At the Dawn of the Theory of Codes

Alexander Barg

Is there anything of which one can say 'Look, this is new'? No, it has already existed, long before our time.

The theory of error-correcting codes was born in 1945 when C. Shannon wrote his landmark paper [1] on the mathematical theory of communication. This, of course, does not mean that there was no notion of the coding of messages before. Although this notion did not take the shape of a mathematical science, it kept producing, from time to time, instructive examples that may be still interesting to the mathematical community because they either present a surprising provisional insight or are of exceptional beauty. Below I intend to discuss some of these episodes. My aim here is not to contest the generally acknowledged priorities, nor do I claim that the discovery of these curiosities is my achievement. Rather I want to bring together a series of mathematical stories that form a part of the early history (or the prehistory) of coding theory.

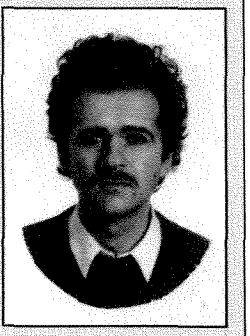
The purposes of the transformation of messages before transmission may be various: to compress the text in order not to send redundant information, or to conceal the sense of the text from an unauthorized user, or to add a few check symbols to correct possible channel errors after the transmission. The theory of error-correcting codes deals with the last problem.

Let F be a finite set (an alphabet) of size $|F| = q$. A (q-ary block) *code* A of length n is a subset of F^n. For q a prime power and $F = \mathbb{F}_q$ a finite field, a *linear code* is a linear subspace of the vector space F^n. Codes are designed for the transmission of messages over noisy channels.

A *channel* is defined as a stochastic mapping $T : F \to F$ with the matrix of transition probabilities $(p(v|u))$, $u, v \in F$, where $p(v|u) = \Pr\{v \text{ is received}|u \text{ is transmitted}\}$ (we do not use the most general definition here). Note that we assume that the information transmission channel is memoryless, i.e., the noise affects the letters of a transmitted word statistically independently. Suppose a codeword (a message) $\mathbf{a} \in A$ is to be transmitted over T letter by letter. Denote by $\mathbf{x} \in F^n$ the received word. To reconstruct a transmitted word from a received one, let us introduce the mapping $D : F^n \to A$ called the *decoder*. The goal of the decoder is to minimize the probability of decoding error, i.e., of an event $D(\mathbf{x}) \neq \mathbf{a}$. It can be shown that if the messages are equiprobable, the error probability is minimized (over all possible decoding rules) by the so-called *maximum-likelihood decoder* D_{ML} defined by the equality

$$\Pr\{\mathbf{x}|D_{\mathrm{ML}}(\mathbf{x})\} = \max_{\mathbf{a}\in A} \Pr\{\mathbf{x}|\mathbf{a}\} \quad \text{for all } \mathbf{x} \in F^n. \qquad (1)$$

Suppose F^n is endowed with a metric d that is matched to the channel in the sense that $\Pr\{\mathbf{x}|\mathbf{y}\} \geqslant \Pr\{\mathbf{z}|\mathbf{y}\}$ im-



Alexander Barg

Alexander Barg is with the mathematical coding-theoretic team of the Institute for Problems of Information Transmission in Moscow. He earned his doctorate from the same institute. His academic interests include algebraic and combinatorial methods in coding theory, complexity of computations, and studies in literary modernism.

plies $d(\mathbf{x},\mathbf{y}) \leq d(\mathbf{z},\mathbf{y})$. Then a decoding rule equivalent to Equation (1) is the so-called *minimum-distance decoder* D_{md} defined by the equality

$$d(\mathbf{x},D_{\text{md}}(\mathbf{x})) = \min_{\mathbf{a}\in A} d(\mathbf{x},\mathbf{a}) \quad \text{for all } \mathbf{x} \in F^n. \tag{2}$$

Consider for example the channel with $p(v|v) = 1 - \epsilon$ and $p(v|u) = \epsilon/(q-1)$ for $v \neq u$; all errors being equiprobable, this is called the *q-ary symmetric channel* (QSC). It is easy to check that for $q\epsilon < q - 1$ the Hamming metric $d(\mathbf{a},\mathbf{b}) = \#\{j : a_j \neq b_j\}$ is matched to this channel. For this reason, the QSC is the most popular channel model among coding theorists. For further details, we refer to the two classical treatises in modern information theory: [2], chap. 5, and [3], chap. 1.

As a rule, the algorithmic implementation of D_{ML} as well as of D_{md} requires the inspection of a large subset of code words and is, therefore, computationally intractable. For example, the general problem of minimum-distance decoding is known to be NP-hard. Because of this, one often studies the implementation of a less powerful decoding mapping, namely, the bounded-distance decoding. By

$$d(A) = \min_{\substack{a',a''\in A \\ a'\neq a''}} d(a',a''),$$

we denote the minimum distance within a code A (hereafter, *code distance*). Let $\delta = 2t + 1$ and let $S(\mathbf{a},t) = \{\mathbf{x} \in F^n | d(\mathbf{a},\mathbf{x}) \leq t\}$. The *bounded-distance decoding* D_δ is a partial mapping $D_\delta : \cup_{\mathbf{a}\in A} S(\mathbf{a},t) \to A$ defined only on vectors that are suitably close to the code. For this mapping to be well-defined, we need that the spheres $S(\mathbf{a},t)$ be disjoint, or in other words that $\delta \leq d(A)$. Usually, one says that the decoding D_δ corrects up to t errors.

Denote the triple of code parameters by $[n,M,d]$ = [length, size, Hamming distance]. If F is a finite field and a code A is linear, then $k = \log_q M$ is its dimension. In this case, one may think of a code as a linear bijective mapping $f_A : F^k \to A \subset F^n$ from the set of k-letter messages onto the set of n-letter code words. For this reason, k is sometimes called the *number of information symbols*, whereas the remaining $n - k$ symbols are redundant and provide the error correction. Usually they are called *check symbols*. In the linear case, the norm $\|\mathbf{x}\|$ that corresponds to the Hamming distance is called the *Hamming weight* and denoted by $\text{wt}(\cdot)$. So $\text{wt}(\mathbf{x}) = \#\{j : x_j \neq 0\}$.

The main problems of coding theory are related to the construction of codes with large size and distance. Evidently, these two objectives come into conflict. A natural question is: How large can the distance of a code of length n with M words be? or, how many words can there be in a code of length n and with distance d? The answer is given by the following statement.

THEOREM 1.

$$M \leq q^{n-d+1}. \tag{3}$$

Proof. Write all M code words in the rows of an $M \times n$ matrix. Delete any $d - 1$ columns. The rows of length $n - d + 1$ of the remaining submatrix are still distinct; on the other hand, there are at most q^{n-d+1} different words of this length.

This theorem was proved in [4] and is known as the Singleton bound. Singleton proved his theorem only for codes with integer k (for example, for linear codes). Surprisingly, we find the general bound (3) already in [5] that dates back to 1932. Though the authors do not provide a proof, the hints that are given lead us to think that they had in mind exactly the cited argument. In [5], the problem of constructing a maximal code was studied for a particular reason that we consider in the next section.

Commercial Codes

Historically, the first codes were intended to encipher plaintext of a dispatch in order to hide its sense from a third party. The story of these codes is presented in a more than thousand page volume [6]. However, Chapter 22 of it (entitled "Sideshows") is devoted to non-secret code systems primarily of commercial use. These codes became common already by 1825, after Claude Chappe in 1794 constructed a semaphore system linking the main cities of France with hilltop towers, signals being repeated from tower to tower. An immense impetus to the promotion of commercial codes was given in 1866 by the laying of the Atlantic cable. The problem of inevitable transmission errors was understood by the code compilers even before 1877, when the United States Supreme Court considered the case of the Philadelphia wool dealer Frank J. Primrose who sued a telegraph company for $20,000 he lost due to such an error. The codes for business transmission were constructed so as to rule out two words that differ by less than two letters. As we would say, these codes had the minimum distance 2. This is obviously insufficient to correct even a single error, though if a received word does not belong to the code, the receiver concludes that the transmission went abnormally. A code with distance 2 detects all single errors (and probably some double, triple, . . . ones).

Originally, commercial codes consisted of codewords of different length, and the restriction meant that every subset of code words of equal length had the minimum distance 2. However, in the first quarter of our century practically all modern cable and telegraph codes were based on the five-letter codeword because five-letter groupings met all the various telegraph companies' criteria of what would be counted as a word. Later, in 1923, A. C. Meisenbach published the Acme Commodity and Phrase Code that together with single

"Hamming" errors was capable of detecting single transpositions, i.e., errors of the form *abcda* → *acbda*. This means that

a word **x** *obtained from a codeword* **a** *after a substitution of one letter for another* or *after a transposition of two adjacent letters does not belong to a code.* (*)

The idea of adding transpositions to Hamming errors was quite appropriate because together these two types of errors account for more than 0.9 of all operator errors. Though the function $d^*(\mathbf{a},\mathbf{b}) = \min\{d(\mathbf{a},\mathbf{b}), I(\mathbf{a},\mathbf{b})\}$, where $I(\mathbf{a},\mathbf{b})$ is the minimum number of transpositions of adjacent symbols that turns **a** into **b**, is not a metric, it is possible to consider the following problem [7]: What is the maximum size of a code that satisfies the imposed restrictions? The Acme code A over $F = \{a, b, c, \ldots\}$ with $q = 26$, of length 5, and with $|A| = 100{,}411$ was in this sense a poor suggestion. Later, this problem was examined in [5]. Starting from Theorem 1, the authors constructed a maximum code of $26^4 = 456{,}976$ words with Hamming distance 2. After a careful hand analysis of the constructed code, they ruled out 16,925 further words to arrive at a code of 440,051 five-tuples that could detect one-place errors and single transpositions.

At present, general methods of constructing codes that detect single substitutions and single transpositions are known [8, 9]. For arbitrary $q \geq 3$ and $2 \leq n \leq q$, these methods yield codes of length n with $n - 1$ information symbols and a single check symbol. In view of (3), we may observe that the transposition detection requirement entails no additional redundancy. Let $W = \{(x_1, \ldots, x_{n-1})\}$ be the dictionary of all q^{n-1} information words. To each word we must add a check symbol x_n so that (*) holds [to construct a (*)-code]. The first two observations are obvious.

OBSERVATION 1. *Let $q = p^s$, where $s \geq 1$ for prime $p > 2$ and $s \geq 2$ for $p = 2$. Let $a_1, \ldots, a_{n-1}$ be distinct nonzero elements of $\mathbb{F}_q$. Then choosing $x_n = \sum_{i=1}^{n-1} a_i x_i$ yields a (*)-code.*

OBSERVATION 2. *If there exist methods of constructing (*)-codes for q_1 and for q_2, then there also exists such a method for $q = q_1 q_2$.*

Proof. Represent every symbol x_i, $1 \leq i \leq n - 1$, uniquely as a pair $(x_i^{(1)}, x_i^{(2)})$, where $x_i^{(1)}$ is a digit base q_1 and $x_i^{(2)}$ is a digit base q_2. Compute separately $x_n^{(j)}$ from $(x_1^{(j)}, \ldots, x_{n-1}^{(j)})$, $j = 1, 2$, where $x_n^{(j)}$ is a digit base q_j, and then restore the symbol x_n base q.

Thus, the only case left is $q = 2s$, s odd. A nice way to treat this problem was suggested in [9]. Consider the dihedral group $\mathbf{D}_s$ of order q whose elements are given by pairs (e,x) with $e \in \{-1,1\}$ and $x \in \{0, 1, \ldots, s - 1\}$. The group law is given by $(e,x) * (f,y) = (ef, ey + x)$. For $a, b \in \mathbb{Z}/(s)$, $a \neq 0$, let us define a permutation of the elements of $\mathbf{D}_s$ by

$$\tau(e,x) = (e, e(a - x) + b).$$

Finally, form

$$x_n = [\tau^{n-1}(x_1) * \tau^{n-2}(x_2) * \cdots * \tau(x_{n-1})]^{-1},$$

where $\tau^m = \tau \circ \tau^{m-1}$ and the inverse is taken in $\mathbf{D}_s$. Proving that the code thus defined satisfies (*) amounts to straightforward calculations [9].

Note that for $q = 2$, the problem of finding the maximum size of a (*)-code remains unsolved.

Decoding of BCH Codes and a System of Nonlinear Equations

In this section, we describe various approaches to the solution of a nonlinear system that attracted the attention of mathematicians during the last 200 years starting in the 1790s. The first to consider it was de Prony [10], who was led to it by an interpolation problem (this was observed in [11]; see also [12]). After that, it has been considered (seemingly, for its own sake) by Srinivasa Ramanujan [13]; finally, it arose in a problem of decoding a class of algebraic codes (see [14–16], and also [17]). Here we derive the system in question in the frame of coding theory.

Gaspard-Clair-François-Marie baron Riche de Prony (1755–1839) was a well-known French engineer in public service. In different periods of his 40-year activity he was engaged in various projects, from relating the Greenwich meridian to that of Paris to measuring the maximum power of the steam engine, and from measuring the speed of sound to compiling trigonometric tables up to 19 decimals. Therefore, it is less surprising that he once came upon the problem we are going to consider, than that he found a solution that later became the frame of reference for the decoding of a class of algebraic codes.

Let q be a prime power and suppose $F = \mathbb{F}_q$ is a finite field of q elements. Let $n = q^m - 1$, $Q = q^m$; denote by α a primitive element of $\mathbb{F}_Q$. Fix a basis of $\mathbb{F}_Q$ over $\mathbb{F}_q$. For a positive integer $\delta \geq 2$, consider the matrix

$$H = \begin{pmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1} \\ 1 & \alpha^2 & \alpha^4 & \cdots & \alpha^{2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \alpha^{\delta-1} & \alpha^{2(\delta-1)} & \cdots & \alpha^{(n-1)(\delta-1)} \end{pmatrix},$$

and substitute for every q^m-ary entry of H the corresponding q-ary column vector of length m. Denote the resulting matrix by H^*. Consider the subset B of vectors $\mathbf{a} \in F^n$ that satisfy the following linear system:

$$H^*\mathbf{a}^T = 0. \tag{4}$$

As the rows of H^* are not necessarily linearly independent, the dimension $k = \dim(B) \geqslant n - m(\delta - 1)$.

It is also possible to describe B as the set of vectors $\mathbf{a} \in \mathbb{F}_q^n$ that satisfy the following δ relations in $\mathbb{F}_Q$:

$$H\mathbf{a}^T = 0. \tag{4'}$$

Equation (4') is more convenient than Equation (4) because it enables us to use the properties of H. For example, any $\delta - 1$ distinct columns of H form a Vandermonde determinant. This implies that w, the minimum Hamming weight of a nonzero solution of Equation (4'), is greater than or equal to δ. The $[n,q^k,d \geqslant \delta]$ code B is called a (narrow-sense primitive) BCH code (after the names of its discoverers; see [18, 19]).

The BCH codes are extremely popular because of their remarkable algebraic properties ([20], Chaps. 7–11, [17], Chaps. 5–11) that allow implementing the decoding D_δ by a simple polynomial algorithm. The domain of D_δ is the set of all vectors of F^n that are at distance $\leqslant t = \lfloor(\delta - 1)/2\rfloor$ from B. For any $\mathbf{x} \in F^n$ and $\mathbf{a} \in B$, the condition $d(\mathbf{x},\mathbf{a}) \leqslant t$ implies $D_\delta(\mathbf{x}) = \mathbf{a}$, that is, D_δ corrects up to t errors.

Suppose a vector received from the channel is equal to $\mathbf{x} = \mathbf{a} + \mathbf{e}$, where $\mathrm{wt}(\mathbf{e}) = \nu \leqslant t$. Denote by $\mathbf{S} = (S_1,\ldots, S_{2t})$ the $2t$-dimensional vector over $\mathbb{F}_Q$:

$$\mathbf{S} = H\mathbf{x}^T = H\mathbf{e}^T. \tag{5}$$

Usually $\mathbf{S}$ is called a *syndrome*. Clearly, for $\mathbf{S} = 0$, we must set $D_\delta(\mathbf{x}) = \mathbf{x}$. Otherwise, consider system (5) in greater detail. Denote by $X_1 = \alpha^{i_1}, \ldots, X_\nu = \alpha^{i_\nu}$ the numbers of nonzero coordinates in $\mathbf{e}$, called the *error locations*. Suppose the values of these nonzero coordinates (errors) are $Y_1, \ldots, Y_\nu \in \mathbb{F}_q$. Then system (5) turns into

$$\begin{aligned} Y_1X_1 + \cdots + Y_\nu X_\nu &= S_1, \\ Y_1X_1^2 + \cdots + Y_\nu X_\nu^2 &= S_2, \\ &\vdots \\ Y_1X_1^{2t} + \cdots + Y_\nu X_\nu^{2t} &= S_{2t}. \end{aligned} \tag{6}$$

We are to solve system (6) in $\mathbb{F}_Q$ with respect to the unknowns X_i, Y_i and to find the correct value of $\nu \leqslant t$.

As mentioned above, system (6) has been independently considered in [10], [13] and [14, 15]. De Prony obtained it in the course of studying a problem of "curve fitting" over reals (that is, of interpolation) with respect to $2t$ equidistant points in the class of exponential functions of the form $f(x) = \Sigma_{i=1}^t Y_iX_i^x$. Finding the unknown coefficients Y_i, X_i again leads to system (6) with $\nu = t$. Surprisingly, both de Prony and Peterson, Gorenstein, and Zierler suggested one and the same method of solving system (6) that is appropriate for an arbitrary field (save one step mentioned below). The method also yields the correct value of ν.

Denote by

$$\sigma(x) = \prod_{i=1}^{\nu} (1 - xX_i) \tag{7}$$

the polynomial with roots equal to the reciprocals of the error locations X_i. We are going to find the coefficients $\sigma_1, \ldots, \sigma_\nu$ of this polynomial. Multiply Equation (7) by $Y_lX_l^{j+\nu}$ and set $x = X_l^{-1}$:

$$Y_lX_l^{j+\nu} + \sigma_1Y_lX_l^{j+\nu-1} + \cdots + \sigma_\nu Y_lX_l^j = 0. \tag{8}$$

Summing the Equations (8) over $l = 1, \ldots, \nu$ and taking into consideration system (6), we arrive at the recurrence relation

$$\sigma_1S_{j+\nu-1} + \cdots + \sigma_\nu S_j = -S_{j+\nu}. \tag{9}$$

When j runs from 1 to ν, Equation (9) generates a system of linear equations in σ_i:

$$\begin{bmatrix} S_1 & S_2 & \ldots & S_\nu \\ S_2 & S_3 & \ldots & S_{\nu+1} \\ \vdots & \vdots & & \vdots \\ S_\nu & S_{\nu+1} & \ldots & S_{2\nu-1} \end{bmatrix} \begin{bmatrix} \sigma_\nu \\ \sigma_{\nu-1} \\ \vdots \\ \sigma_1 \end{bmatrix} = - \begin{bmatrix} S_{\nu+1} \\ S_{\nu+2} \\ \vdots \\ S_{2\nu} \end{bmatrix}. \tag{10}$$

It is not difficult to prove (see [17], Theorem 7.2.2) that the determinant of this system is zero if ν is greater than the actual number of errors and nonzero if ν is equal to this number. Therefore, computing determinants of (10) for $\nu = t, t - 1, \ldots$, we can find the number of errors. Clearly, de Prony's solution did not need this step because in his problem $\nu = t$. Solving linear system (10), we find the coefficients of the polynomial $\sigma(x)$. The problem is now to find its roots. At this point, de Prony applied numerical methods for solving equations of high degree that had just been devised by Lagrange (see [12]). In coding theory, it is possible to inspect all elements of the (finite) field of constants $\mathbb{F}_Q$ to find the error locations X_i, $1 \leqslant i \leqslant \nu$ (the so-called "Chien search"). After this has been done, finding values of Y_i from the by-now linear system (6) presents no further difficulties in principle, but it requires some additional computational effort.

This drawback of the proposed solution led G. D. Forney [21] to a simplification based on a relation between $\sigma(x)$ and $S(x) = \Sigma_{i=1}^{2t} S_ix^i$. Define the *polynomial of error values*

$$\omega(x) = \sum_{i=1}^{\nu} Y_iX_ix \prod_{j\neq i} (1 - X_jx).$$

Then the following equality holds:

$$\omega(x) \equiv S(x)\sigma(x) \mod x^{2t}. \tag{11}$$

Indeed,

$$S(x)\sigma(x) = \sum_{j=1}^{2t} \sum_{i=1}^{\nu} Y_i X_i^j x^j \prod_{l=1}^{\nu} (1 - X_l x)$$

$$= \sum_{i=1}^{\nu} Y_i X_i x \prod_{l \neq i} (1 - X_l x) \left[(1 - X_i x) \sum_{j=1}^{2t} (X_i x)^{j-1} \right].$$

The term in brackets equals 1 modulo x^{2t} and Equation (11) follows.

Since deg $\omega(x) \leq t$, to find $\omega(x)$ from Equation (11), we need only the first t coefficients of $S(x)$. Once we have found $\omega(x)$, the error values Y_i are readily found from the relation

$$Y_i = \frac{\omega(X_i^{-1})}{\prod_{j \neq i}(1 - X_j X_i^{-1})} = -\frac{X_i \omega(X_i^{-1})}{\sigma'(X_i^{-1})}.$$

Thus, one can observe that Forney's algorithm for the computation of error values is, in essence, Lagrange interpolation. Finally, note that because this algorithm relies on the code construction (4), it guarantees the realization of the designated distance δ, not the true distance d.

Relation (11) plays the principal role in the decoding of BCH codes. For this reason, it is referred as the "key equation" in coding theory [16] (whereas in numerical mathematics, it is known as the Padé equation). The aim of our exposition was to remark that this relation was also central in the method of solving the system (6) proposed in 1912 by Ramanujan. His idea of finding Y_i and X_i was much the same as the one here and involved the relations between power sums ΣX_i^j and symmetric functions $\sigma_j(X_1, \ldots, X_\nu)$ known as Newton identities. Here we applied these identities in the generalized form (9). It is interesting to note that G. H. Hardy, in the foreword to the edition of Ramanujan's collected works, ranked this small paper among Ramanujan's major achievements.

Concluding Remarks. 1. It follows from (9) that one may view the problem of finding the coefficients of $\sigma(x)$ as the problem of synthesis of the shortest linear feedback shift register with feedback coefficients $-\sigma_1, \ldots, -\sigma_\nu$ that generates a given syndrome sequence. An algorithm that solves the problem in this form has been proposed in [16]. The "modern" description of this algorithm is due to J. Massey. We refer to [17] for a detailed formulation and discussion of the Berlekamp–Massey algorithm (BMA) which is computationally simpler than the procedure described above. Recently, the BMA was generalized to 2- and N-dimensional syndrome arrays [22] and applied to the decoding of algebraic-geometric codes.

2. The problem of solving the system (10) of the so-called Yule–Walker equations (which is to say, inverting Toeplitz and Hankel matrices, or solving discrete-time Wiener–Hopf equations) has been extensively studied, not only in coding theory but also in digital signal processing and in numerical methods. Since the 1947 paper [23], numerous algorithms have been suggested to solve this problem (see, e.g., W. Trench's article [24], whose algorithm is close to the BMA), many of them founded on Euclid's algorithm. The fastest version of the BMA, which applies the fast Fourier transform in finite fields, has the time complexity $O(n \log^2 n)$. We refer to [25] for an overview and comparison of these methods.

Perfect Coverings of Hamming Space and a Football-Pool Problem

Let $q = p^m$ be a prime power. Here we consider a problem of constructing a covering of the vector space $\mathbb{F}_q^n$ with Hamming metric. A subset $A \subset \mathbb{F}_q^n$ is called an *R-covering* if for every $\mathbf{x} \in \mathbb{F}_q^n$ there exists $\mathbf{a} \in A$ such that

$$d(\mathbf{a},\mathbf{x}) \leq R. \tag{12}$$

The parameter R is called the *covering radius* of A. If A is a covering, then, of course, for some $\mathbf{x}$ there may be more than one vector from A with property (12). However, if A is at the same time a code with minimum distance d and $t = \lfloor (d - 1)/2 \rfloor$ ($t \leq R$), then all the vectors $\mathbf{x}$ with $d(\mathbf{x},A) \leq t$ are covered only once. Finally, if $t = R$, the code A forms a perfect covering and for every $\mathbf{x} \in \mathbb{F}_q^n$, there exists exactly one $\mathbf{a} \in A$ such that inequality (12) holds. In this case, A is called a *perfect code*. Observe that perfect codes necessarily have odd distances.

One can easily compute the size of a perfect $[n,M,d = 2t + 1]$ code A:

$$M = |\mathbb{F}_q^n| \Big/ \#\{\mathbf{x} | d(\mathbf{x},A) \leq t\} = q^n \Big/ \sum_{i=0}^{t} \binom{n}{i} (q - 1)^i. \tag{13}$$

Because no code can be of size greater than the right-hand side of Equation (13), this equality presents an upper bound to the cardinality $|A|$ of a code of length n having distance d.

Let us give some examples of perfect codes:

1. q-ary linear $[n = (q^m - 1)/(q - 1), q^{n-m}, 3]$ Hamming code $\mathcal{H}_m$;
2. binary linear $[23, 2^{12}, 7]$ Golay code $\mathcal{G}_{23}$;
3. ternary linear $[11, 3^6, 5]$ Golay code $\mathcal{G}_{11}$.

Surprisingly, any nontrivial perfect code over a finite field either has the parameters coinciding with those of $\mathcal{H}_m$ or is equivalent to one of the Golay codes. This theorem (proved independently by two different teams of researchers in 1973) can be found in [20], Sec. 6.10.

Both Golay codes are plainly linear. For example, the code $\mathcal{G}_{11}$ is spanned by the rows of the matrix $[I_6|P]$, where

$$P = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & - & - & 1 \\ 1 & 0 & 1 & - & - \\ - & 1 & 0 & 1 & - \\ - & - & 1 & 0 & 1 \\ 1 & - & - & 1 & 0 \end{bmatrix}$$

(as usual, we write $-$ instead of -1). The ternary linear code $\mathcal{G}_{12}$ generated by the matrix $G = [I_6|S]$, where

$$S = \begin{bmatrix} 0 & \\ 1 & \\ \vdots & P \\ 1 & \end{bmatrix},$$

is not perfect but helps a lot in studying the properties of $\mathcal{G}_{11}$ by establishing a number of deep connections, some of which are mentioned at the end of this section. In fact, $\mathcal{G}_{12}$ can be obtained from $\mathcal{G}_{11}$ after extending each code word $\mathbf{a}$ by adding the overall parity check. Define the inner product $\langle \mathbf{a},\mathbf{b} \rangle = \Sigma\, a_i b_i \bmod 3$ of two vectors $\mathbf{a},\mathbf{b} \in \mathbb{F}_q^n$. A code $A \subset \mathbb{F}_q^n$ is called *weakly self-dual* (or self-orthogonal) if $A \subset A^+$ and *self-dual* if $A = A^+$. Clearly, a weakly self-dual linear code with $\dim A = n/2$ is self-dual. Because $SS^T = -I_6 \pmod 3$, every two rows of G are orthogonal. We conclude that $\mathcal{G}_{12}$ is self-dual (and so is $\mathcal{G}_{24}$, the extended binary Golay code). Therefore, the Hamming weight of every $\mathbf{a} \in \mathcal{G}_{12}$ is divisible by 3. The weight distribution of $\mathcal{G}_{12}$ is the following:

$$M_0 = 1, \quad M_6 = 264, \quad M_9 = 440, \quad M_{12} = 24, \qquad (14)$$

where M_i is the number of vectors of weight i.

It can be shown that the largest minimum weight d a self-dual $[n,3^{n/2},d]$ code A over $\mathbb{F}_3$ can have equals $3[n/12] + 3$. So the code $\mathcal{G}_{12}$ is extremal in this sense. It appears that the weight distribution of any extremal self-dual code must coincide with the one described in (14). Moreover, one can see that the 132 distinct supporting sets of the words of weight 6 in A form the Steiner system $S(5,6,12)$. From this, it is not very difficult to prove the following result.

THEOREM 2 (Theorem 98 in [27]). *A ternary self-dual* $[12,3^6,6]$ *code is unique.*

This means that any ternary self-dual [12,729,6] code can be obtained from $\mathcal{G}_{12}$ after a suitable permutation of coordinates.

The following theorem is also valid (though considerably more difficult to prove):

THEOREM 3 ([20], Sec. 20.8). *A ternary* $[12,3^6,6]$ *code is unique.*

COROLLARY. *The Golay code* $\mathcal{G}_{11}$ *is unique.*

The problem of constructing a good (not necessarily perfect) covering of the Hamming space $\mathbb{F}_3$ is also called the football-pool problem. To explain this second name, consider a lottery that invites its participants to forecast the correct results of a series of n (football or other) matches between known teams. The objective is to form a ternary n-vector $\mathbf{x}$ with coordinates equal to Win, Draw, and Lose that is close in Hamming metric to the correct vector $\mathbf{a}$ that becomes known later. A guess $\mathbf{x}$ is *good* (and is paid for) if $d(\mathbf{x},\mathbf{a}) \leq t$, where t is a threshold imposed by the lottery rules. A player is charged for each forecast.

Suppose a person has definitely decided to win or at least not to lose, that is, to complete a reasonable set A of guesses at least one of which is good. Clearly, the best choice for A would be a perfect covering of $\mathbb{F}_3^n$. This does not mean that he or she will necessarily get back more than the invested sum. (Seemingly, a better strategy would be to choose a random subset of A; we shall not go into details.)

Suppose a football pool consists of 12 matches and $t = 2$. Usually, even a lottery novice is sure about the result of one game of the pool. The code $\mathcal{G}_{11}$ provides a very good play system for the remaining 11 matches. This connection might seem strained were it not for events in 1947 in Finland, where Juhani Virtakallio published the code $\mathcal{G}_{11}$ in issues 27, 28, and 33 of the Finnish football pool magazine *Veikkaaja*. In the accompanying text he diffidently reports:

> The following system with 729 columns [=codewords] was born in my brain during a period of depression in football-pool prizes. Because the prizes were too small at that time to compensate the investments that would have been required if the system had been used week after week, the system remained unpublished and was forgotten among other systems. When during the last winter the football-pool prizes reached a peak, there was talk with the editors about publishing the system but they could not fit the 729 columns into the magazine. Only now, when I discovered a method to obtain the required saving of space, does this system get a chance to enrich the possibilities of players, and perhaps the players themselves.
>
> If the match chosen to be the sure match has been forecast correctly, the system guarantees at least 10 correct results. In the model we only present how to forecast the 11 other matches, the sure match has not been written down. (cited from [26])

Whereas the discovery of the code $\mathcal{G}_{11}$ a year and a half before it was constructed by M. Golay [28] is remarkable in itself; it becomes even more surprising in view of the corollary, as well as the fact that it is one of the few existing perfect codes over finite fields.

The Golay codes have a number of other interesting and deep properties. For example, consider the group

$(\text{Aut } \mathcal{G}_{12})^+ = (\text{Aut } \mathcal{G}_{12})/\{\pm 1\}$, where Aut $\mathcal{G}_{12}$ is the automorphism group of the Golay code $\mathcal{G}_{12}$. It is known that

$$\text{Aut } \mathcal{G}_{12} \cong \mathcal{M}_{12},$$

a Mathieu group, a 5-transitive sporadic group of order 95,040. The automorphism group of $\mathcal{G}_{11}$ is isomorphic to a subgroup of $\mathcal{M}_{12}$ that fixes a certain coordinate. $\mathcal{M}_{12}$ is a subgroup of another Mathieu group, $\mathcal{M}_{24}$, that itself is the automorphism group of the Golay code $\mathcal{G}_{24}$. This code is obtained from the code $\mathcal{G}_{23}$ after adding an overall parity check. The code $\mathcal{G}_{24}$ is closely related to the Leech lattice $\Lambda_{24} \subset \mathbb{R}^{24}$ that is known as a very good (indeed, the best possible) packing of unit spheres in 24-dimensional Euclidean space. A parti-colored mosaic of the properties of Golay codes, the Leech and kindred lattices, and related mathematical results forms the main contents of the voluminous book [30].

The football-pool problem gave birth to a number of articles [30–33] devoted to the construction of good covering codes. Suppose we know for certain that in b of n matches of the pool, one of the teams is likely not to lose. Then, in these b coordinates, only two possibilities, W and D, are left. This leads to the problem of constructing covering codes in the "mixed" space $M(b,n) = \mathbb{F}_2^b \times \mathbb{F}_3^{n-b}$. The three most recent articles in the cited list are devoted to the construction of mixed covering codes in the space $F_1 \times F_2 \times \cdots \times F_n$, where $|F_i| = q_i \geq 2$. In particular, [31] contains a large table of the best-known covering codes in $M(b,n)$ for $1 \leq n \leq 13$ and $0 \leq b \leq 13$. Needless to say, many codes cited in it originate from *Veikkaaja, Veikkaus-Lotto*, and 11 other magazines and brochures devoted to football-pool systems.

References

1. C. E. Shannon, A mathematical theory of communication I, II, *Bell Syst. Tech. J.* 27 (1948), 379–423, 623–656; reprinted in C. E. Shannon and W. Weaver, *A Mathematical Theory of Communication*, Urbana: University of Illinois Press (1949).
2. R. G. Gallager, *Information Theory and Reliable Communication*, New York: Wiley (1968).
3. I. Csiszár and J. Körner, *Information Theory. Coding Theorems for Discrete Memoryless Systems*, New York: Academic Press (1981).
4. R. C. Singleton, Maximum distance q-nary codes, *IEEE Trans. Inform. Theory* IT-10(2) (1964), 116–118.
5. W. F. Friedman and C. J. Mendelsohn, Notes on codewords, *Amer. Math. Monthly* 39 (1932), 394–409.
6. D. Kahn, *The Codebreakers: The Story of Secret Writing*, New York: Macmillian (1968).
7. G. Simmons, How good is the Acme code? *Fourth Joint Swedish-Soviet International Workshop on Information Theory, August–September 1989, Gotland, Sweden, Open Problems*, pp. 24–30.
8. J. Verhoeff, *Error Correcting Decimal Codes*, Mathematical Center Tracts No. 29, Amsterdam: Math. Centrum (1969).
9. H. P. Gumm, A new class of check-digit methods for arbitrary number systems, *IEEE Trans. Inform. Theory* IT-31(1) (1985), 102–105.
10. M. R. de Prony, Essai expérimentalle et analytique, *J. École Polytech. (Paris)* 1 (1795), 24–76.
11. J. K. Wolf, Decoding of Bose–Chaudhuri–Hochquenghem codes and Prony's method for curve fitting, *IEEE Trans. Inform. Theory* IT-13(4) (1967), 608.
12. R. Hill, Error-correcting codes I, II, *Math. Spectrum* 22(3) (1989–1990), 94–102; 23(1) (1990–1991), 14–22.
13. S. Ramanujan, Note on a set of simultaneous equations, *J. Indian Math. Soc.* 4 (1912), 94–96.
14. W. W. Peterson, Encoding and error correction procedures for the Bose–Chaudhuri codes, *IRE Trans. Inform. Theory* IT-6 (1960), 459–470.
15. D. C. Gorenstein and N. Zierler, A class of error-correcting codes in p^m symbols, *SIAM J.* 9 (1961), 207–214.
16. E. R. Berlekamp, *Algebraic Coding Theory*, New York: McGraw-Hill (1968).
17. R. Blahut, *Theory and Practice of Error Control Codes*, Reading, MA: Addison-Wesley (1984).
18. R. C. Bose and D. K. Ray-Chaudhuri, On a class of error-correcting binary group codes, *Inform. Control* 3 (1960), 68–79.
19. A. Hochquenghem, Codes correcteurs d'erreurs, *Chiffres* 2 (1959), 147–156.
20. F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, Amsterdam: North-Holland (1977).
21. G. D. Forney, On decoding BCH codes, *IEEE Trans. Inform. Theory* IT-11(4) (1965), 549–557.
22. S. Sakata, Decoding binary 2-D cyclic codes by the 2-D Berlekamp–Massey algorithm, *IEEE Trans. Inform. Theory* IT-37(4) (1991), 1200–1203.
23. N. Levinson, The Wiener RMS error criterion in filter design and prediction, *J. Math. Phys.* 25 (1947), 261–278.
24. W. F. Trench, An algorithm for the inversion of finite Toeplitz matrices, *SIAM J.* 12 (1964), 512–522.
25. Y. Sugiyama, An algorithm for solving discrete-time Wiener–Hopf equations based upon Euclid's algorithm, *IEEE Trans. Inform. Theory* IT-32(3) (1986), 394–409.
26. I. Honkala, On the early history of the ternary Golay code, an appendix to [31], unpublished manuscript.
27. V. Pless, *Introduction to the Theory of Error-Correcting Codes*, 2nd ed., New York: Wiley (1989).
28. M. Golay, Notes on digital coding, *Proc. IEEE* 37 (1949), 637.
29. J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, New York: Springer-Verlag (1988).
30. H. J. L. Kamps and J. H. van Lint, The football pool problem for 5 matches, *J. Comb. Theory, Ser. A* 3 (1967), 315–335.
31. H. Hämäläinen and S. Rankinen, Upper bounds for football pool problems and mixed covering codes, *J. Comb. Theory, Ser. A* 56 (1991), 84–95.
32. J. H. van Lint, Jr. and G. J. M. van Wee, Generalized bounds on binary/ternary mixed packing and covering codes, *J. Comb. Theory, Ser. A* 57 (1991), 130–143.
33. G. J. M. van Wee, Bounds on packings and coverings by spheres in q-ary and mixed Hamming spaces, *J. Comb. Theory, Ser. A* 57 (1991), 117–129.

Institute for Problems of Information Transmission
Ermolovoy 19
Moscow, GSP-4 101447
Russia