Problems and Conjectures in Algebraic and Combinatorial Coding Theory

Wolfgang Willems

Error correction in data transmission has challengend electrical engineering since the beginnings of the fifties of the last century when Shannon wrote his landmark paper [21]. Clearly, most of the problems which had or still have to be solved are of technical nature and they belong to the area of engineering. However, it turned out that many problems are of pure mathematical nature and meanwhile coding theory is no longer exclusively an area of engineering; at least part of it is getting more and more of interest in mathematics. In this note we collect some open problems which have exceedingly attracted mathematicians. Many of them are long standing questions and we are still far away from any complete answer.

In order to state the problems and conjectures we fix some notations. For a finite alphabet K and $n \in \mathbb{N}$ we put

$$K^n = \{a = (a_1, \dots, a_n) \mid a_i \in K\}.$$

On the space K^n the *Hamming distance* between vectors a and b is defined by

$$\mathrm{d}(a, b) = |\{i \mid a_i \neq b_i\}|.$$

A set $C \subseteq K^n$ is called an (n, M, d)-*code* or a *code with parameters* (n, M, d) if $|C| = M$ and

$$d = \min\{\mathrm{d}(c, c') \mid c, c' \in C,\ c \neq c'\}$$

is the *minimum distance of* C. In case K is a finite field and C is a K-linear subspace of K^n of dimension k we write $[n, k, d]$ instead of $(n, |C|, d)$. To indicate the size $q = |K|$ of the underlying field resp. alphabet we sometimes write $[n, k, d]_q$ resp. $(n, |C|, d)_q$.

1. Perfect codes

A code $C \subseteq K^n$ is called *perfect* if the full ambient space K^n can be covered by pairwise disjoint spheres $B_e(c) = \{a \mid a \in K^n,\ \mathrm{d}(c, a) \leq e\}$ around the code words $c \in C$. If $|K| = q$ this obviously affords the *sphere packing equation*

$$q^n = |C| \sum_{j=0}^{e} \binom{n}{j} (q-1)^j. \tag{1}$$

Note that $d = 2e + 1$ if $1 < |C| < |K|^n$. (In the perfect cases where C is trivial, i.e. $|C| = 1$ or $C = K^n$ we are not interested.) Now we may naturally ask

Problem 1. *Which parameters* $(n, |C|, d = 2e + 1)_q$ *satisfy the sphere packing equation* (1)*?*

Problem 2. *For which parameters* $(n, |C|, d = 2e + 1)_q$ *does there even exist a perfect code?*

First note that there are parameters, for instance $(90, 2^{78}, 5)_2$, which satisfy the sphere packing equation but there does not exist a perfect code with these parameters. Problem 1 seems to be one of the number theoretical problems which we easily can state but never can solve, at least not at the moment. In contrast, the solution of Problem 2 seem to be more feasible since the existence of a perfect $(n, |C|, d = 2e + 1)_q$-code implies that the Lloyd-polynomial

$$L_e(x) = \sum_{j=0}^{e} (-1)^j (q-1)^{e-j} \binom{x-1}{j} \binom{n-x}{e-j}$$

has e different integer roots in the interval $[1, n]$ (see [15]). Applying this fact one can prove

Theorem (Hong, [11]). *If* $e \geq 3$*,* $n \geq e + 1$ *and* $q \geq 3$ *then there does not exist a perfect* $(n, |C|, d = 2e + 1)_q$*-code.*

Note that for $e = 3$ and $q = 2$ we have the perfect binary $[23, 12, 7]$ Golay code, for $e = 2$ and $q = 3$ the perfect ternary $[11, 6, 5]$ Golay code and for $e = 1$ the perfect $[n = \frac{q^k-1}{q-1}, n - k, 3]_q$ Hamming code where q is a power of a prime. Furthermore, the binary repetition code of odd length is perfect. Finally, observe that are also non-linear codes with the same parameters as for the Hamming codes (see [25], p. 33).

Problem 3. *Does there exist for* $e = 1$ *resp.* $e = 2$ *perfect codes over an alphabet* K *where* $|K|$ *is not a power of a prime?*

2. The Gilbert-Varshamov bound

Let $A_q(n, d)$ be defined as the largest integer M for which a code C over an alphabet K with $|K| = q$ exists where $|C| = M$ and C is of minimum distance at least d. With the notation

$$\alpha_q(\delta) = \limsup_{n\to\infty} \frac{1}{n} \log_q A_q(n, \delta n) \quad (0 \leq \delta \leq 1)$$

we have

Theorem (Gilbert-Varshamov bound, [9], [24])

$$\alpha_q(\delta) \geq 1 - H_q(\delta) \tag{2}$$

for $0 \leq \delta \leq (q-1)/q$ *where the* entropy function $H_q(x)$ *is given by*

$$\begin{cases} 0, & \textit{for } \ x = 0 \\ x \log_q(q-1) - x \log_q x - (1-x)\log_q(1-x), & \textit{for } \ 0 < x \leq (q-1)/q. \end{cases}$$

For more than 30 years this was the best known lower asymptotic bound. Using deeper results from algebraic curves Tsfasman, Vlăduţ and Zink improved 1982 the bound to

Theorem [23], [8]. *If q is a power of a prime and a square then*

$$\alpha_q(\delta) \geq (1 - \frac{1}{\sqrt{q}-1}) - \delta \quad \textit{for} \quad 0 \leq \delta \leq 1 - \frac{1}{\sqrt{q}-1}.$$

In case $q \geq 49$ this implies that $\alpha_q(\delta) > 1 - H_q(\delta)$ for all δ in a suitable subinterval of $[0, 1 - \frac{1}{\sqrt{q}-1}]$. Meanwhile, for non-linear codes, there are even better bounds [26], [5]. However, for $q \leq 45$, no improvements of the Gilbert-Varshamov bound are known. Thus we may ask

Problem 4. *Can we improve the Gilbert-Varshamov bound for small q?*

For $q = 2$ it is conjectured that we can not. Therefore we may state

Conjecture 1. *The Gilbert-Varshamov bound is asymptotically exact for binary codes.*

We would like to mention here that the Gilbert-Varshamov bound on the size of binary codes (the results above are on the rate) is not exact due to Jiang and Vardy [12] in 2004.

3. Asymptotically good codes

A class $\mathcal{C}$ of linear codes is called *asymptotically good* if there exists a sequence of $[n_i, k_i, d_i]_q$ codes $C_i \in \mathcal{C}$ with the following properties

(i) $n_1 < n_2 < n_3 < ....$

(ii) $\frac{k_i}{n_i} > \alpha > 0$ for all i

(iii) $\frac{d_i}{n_i} > \beta > 0$ for all i.

This means that neither the rate nor the relative minimum distance of the sequence tend to 0. It is easy to show that asymptotically good codes exist. However, the construction of an explicit sequence was first given by Justesen in [13]. The class of BCH codes, which is a subclass of the class of cyclic codes, is asymptotically bad [14], and as a long standing question we have

Problem 5. *Is the class of cyclic codes asymptotically good?*

There is some evidence that there are asymptotically good sequences of binary cyclic codes for which all lengths are primes if there is any asymptotically good sequence of binary cyclic codes (see [18]). A canonical class of cyclic codes for which all lengths are primes is the class of quadratic residue codes. Thus we may ask

Problem 6. *Is the class of binary quadratic residue codes asymptotically good?*

At this point note that 2-quasi-cyclic binary codes achieve asymptotically a bound better than the Gilbert-Varshamov bound (2), a recent result of Gaborit and Zemor [7]. If we add the assumption that all codes in the sequence have an automorphism group which acts transitively on the positions the class is still asymptotically good [19]. Note that 2-quasi-cyclic transitive codes are very close to cyclic codes.

4. MDS codes

A linear $[n, k, d]$-code C is an MDS-code (*Maximum Distance Separable*) if $d = n-k+1$. Thus MDS codes attain the Singleton bound. Canonical examples for MDS codes are the Reed-Solomon codes or their generalized versions. We put

$$m(k, q) = \max\{n \mid \text{there is an } [n, k]_q \text{ MDS code}\}.$$

For $k \geq q$ one easily proves that $m(k, q) = k + 1$. All known examples indicate that the following conjecture may have an affirmative answer.

Conjecture 2. *For $2 \leq k \leq q-1$ we have*

$$m(k, q) = \begin{cases} q+2, & \textit{if } q \textit{ is even and } k = 3 \textit{ or } k = q-1 \\ q+1, & \textit{otherwise.} \end{cases}$$

The conjecture has been proved by Segre, Casse, Hirschfeld et al. for $k \leq 5$, $q \leq 27$, $\sqrt{q} > 4k-9$ and q odd and in many other cases. In particular the exceptional case $m(k, q) = q+2$ occurs exactly for $k = 3$ and $k = q-1$ in even characteristic.

MDS codes are strongly related to *arcs* in finite geometry. For $k \geq 2$, an arc in the projective space $PG(k-1, q)$ is a set of $n \geq k$ points in $PG(k-1, q)$ such that no subset of k points is contained in a hyperplane. With this notation we may reformulate the conjecture to

Conjecture 2'. *If $\mathcal{A}$ is a maximal arc in the projective space $PG(k-1, q)$ with $3 < k < q-1$ then $|\mathcal{A}| = q+1$.*

Part of the difficulty in proving Conjecture 2' goes back to the fact that there are arcs which are not normal rational curves. Note that Reed-Solomon codes are rational curves from the point of view of geometry.

5. Extremal Codes

By results of Mallows-Sloane [17] and Rains [20], a binary self-dual code C of length n and minimum distance d satisfies the inequality

$$d \leq 4\left\lfloor \frac{n}{24} \right\rfloor + 4, \tag{3}$$

with the exception

$$d \leq 4\left\lfloor \frac{n}{24} \right\rfloor + 6 \quad \text{if } n \equiv 22 \bmod 24. \tag{4}$$

In case equality holds in (3) or (4) we say that C is *extremal*. The weights of code words of a self-dual code are always divisible by 2. Thus we may distinguish between *singly-even* codes, i.e. there are code words of weight not divisible by 4 and *doubly-even* codes in which all weights are multiples of 4. The length n of a doubly-even code is always divisible by 8, due to Gleason's Theorem [10]. Furthermore, extremal codes do not exist for large n as Mallows and Sloane have already observed in [17]. However, they did not state any explicit bound. By results of Ma [16] (based on invariant theory) and Duursma [4] (more elementary) we know that an extremal doubly-even code has a length bounded by $n \leq 3928$. The largest extremal code which has been constructed so far is of length $n = 136$. Thus there is a big gap been the known bound and what we are able construct.

Problem 7. *Can we improve for extremal doubly-even codes the bound on the length, i.e.* $n = 3928$*?*

Since our definition of extremal singly-even codes differs from earlier definitions given in the literature no explicit bound for such codes seem to be known.

Problem 8. *Can we find a good bound for extremal singly-even codes?*

For several reasons extremal codes of length $n = 24m$ are of particular interest (see also section 6). They are always doubly even, by [20]. The best bound on the length we have here (see [27] or [4]) is $n \leq 3672$ or equivalently $m \leq 153$, but we know only two examples. For $m = 1$ the $[24, 12, 8]$ extended Golay code with the simple automorphism group M_{24} (a Mathieu group) and for $m = 2$ the $[48, 24, 12]$ extended quadratic residue code with the simple automorphism group $\mathrm{PSL}(2, 47)$. Already 1973 Sloane asked in [22] the question

Problem 9. *Does there exist an extremal code of length* 72*?*

Attempting to construct such a code a non-trivial automorphism might be helpful. In contrast to the two known examples $m = 1$ and $m = 2$ the automorphism group - if the code exists - is solvable of order ≤ 36 and the only primes dividing $|\mathrm{Aut}(C)|$ are $2, 3, 5$ and 7 (see [2]). The primes which may occur in the order of the automorphism group of a putative extremal code of length 96 are $2, 3$ and 5 (see [3]). However, in both cases the automorphism group still might be trivial. This leads to

Problem 10. *Is the automorphism group of an extremal code of length* $24m$ *for* $m \geq 3$ *trivial?*

If so, then codes - if they exist - do not carry symmetries. They are purely combinatorial objects and then it may be extremly difficult to construct them.

6. Codes and combinatorial t-designs

Constructing combinatorial objects like t-designs is often a challenging task. A pair $(S, \mathcal{B})$ is called a *t-design*, more precisely a t-(v, k, λ) *design*, if the following three conditions hold. (i) S is a finite set of size $|S| = v$. (ii) $\mathcal{B}$ is a set of subsets of S which are called blocks, each block of size k. (iii) Any t-subset of S is contained in exactly λ blocks.

For particular parameters such designs naturally occur in linear codes. For instance,

Theorem (Assmus-Mattson, [1]). Let C be an extremal binary self-dual and doubly-even code of length $n = 24m$. Then, for all integers i with $4 \mid i$ and $4m + 4 \leq i \leq n - (4m + 4)$, the code words of fixed weight i form a 5-design.

Let C be a code as in the Assmus-Mattson Theorem. Then the parameters of C are $[24m, 12m, 4m + 4]$. Furthermore, by Mallows-Sloane [17], the number of codewords of weight $4m + 4$ is

$$\binom{n}{5}\binom{5m-2}{m-1}/\binom{4m+4}{5},$$

and one easily sees that the corresponding 5-design has parameters

$$5 - (24m, 4m + 4, \binom{5m-2}{m-1}).$$

As mentioned in the previous section we do not know whether an extremal self-dual doubly-even code of length $n = 24m$ exists for $m > 2$. Thus we may ask

Problem 11. *Does there exist* 5-$(24m, 4m + 4, \binom{5m-2}{m-1})$ *designs for* $m > 2$*?*

In case $m = 1$, i.e. for the Golay code, the Assmus-Mattson Theorem yields a 5-$(24, 8, 1)$ design. A t-$(v, k, 1)$ design (i.e. $\lambda = 1$) is called a *Steiner system* which we denote by $S(v, k, t)$. Till this day no Steiner systems are known for $t > 5$. Thus we have

Problem 12. *Does there exist an* $S(v, k, t)$ *Steiner system for* $t > 5$*?*

For $x = (x_1, \ldots, x_n) \in K^n$ the support of x is defined by $\{i \mid x_i \neq 0\}$. With this notation one might try to attack Problem 12 as follows.

Theorem (Faldum-Willems, [6]) *Let* C *be a self-dual* $[2d, d, d]$ *code over a finite field* F_q *which does not have code words of weight* $d + 1 \geq 3$*. Then the supports of all codewords of weight* d *form a Steiner system with* $t = d - 1$*.*

In [6] it has been shown that the weight distribution of a self-dual $[2d, d, d]$ code with no code words of weight $d + 1$ is uniquely determined. Furthermore, $d+1$ must be a prime, hence $d = p-1$. As a consequence of the Griesmer bound the minimum distance of a $[2d, d, d]$ code over $\mathbb{F}_q$ with $d \geq 2$ satisfies $d \leq 2q$. The case $d = 2q$ leads up to equivalence to the extended binary $[8, 4, 4]$ Hamming code or to the extended ternary $[12, 6, 6]$ Golay code. The corresponding Steiner systems are $S(8, 4, 3)$ and $S(12, 6, 5)$. In order to find Steiner systems with $t > 5$ we may therefore assume that $5 < d - 1 = p - 2$, hence $p \geq 11$. This leads to

Problem 13. *Does there exist a self-dual* $[2p - 2, p - 1, p - 1]$ *code over a finite field which has no codewords of weight* p *where* $p \geq 11$ *is a prime.*

Literatur

[1] E.F. Assmus and H.F.Mattson. New 5-designs. *J. Comb. Theory* **6**, 122-151 (1969).

[2] S. Bouyuklieva, E.A. O'Brien and W. Willems. The automorphism group of a binary self-dual doubly even $[72, 36, 16]$ code is solvable. *IEEE Transactions on Inform. Theory* **52**, 4244-4248 (2006).

[3] R. Dontcheva. Binary self-dual codes with automorphism of composite order. *IEEE Trans. Inform. Theory* **48**, 557-561 (2002).

[4] I.M. Duursma. Extremal weight enumerators ans ultraspherical polynomials. *Discrete Math.* **268**, 103-127 (2003).

[5] N. Elkies. Still better codes from modular curves. **arXiv***:math.NT/0308046.* 2003

[6] A. Faldum and W. Willems. Codes of small defect. *Designs, codes and cryptography.* **10**, 341-350 (1997).

[7] P. Gaborit and G. Zemor. Asymptoptic improvement of the Gilbert-Varshamov bound for binary linear codes. *Proceed. ISIT 2006, Seattle.* 287-291 (2006).

[8] A. Garcia and H. Stichtenoth, A tower of Artin-Schreier extensions of function fields attaining the Drinfeld-Vlăduţ bound. *Invent. Math.* **121**, 211-222 (1995).

[9] E.N. Gilbert. A comparison of signalling alphabets. *Bell Syst. Tech. J.* **31**, 504-522 (1952).

[10] A.M. Gleason. Weight polynomials of self-dual codes and the MacWilliams identities. In *Actes Congrès Internat. Math.* Tome **3**, 211-215 (1970).

[11] Y. Hong. On the nonexistence of nontrivial perfect e-codes and tight $2e$-designs in Hamming schemes $H(n, q)$ with $e \geq 3$ and $q \geq 3$. *Graphs and Combinatorics* **2**, 145-164 (1986).

[12] T. Jiang and A. Vardy. Asymptotic improvement of the Gilbert-Varshamov bound on the size of binary codes. *IEEE Trans. Inform. Theory.* **50**, 1655-1664 (2004).

[13] J. Justesen. A class of constructive asymptotically good algebraic codes. *IEEE Trans. Inform. Theory* **18**, 652-656 (1972).

[14] S. Lin and E.J. Weldon. Long BCH codes are bad. *Inform. and Control.* **11**, 445-451 (1967).

[15] S.P. Lloyd. Binary block coding. *Bell System Tech. J.* **36**, 517-535 (1957).

[16] X. Ma. Nonexistence of extremal doubly even self-dual codes with large length. *Discrete Math.* **185**, 265-274 (1998).

[17] C.L. Mallows and N.J.A.Sloane. An upper bound for self-dual codes. *Inform. and Control* **22**, 188-200 (1973).

[18] C. Martínez-Pérez and W. Willems. Is the class of cyclic codes asymptotically good? *IEEE Trans. Inform. Theory* **52**, 696-700 (2006).

[19] C. Martínez-Pérez and W. Willems. Self-dual doubly-even 2-quasi-cyclic transitive codes are asymptotically good. *IEEE Trans. Inform. Theory* **53**, 4302-4308 (2007).

[20] E.M. Rains. Shadow bounds for self-dual codes. *IEEE Trans. Inform. Theory.* **44**, 134-139 (1998).

[21] C. Shannon. A mathematical theory of communication. *Bell System Tech. J.* **27**, 379-423, 623-656 (1948).

[22] N.J.A. Sloane. Is there a $(72, 36), d = 16$ self-dual code? *IEEE Trans. Inform. Theory* **19**, 251 (1973).

[23] M.A. Tsfasman, S.G. Vlăduţ and T. Zink. Modular curves, Shimura curves, and Goppa codes, better than Varshamov-Gilbert bound. *Math. Nachrichten* **109**, 21-28 (1982).

[24] R.R. Varshamov. Estimate of the number of signals in error correcting codes. *Dokl. Akad. Nauk SSSR* **117**, 739-741 (1957).

[25] W. Willems. Codierungstheorie. DeGruyter, Berlin/New York, 1999.

[26] C. Xing. Nonlinear codes from algebraic curves improving the Tsfasman-Vlăduţ-Zink bound. *IEEE Trans. Inform. Theory.* **49**, 1653-1657 (2003).

[27] S. Zhang. On the nonexistence of extremal self-dual codes. *Disc. Appl. Math.* **91**, 103-127 (1999).