown that such code families do exist.

Theorem 15.1: *Let*

$$E_q(\delta) := h_2(\delta) - \frac{\log q}{\sqrt{q} - 1} - \log \frac{q}{q - 1}.$$

Let $q = 2^{2s}, s = 3, 4, \ldots$ be fixed and let $\delta_1 < \delta_2$ be the zeros of $E_q(\delta)$. Then for any $0 < \delta_1 < \delta < \delta_2 < 1/2$ there exists a sequence of binary linear codes $\{\mathcal{C}_i, i = 1, 2, \ldots\}$ of length $n = qN, N \to \infty$, and distance $d_i = n\delta/2$ such that

$$\log A_{d_i} \geq N E_q(\delta) - o(N).$$

The idea of the proof is as follows. Let X be a (smooth projective absolutely irreducible) curve of genus g over $\mathbb{F}_q$, $q = 2^{2s}$. Let $N = N(X) := \sharp X(\mathbb{F}_q)$ be the number of $\mathbb{F}_q$-rational points of X and suppose that X is such that $N \geq g(\sqrt{q} - 1)$ (*e.g.*, X is a suitable modular curve). The set of $\mathbb{F}_q$-rational effective divisors of degree $a \geq 0$ on X is denoted by $Div_a^+(X)$. Recall that $Div_a^+(X)$ is a finite set. For $D \in Div_a^+(X)$ let $\mathcal{C} = \mathcal{C}(D)$ be an $[N, K, d(\mathcal{C})]$ geometric Goppa code constructed in a usual way [49]. Then $K \geq a - g + 1$ and $d(\mathcal{C}) \geq N - a$.

Note that once X and a are fixed, the estimates of the code parameters given above do not depend on the choice of the divisor D. It is conceivable that for some divisors D the code $\mathcal{C}(D)$ will have better parameters. That this is the case was shown by Vlăduţ in 1987 [49]. This result was obtained by computing the average parameters of codes over $Div_a^+(X)$. The same idea applies to the weight spectrum of the Goppa codes: one can compute the average weight spectrum of the code $\mathcal{C}(D), D \in Div_a^+(X)$, and then prove that there exists a code whose weight spectrum is at least as good as the average value. This code is then concatenated with a binary $[n = q - 1, 2s, q/2]$ simplex code. This results in a binary code whose number of minimum-weight codewords is given in Theorem 15.1.

16. Covering Radius of Linear Codes

The polynomial method can be used to derive bounds on the covering radius of linear codes.

Theorem 16.1: [16, p. 230] *Let $\mathcal{C}$ be a linear code with dual distance d'. Let r be an integer and $f(x) = \sum_{i=0}^{n} f_i K_i(x)$ be a polynomial such that $f(i) \le 0, i = r+1, \ldots, n,$ and*

$$f_0 > \sum_{j=d'}^{n} |f_j| A'_j.$$

Then $r(\mathcal{C}) \le r$.

Proof: Let $\mathcal{C}$ be a code of size M and $\mathcal{D}(x) = \mathcal{C} + x$ be the translation of $\mathcal{C}$ by a vector x and let $A(x) = (A_i(x), i = 0, 1, \ldots, n)$ be the *weight distribution* of $\mathcal{D}(x)$. Let $A'(x) = (A'_i(x), i = 0, 1, \ldots, n)$ be the MacWilliams transform of the vector $A(x)$:

$$A'(x) = (M)^{-1} A(x) \mathbf{K},$$

where $\mathbf{K}$ is the Krawtchouk matrix. Note that the components of $A'(x)$ can be negative.

Also let $\mathcal{C}'$ be the dual code of $\mathcal{C}$ and $(A'_i, 0 \le i \le n)$ be its weight distribution. It is known that for any x

$$|A'_i(x)| \le A'_i \quad (i = 0, \ldots, n).$$

Therefore, compute

$$
\begin{aligned}
M^{-1} \sum_{i=0}^{n} f(i) A_i(x) &= M^{-1} \sum_{i=0}^{n} \sum_{j=0}^{n} f_j K_j(i) A_i(x) \\
&= \sum_{j=0}^{n} f_j M^{-1} \sum_{i=0}^{n} A_i(x) K_j(i) \\
&= \sum_{j=0}^{n} f_j A'_j(x) = f_0 + \sum_{j=d'}^{n} f_j A'_j(x) \\
&\ge f_0 - \sum_{j=d'}^{n} |f_j| A'_j.
\end{aligned}
$$

If the last expression is positive, then so is the sum $\sum_{i=0}^{n} f(i)A_i(x)$. Further, if $f(i) < 0$ for $i = r + 1, \ldots, n$, then there exists an $i \leq r$ such that (for every x) $A_i(x) > 0$, *i.e.*, $r(\mathcal{C}) \leq r$. □

Applications of this theorem rely on bounds on the dual weight distribution of a linear code with dual distance d', *i.e.*, of the weight distribution (A_i) of a linear code with distance d'. Since $A_i < M(J^{n,i}; d')$, any upper bound on the size of a constant weight code can be used to obtain a bound on $r(\mathcal{C})$. Good bounds are obtained if we again take the polynomial $W_t(x)$ (Section 13). Another possibility is to use the results of Section 13 and use asymptotics of its coefficients from Theorem 13.1(iii). These methods yield the best asymptotic upper bounds on r in terms of δ' currently known (see [1] and references therein).

Further reading. In many cases this article provides only a starting point of a large topic in coding and information theory. A lot more information is found in the books and articles referenced in the main text. In addition, the following textbooks or monographs offer a general introduction and more expanded context pertinent to the subjects of this chapter: combinatorial coding theory [5], [38], [50], error exponents [20], [51], algebraic-geometric codes [47], [49], tables of short codes and general reference source [41], covering radius [16], spherical codes and bounds [18], [24], orthogonal polynomials [40], [48].

References

[1] A. Ashikhmin and A. Barg, *Bounds on the covering radius of linear codes*, Des. Codes Cryptogr., to appear.

[2] A. Ashikhmin, A. Barg, and S. Litsyn, *A new upper bound on the reliability function of the Gaussian channel*, IEEE Trans. Inform. Theory **46** (2000), no. 6, 1945–1961.

[3] ———, *Estimates of the distance distribution of codes and designs*, IEEE Trans. Inform. Theory **47** (2001), no. 3, 1050–1061.

[4] A. Ashikhmin, A. Barg, and S. Vlăduţ, *Linear codes with exponentially many light vectors*, Journal of Combin. Theory, Ser. A **96** (2001), no. 2, 396–399.

[5] E. F. Assmus, Jr. and J. D. Key, *Designs and Their Codes*, Cambridge Tracts in Mathematics, vol. 103, Cambr. Univ. Press, Cambridge, 1992.

[6] A. Barg, *On error bounds for spherical and binary codes*, preprint, 2001.

[7] ———, *On some polynomials related to the weight enumerators of linear codes*, SIAM J. Discrete Math. **15** (2002), no. 2, 155–164.

[8] A. Barg and G. D. Forney, Jr., *Random codes: Minimum distances and error exponents*, IEEE Trans. Inform. Theory **48** (2002), no. 9.

[9] A. Barg and D. B. Jaffe, *Numerical results on the asymptotic rate of binary codes*, Codes and Association Schemes (A. Barg and S. Litsyn, eds.), DIMACS series, vol. 56, AMS, Providence, R.I., 2001, pp. 25–32.

48   *A. Barg*

[10] A. Barg and D. Nogin, *Bounds on packings of spheres in the Grassmann manifolds*, IEEE Trans. Inform. Theory **48** (2002), no. 9.

[11] L. A. Bassalygo, S. I. Gelfand, and M. S. Pinsker, *Simple methods for deriving lower bounds in the theory of codes*, Problemy Peredachi Informatsii **27**, no. 4, 3–8 (in Russian) and 277–281 (English transaltion).

[12] E. R. Berger, *Some additional upper bounds for fixed-weight codes of specified minimum distance*, IEEE Trans. Inform. Theory **13** (1967), no. 2, 307–308.

[13] E. R. Berlekamp, *The performance of block codes*, Notices of the AMS (2002), no. 1, 17–22.

[14] R. E. Blahut, *Hypothesis testing and information theory*, IEEE Trans. Inform. Theory (1974), no. 4, 405–417.

[15] T Britz, *MacWilliams identities and matroid polynomials*, Electron. J. Combin. **9** (2002), #R19.

[16] G. Cohen, I. Honkala, S. Litsyn, and A. Lobstein, *Covering Codes*, Elsevier Science, Amsterdam, 1997.

[17] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, *Packing lines, planes, etc.: Packings in Grassmannian spaces*, Experimental Mathematics **5** (1996), no. 2, 139–159.

[18] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, Springer-Verlag, New York-Berlin, 1988.

[19] I. Csiszár, *The method of types*, IEEE Trans. Inform. Theory **44** (1998), no. 6, 2505–2523, Information theory: 1948–1998.

[20] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Channels*, Akadémiai Kiadó, Budapest, 1981.

[21] P. Delsarte, *Bounds for unrestricted codes, by linear programming*, Philips Res. Rep. **27** (1972), 272–289.

[22] ———, *An algebraic approach to the association schemes of coding theory*, Philips Research Repts Suppl. **10** (1973), 1–97.

[23] P. Delsarte, J. M. Goethals, and J. J. Seidel, *Spherical codes and designs*, Geometriae Dedicata **6** (1977), 363–388.

[24] T. Ericson and V. Zinoviev, *Codes on Euclidean Spheres*, Elsevier Science, Amsterdam e. a., 2001.

[25] R. G. Gallager, *Low-Density Parity-Check Codes*, MIT Press, Cambridge, MA, 1963.

[26] ———, *Information Theory and Reliable Communication*, John Wiley & Sons, New York e.a., 1968.

[27] M. Greferath and S. E. Schmidt, *Finite-ring combinatorics and MacWilliams' equivalence theorem*, J. Combin. Theory Ser. A **92** (2000), no. 1, 17–28.

[28] G. Kabatyansky and V. I. Levenshtein, *Bounds for packings on the sphere and in the space*, Problemy Peredachi Informatsii **14** (1978), no. 1, 3–25.

[29] G. Kalai and N. Linial, *On the distance distribution of codes*, IEEE Trans. Inform. Theory **41** (1995), no. 5, 1467–1472.

[30] T. Kløve, *Support weight distribution of linear codes*, Discrete Mathematics **106/107** (1992), 311–316.

[31] I. Krasikov, *Nonnegative quadratic forms and bounds on orthogonal polynomials*, J. Approx. Theory **111** (2001), no. 1, 31–49.

[32] I. Krasikov and S. Litsyn, *Survey of binary Krawtchouk polynomials*, Codes

and Association Schemes (Piscataway, NJ, 1999) (A. Barg and S. Litsyn, eds.), Amer. Math. Soc., Providence, RI, 2001, pp. 199–211.

[33] V. I. Levenshtein, *Upper-bound estimates for fixed-weight codes*, Problemy Peredachi Informatsii **7** (1971), no. 4, 3–12.

[34] _____, *On the minimal redundancy of binary error-correcting codes*, Information and Control **28** (1975), no. 4, 268–291, Translated from the Russian (Problemy Peredachi Informatsii **10** (1974), no. 2, 26–42).

[35] _____, *Krawtchouk polynomials and universal bounds for codes and designs in Hamming spaces*, IEEE Trans. Inform. Theory **41** (1995), no. 5, 1303–1321.

[36] S. Litsyn, *New upper bounds on error exponents*, IEEE Trans. Inform. Theory **45** (1999), no. 2, 385–398.

[37] F. J. MacWilliams, *A theorem in the distribution of weights in a systematic code*, Bell Syst. Techn. Journ. **42** (1963), 79–94.

[38] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, 3rd ed., North-Holland, Amsterdam, 1991.

[39] R. J. McEliece, E. R. Rodemich, H. Rumsey, and L. R. Welch, *New upper bound on the rate of a code via the Delsarte-MacWilliams inequalities*, IEEE Trans. Inform. Theory **23** (1977), no. 2, 157–166.

[40] A. F. Nikiforov, S. K. Suslov, and V. B. Uvarov, *Classical Orthogonal Polynomials of a Discrete Variable*, Springer-Verlag, Berlin, 1991.

[41] V. Pless and W. C. Huffman (eds.), *Handbook of Coding Theory*, vol. 1,2, Elsevier Science, Amsterdam, 1998, 2169 pp.

[42] G. Sh. Poltyrev, *Bounds on the decoding error probability of binary linear codes via their spectra*, IEEE Trans. Inform. Theory **40** (1994), no. 4, 1284–1292.

[43] _____, *On coding without restrictions for the AWGN channel*, IEEE Trans. Inform. Theory **40** (1994), no. 4, 409–417.

[44] A. Samorodnitsky, *On the optimum of Delsarte's linear program*, Journal of Combin. Theory, Ser. A **96** (2001), no. 2, 261–287.

[45] C. E. Shannon, *Probability of error for optimal codes in a Gaussian channel*, Bell Syst. Techn. Journ. **38** (1959), no. 3, 611–656.

[46] J. Simonis, *MacWilliams identities and coordinate partitions*, Linear Alg. Appl. **216** (1995), 81–91.

[47] S. A. Stepanov, *Codes on Algebraic Curves*, Kluwer Academic, New York, 1999.

[48] G. Szegö, *Orthogonal Polynomials*, Colloquium Publications, vol. 23, AMS, Providence, RI, 1975.

[49] M. Tsfasman and S. Vlăduţ, *Algebraic-Geometric Codes*, Kluwer, Dordrecht, 1991.

[50] J. H. van Lint, *Introduction to Coding Theory*, 3rd ed., Springer-Verlag, Berlin e. a., 1999, (1st ed. 1981).

[51] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill, 1979.