Mathematische Aspekte der Codierungstheorie

Wolfgang Willems

I. Einleitung

Die Theorie der fehlerkorrigierenden Codes wurde im Jahre 1945 geboren, als Claude Shannon die wegweisende Arbeit *A mathematical theory of communication* [32] schrieb. Sie beschäftigt sich mit der Sicherung von Daten gegen zufällige Fehler bei der Übertragung über einen unzuverlässigen Kanal. In seiner Arbeit hat Shannon, grob gesprochen, folgendes gezeigt:

Man kann Information bis zu einer gewissen Rate, die von der Kapazität des Kanals abhängt, fast ohne Informationsverlust übertragen, d.h. mit beliebig kleiner Fehlerwahrscheinlichkeit.

Der Nachteil dieses bahnbrechenden Ergebnisses, an das vor Shannon niemend so recht geglaubt hat: es beruht auf einem reinen Existenzbeweis und macht keinerlei Aussage, wie man dies praktisch bewerkstelligen kann.

In ihrem historischen Überblicksvortrag [25], den Jessie MacWilliams 1968 in Wisconsin gegeben hat, zitiert sie Marcel Golay, ebenfalls einer der Väter der Theorie:

The upper bound given by Shannon to the transmission capacity of a noisy discrete channel has challenged the mathematicians, who have accepted the challenge, to devise error correcting codes.

Dann fährt sie, die Mathematikerin, fort:

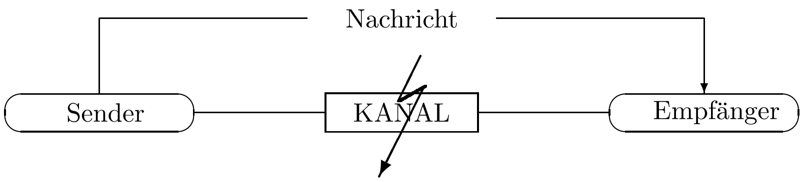
I think we all agree with the definition contained in Golay's statement and all disagree with that what he says about mathematicians. Very few mathematicians have even noticed the challenge.

MacWilliams hatte recht, und es hat lange gedauert, bis wir Mathematiker die Herausforderung akzeptiert haben. Dies mag zum Teil daran gelegen haben, dass *Diskrete Mathematik* damals noch nicht im Trend der Zeit lag, aber auch daran, dass wir nicht die große Bandbreite der mathematischen Methoden erahnt haben, die heute diesem Gebiet zugrunde liegt. Dass neben ansprechender Mathematik der direkte Anwendungsbezug hinzukommt, hat sicherlich dazu beigetragen, dass die Codierungstheorie zu einem festen Bestandteil der Mathematik geworden ist.

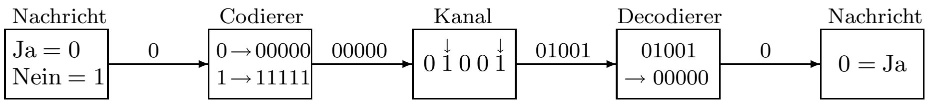
In diesem Beitrag können wir natürlich nur auf wenige Facetten der Codierungstheorie eingehen; insbesondere sagen wir nichts zu technischen Fragestellungen. Was die Codierungstheorie im täglichen Leben zu leisten vermag, kennt jeder (aber nur wenige wissen es) von seinem CD-Player [22], der etwa 4000 in Folge zerstörte Audiobits rekonstruieren kann, eine enorme Leistung. Uns interessiert die Mathematik, insbesondere die Frage: *Was kann die Mathematik in der Codierunstheorie leisten?* und ebenso: *Was bringen codierungstheoretische Methoden für die Mathematik?*

II. Problemstellung

Über einen Kanal (wie Telefonleitung, Atmosphäre, magnetisches Band, CD) sollen digitale Daten von einem Sender zu einem Empfänger übertragen werden. Dabei verursache der Kanal zufällige Störungen in den Daten, bedingt etwa durch atmosphärisches Rauschen, Interferenzen, Änderung der Magnetisierung, Kratzer auf der CD usw.



Aufgabe der Codierungstheorie ist es, die bei der Übertragung gestörten Daten zu korrigieren. Ein naives Verfahren, welches man anwenden könnte, ist die n-fache Wiederholung der gesendeten Nachricht.



In diesem Beispiel wiederholt der Codierer die Nachricht 0 fünfmal. Der Kanal stört das gesendete Wort 00000 zu 01001. Der Decodierer entschlüsselt nun das empfangene Wort 01001 zu dem *Codewort*, welches zu 01001 am nächsten liegt, also 00000.

Durch die 5-fache Wiederholung kann man, wie man unmittelbar sieht, bis zu zwei Fehler korrigieren. Ein derartiger *Wiederholungscode* ist in der Regel zu aufwändig, d.h. für die Praxis zu teuer. Will man e Fehler mit diesem Verfahren korrigieren, so muß $n \geq 2e+1$ sein. Ein Bit Information erfordert also $n-1 \geq 2e$ *redundante* Bits. Hier drängt sich die Frage auf: *Wie kann man die Korrektur von e Fehlern effizienter bewerkstelligen?*

Die Nachricht bestehe nun aus Paaren von Bits, denen durch die folgende Festsetzung 5-Tupel als Codeworte zugeordnet werden.

Nachricht	Codewort
0 0	0 0 0 0 0
0 1	0 1 1 0 1
1 0	1 0 1 1 0
1 1	1 1 0 1 1

Die Korrektur eines Fehlers ist hier möglich, denn jedes vom Empfänger erhaltene Wort mit höchstens einem Fehler ist näher zum tatsächlich gesendeten Codewort als zu allen anderen.

Ein Wiederholungscode hingegen benötigt dafür 6-Tupel, also ein Bit mehr Redundanz.

III. Der Hammingraum, Perfektheit und Zahlentheorie

Statt binärer Folgen, wie im Abschnitt II beschrieben, wollen wir allgemeiner Folgen mit Einträgen aus einem festen endlichen Körper $K = \mathbb{F}_q$ mit q Elementen zulassen. Man beachte, dass $\mathbb{Z}/2\mathbb{Z} = \{0, 1\}$ ein solcher Körper ist; aber auch die Restklassen $\mathbb{Z}/p\mathbb{Z}$ für eine beliebige Primzahl p. Setzen wir nun

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

so können wir jedem Vektor $a = (a_1, \ldots, a_n) \in K^n$ das sogenannte *Hamming-Gewicht*

$$\mathrm{wt}(a) = |\{i \mid a_i \neq 0\}|$$

zuordnen. Den Abstand zwischen zwei Vektoren $a, b \in K^n$ messen wir mittels

$$\mathrm{d}(a, b) = \mathrm{wt}(a - b)$$

und nennen ihn den *Hamming-Abstand*. Er gibt also an, an wie vielen Stellen sich a und b unterscheiden. Durch $\mathrm{d}(\cdot, \cdot)$ wird eine Metrik auf K^n definiert.

Eine Teilmenge $C \subseteq K^n$ heißt ein $(n, M, d)_q$-Code oder Code mit den Parametern $(n, M, d)_q$, falls $|C| = M$ und

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

Den Parameter d, der den minimalen Abstand zwischen zwei verschiedenen Codeworten angibt und Auskunft über die Anzahl der möglichen Fehlerkorrekturen macht, wie wir später sehen werden, nennt man die *Minimaldistanz* von C. Die Elemente aus C heißen *Codeworte*. Ist C sogar linear, also ein k-dimensionaler Unterraum von K^n, so sprechen wir von einem $[n, k, d]_q$-Code. Gegenüber den nichtlinearen Codes haben lineare Codes den Vorteil, dass man zur Beschreibung nur eine Basis angeben muß, also im Rechner wenig Speicherplatz benötigt. Mehr noch, und dies ist entscheidend, wird die i.a. aufwändige Berechnung der Minimaldistanz d wesentlich reduziert, denn wegen $\mathrm{d}(c, c') = \mathrm{wt}(c - c')$ für $c, c' \in C$ gilt:

Die Minimaldistanz eines linearen Codes ist gleich dem minimalen Gewicht der von Null verschiedenen Codeworte.

Der Aufwand der Berechnung von d reduziert also von $M(M-1)/2$ Abstandsberechnungen $\mathrm{d}(c, c')$ für $c \neq c' \in C$ im nichtlinearen Fall auf $M - 1$ Gewichtsbestimmungen $\mathrm{wt}(c)$ für $0 \neq c \in C$. Eine Beschränkung auf lineare Codes hat natürlich zur Folge, dass man auch etwas verliert, denn zu vorgegebenen n, d, q kann man oft nichtlineare Codes finden, die mehr Elemente haben als lineare. Diesen Nachteil, der sich in den Anwendungen dadurch ausdrückt, dass bei der Verwendung linearer Codes unter Umständen mehr Redundanz $n - \log_q |C|$ übertragen wird, nimmt man häufig aufgrund anderer Vorteile, die sich insbesondere in der Schnelligkeit der Decodieralgorithmen niederschlägt, in Kauf.

In den Anwendungen tragen die Codeworte $c \in C$ die Information. Diese werden über einen Kanal, von dem wir annehmen, dass er nicht fehlerfrei arbeitet, übertragen. Für $a \in K^n$ bezeichnen wir mit $\mathrm{P}(c \mid a)$ die bedingte Wahrscheinlichkeit, dass a empfangen, aber c gesendet wurde. Unter den oft realistischen Annahmen an den Kanal, dass

- jedes $a \in K$ mit gleicher Wahrscheinlichkeit $p \leq (q-1)/q$ verfälscht wird,
- alle möglichen Verfälschungen gleich wahrscheinlich sind,
- und alle Codeworte mit gleicher Wahrscheinlichkeit benutzt werden,

kann man zeigen:

$(*)$ $\quad \mathrm{P}(c \mid a) = \max_{c' \in C} \mathrm{P}(c' \mid a)$ *genau dann, wenn* $\mathrm{d}(a,c) = \min_{c' \in C} \mathrm{d}(a,c')$.

Ein sogenannter *Maximum-Likelihood-Decodierer* sucht nun bei einem empfangenen Wort a dasjenige Codewort c, welches am wahrscheinlichsten gesendet wurde, also nach $(*)$ ein Codewort, welches bezüglich des Hamming-Abstands dem empfangenen Wort a am nächsten ist.

Für $0 \leq r \in \mathbb{R}$ sei nun

$$\mathrm{B}_r(a) = \{b \mid b \in K^n,\ \mathrm{d}(a,b) \leq r\}$$

die r-Kugel vom Radius r um $a \in K^n$. Ist C ein $(n,M,d)_q$ Code und $1+2e \leq d$ mit $e \in \mathbb{N}_0$, so sind die e-Kugeln um die Codeworte $c \in C$ offenbar paarweise disjunkt. Da jede e-Kugel genau $\sum_{j=0}^{e} \binom{n}{j}(q-1)^j$ Elemente aus K^n enthält, erhalten wir

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

Wird $c \in C$ gesendet und passieren im Kanal höchstens e Fehler, so wird das empfangene Wort a, welches in $\mathrm{B}_e(c)$ liegt, zum nächstgelegenen Codewort, also c decodiert; d.h. der Decodierer decodiert richtig. Für die Anwendungen wäre es nun sehr erfreulich, wenn die e-Kugeln den ganzen Raum überdecken würden, so dass der Decodierer stets decodieren kann und bei e Fehlern auch richtig decodiert. Dies erfordert die sogenannte *Kugelpackungsgleichung*

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

Codes, die diese Gleichung erfüllen, nennt man *perfekt*. In diesen Fällen ist $d = 2e+1$, sofern der Code nicht nur aus einem Wort besteht. Triviale Beispiele sind der volle Raum K^n, ein einzelner Vektor und der binäre Wiederholungscode mit n ungerade. Also existieren perfekte Codes zu den trivialen Parametern $(n, |C|, d)_q =$

$$(n, q^n, 1)_q,\ (n, 1, 0)_q,\ (n, 2, n)_2.$$

Die Kugelpackungsgleichung ist aber auch für die nichttrivialen Parameter

$$((q^k-1)/(q-1), q^{n-k}, 3)_q,\ (11, 3^6, 5)_3,\ (24, 2^{12}, 7)_2$$

und

$$(90, 2^{78}, 5)_2$$

erfüllt.

In der berühmten Arbeit [15], die nicht einmal eine Seite lang ist, hat Golay in den ersten drei Fällen der nichttrivialen Parameter sogar lineare Codes angegeben, die $[n = (q^k-1)/(q-1), n-k, 3]_q$-Hamming-Codes ($q, k \geq 2$ beliebig), den ternären $[11, 6, 5]_3$- und den binären $[23, 8, 7]_2$-Golay-Code. Der Hamming-Code läßt sich einfach beschreiben als der Kern einer Matrix, dessen Spalten aus Vertretern aller 1-dimensionalen Unterräume des K^k besteht. Den kleinsten nichttrivalen Hamming-Code C mit den Parametern $[7, 4, 3]_2$, der bereits von Shannon in [32] erwähnt wird, kann man wie folgt konstruieren. Schreiben wir in die sogenannte *Kontrollmatrix* H alle nichttrivialen Vektoren des $\mathbb{F}_2^3$, also

$$H = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix},$$

so gilt $C = \{c \in \mathbb{F}_2^7 \mid Hc^T = 0\} = \operatorname{Kern} H$. Da H den Rang 3 hat, ist dann $\dim C = 7 - 3 = 4$. Offenbar gilt für jedes $0 \neq c \in C$, dass $\mathrm{wt}(c) \geq 3$, und man sieht auch, dass es Vektoren vom Gewicht 3 gibt. Somit hat C die angegebenen Parameter $[7, 4, 3]_2$.

Erstaunlicherweise existiert kein binärer Code C, auch kein nicht-linearer, mit $n = 90, |C| = 2^{78}$ und $e = 2$, welches man mit einem einfachen Abzählargument nachweisen kann. Eine der ersten ernsthaften Herausforderungen an die Mathematik sind nun die natürlichen Fragen:

1. Sind die obigen Parameter die einzigen, die die Kugelpackungsgleichung erfüllen oder gibt es noch andere?

2. Zu welchen Parametern existieren perfekte Codes?

Die Antwort auf die für die Codierungstheorie zentrale Frage 2 nach der Existenz perfekter Codes hat mehr als zwanzig Jahre auf sich warten lassen. Im Jahre 1973 konnten dann Tietäväinen [37], aufbauend auf vielen Vorarbeiten von J.H. van Lint, und unabhängig davon Zinov'ev und Leont'ev [41] zeigen:

Satz 1. (Tietäväinen, van Lint; Zinov'ev, Leont'ev)
Ein nicht-trivialer perfekter Code über $K = \mathbb{F}_q$ hat die gleichen Parameter wie ein Hamming- oder Golay-Code.

Ein enttäuschendes Resultat für die fehlerkorrigierende Datenübertragung. Man hätte für die Anwendungen gerne mehr gesehen. Die Antwort auf die zahlentheoretische Frage 1 ist immer noch offen.

Während es zu den Parametern des Hamming-Codes auch nichtlineare Beispiele gibt - die ersten wurden von Vasil'ev [39] entdeckt - sind Codes zu den Parametern der beiden Golay-Codes stets linear (sofern sie den Nullvektor enthalten) und sogar äquivalent zu diesen. Dabei heißen lineare Codes $C, C' \leq K^n$ äquivalent, falls es eine Matrix $M \in \mathrm{GL}(n, K)$ gibt mit genau einem Eintrag ungleich 0 aus K in jeder Zeile und jeder Spalte, so dass $CM = C'$ gilt. Man beachte, dass die Menge derartiger Matrizen gerade die (lineare) Isometriegruppe des Hammingraums bildet.

Der Unterschied, dass man Frage 2 lösen konnte und Frage 1 bisher nicht, liegt an den Nullstellen des Polynoms

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

Bereits 1957 zeigte Lloyd [23] das folgende Resultat. Dabei ist $q \in \mathbb{N}$ beliebig, also nicht wie vorher eine Primzahlpotenz.

Satz 2. (Lloyd) *Gibt es einen perfekten $(n, |C|, d = 2e+1)_q$-Code, so hat das Polynom $L_e(x)$ genau e verschiedene ganzzahlige Nullstellen im Interval* $[1, n]$.

Nach einem Resultat von Hong [17] aus dem Jahr 1986 hat $L_e(x)$ für die Parameter $e \geq 3, n \geq e+1$ und $q \geq 3$ (q nicht notwendig Primzahlpotenz) Nullstellen, die nicht in $\mathbb{Z}$ liegen. In diesen Fällen reicht also zum Nachweis der Nichtexistenz von perfekten Codes einzig die Annahme der Ganzzahligkeit der Nullstellen von $L_e(x)$. Für $n = 90, |C| = 2^{78}$ und $e = q = 2$ hat das Lloyd-Polynom die Gestalt

$$L_2(x) = 2(x^2 - 91x + 2048).$$

Da die beiden Nullstellen nicht ganzzahlig sind, kann es also nach Satz 2 keinen binären $(90, 2^{78}, 5)$-Code geben, welches, wie oben bereits erwähnt, auch mit einem Abzählargument gezeigt werden kann.

Fragen wir nach den hier verbleibenden zentralen Problemen, so ist sicher die Frage nach den Parametern, die die Kugelpackungsgleichung erfüllen, zu nennen. Aus codierungstheoretischer Sicht bleibt aber auch das Problem der Existenz perfekter Codes mit $e = 1$ und $e = 2$, wenn wir den Körper K durch ein beliebiges endliches Alphabet $\mathcal{A}$, wobei $|\mathcal{A}|$ keine Primzahlpotenz ist, ersetzen.

IV Untere Schranken und algebraische Kurven

Bei vorgegebener Länge n und Minimaldistanz d möchte man möglichst viel Information im Code C unterbringen, um die Redundanz bei der Übertragung zu minimieren. Sei also $A_q(n, d)$ das größte M, für welches es einen Code C über $K = \mathbb{F}_q$ mit $|C| = M$ und Minimalabstand mindestens d gibt. Bereits in den fünfziger Jahren haben Gilbert und Varshamov gezeigt:

$$A_q(n, d) \geq q^n / \sum_{j=0}^{d-1} \binom{n}{j} (q-1)^j.$$

Dies sieht man so: Ist C ein Code mit $|C| = A_q(n,d)$, so liegt jedes $a \in K^n$ in einer $(d-1)$-Kugel um ein geeignetes Codewort, denn sonst könnte man a zu C hinzufügen, entgegen der Maximalität von $|C|$. Also gilt $K^n \subseteq \cup_{c \in C} \mathrm{B}_{d-1}(c)$, und somit $q^n \leq |C| \sum_{j=0}^{d-1} \binom{n}{j} (q-1)^j$.

Für $0 \leq \delta \leq 1$ setzen wir nun

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

Dies besagt: Es gibt eine Folge von Codes C_i mit Parametern $(n, |C_i|, d_i)_q$, so dass $\frac{d_i}{n} \to \delta$ und $\frac{\log_q |C_i|}{n} \to \alpha_q(\delta)$. In Worten: Die sogenannten *relativen Minimaldistanzen* $\frac{d_i}{n}$ konvergieren gegen δ und die sogenannten *Raten* $\frac{\log_q |C_i|}{n}$ gegen $\alpha_q(\delta)$. Mit etwas Analysis und der aus der Informationstheorie bekannten Entropiefunktion

$$H_q(x) = \begin{cases} 0, & \text{für } x = 0 \\ x \log_q(q-1) - x \log_q x - (1-x) \log_q(1-x), & \text{für } 0 \leq x \leq (q-1)/q \end{cases}$$

besagt die von Gilbert-Varshamov angegebene Schranke asymptotisch

Satz 3. (Gilbert-Varshamov-Schranke)

$$\alpha_q(\delta) \geq 1 - H_q(\delta) \quad \textit{für} \quad 0 \leq \delta \leq (q-1)/q.$$

Man kann nun explizit Folgen von Codes, sogar Folgen linearer Codes angeben, für welche Satz 3 mit Gleichheit gilt. Es ist offen, ob dies auch richtig ist, wenn man nur zyklische Codes betrachtet, eine der wichtigsten Klassen in den Anwendungen, da diese sich effizient mittels Schieberegister implementieren lassen. Dabei heißt ein linearer Code C zyklisch, falls mit $(c_1, c_2, \ldots, c_n) \in C$ stets auch $(c_n, c_1, \ldots, c_{n-1})$ in C liegt. Für diese Klasse weiß man nicht einmal, ob sie *asymptotisch gut* ist, d.h., ob es ein $\delta > 0$ gibt mit $\alpha_q(\delta) \geq \epsilon$ bei vorgegebenem $\epsilon > 0$, eine bereits 1966 von Assmus, Mattson und Turyn gestellte Frage.

Die asymptotische Schranke von Gilbert und Varshamov konnte für mehr als 30 Jahre nicht verbessert werden. Erst unter Benutzung tiefliegender Resultate über algebraische Kurven konnten Tsfasman, Vlăduţ und Zink 1982 in [38] zeigen, dass für gewisse δ (sogar für ein ganzes Interval) $\alpha_q(\delta)$ echt oberhalb $1 - H_q(\delta)$ liegt, sofern q ein Quadrat und $q \geq 49$ ist, genauer:

Satz 4. (Tsfasman-Vlăduţ-Zink-Schranke) *Ist q ein Quadrat, so gilt*

$$\alpha_q(\delta) \geq \left(1 - \frac{1}{\sqrt{q} - 1}\right) - \delta \ \textit{für} \ \ 0 \leq \delta \leq 1 - \frac{1}{\sqrt{q} - 1}.$$

Grundlegend für dieses tiefe Ergebnis war die von Goppa 1977 gemachte Entdeckung der algebraisch geometrischen Codes [16]. Dies sind Codes, die man mittels algebraischer Kurven gewinnen kann. Der Vorteil einer derartigen Konstruktion liegt auf der Hand: man hat den vollen Apparat der Theorie der algebraischen Kurven zur Verfügung, wie etwa den Satz von Riemann-Roch [34], der dann für die zugehörigen Codes Aussagen über die Parameter macht.

Die Idee der Goppa-Konstruktion läßt sich einfach beschreiben. Sei dazu $\mathcal{P} = \{P_1, \ldots, P_n\}$ eine Menge von n Punkten in einem geometrischen Objekt $\mathcal{X}$. Sei weiterhin L ein K-Vektorraum von Funktionen von $\mathcal{P}$ in K. Dann definiert

$$C = \{(f(P_1), \ldots, f(P_n)) \mid f \in L\}$$

einen linearen Code in K^n. Eine derartig allgemeine Vorgehensweise hat zur Folge, dass man nur schwerlich die Dimension und die Minimaldistanz von C bestimmen kann. Wählt man für $\mathcal{X}$ jedoch das Nullstellengebilde einer algebraischen Kurve, die über K definiert ist, für $\mathcal{P}$ eine Menge von K-rationalen Punkten, und für L einen Vektorraum von rationalen Funktionen auf $\mathcal{X}$, so geben Eigenschaften der Kurve Auskunft über die Dimension und die Minimaldistanz von C.

Das klassische Beispiel hierzu sind die Reed-Solomon-Codes, die übrigens auch auf CD's zu finden sind, und von Reed und Solomon in den fünfziger Jahren entdeckt wurden. Hier ist $\mathcal{X}$ die affine Gerade über dem algebraischen Abschluß $\overline{K}$ von K und $\mathcal{P} = \{\alpha_1, \ldots, \alpha_n\}$ eine Teilmenge von K. Als L wählen wir den K-Vektorraum $\mathrm{Pol}_{k-1}(K)$ der Polynome vom Grad kleiner oder gleich $k-1$. Für $k \leq n$ lassen sich die Parameter des Reed-Solomon Codes

$$\mathrm{RS}(n,k) = \{(f(\alpha_1), \ldots, f(\alpha_n)) \mid f \in \mathrm{Pol}_{k-1}(K)\}$$

direkt angeben. Diese sind $[n, k, n-k+1]_q$, denn: Die Abbildung

$$f \to (f(\alpha_1), \ldots, f(\alpha_n))$$

für $f \in \mathrm{Pol}_{k-1}(K)$ ist offenbar injektiv, also gilt $\dim \mathrm{RS}(n,k) = k$. Da ein Polynom vom Grad kleiner oder gleich $k-1$ höchstens $k-1$ Nullstellen hat, hat jeder Vektor $\neq 0$ ein Gewicht größer oder gleich $n-(k-1) = n-k+1$. Somit gilt für die Minimaldistanz $d \geq n-k+1$. Es gilt sogar Gleichheit, denn stets ist $d \leq n-k+1$, welches unmittelbar aus der Singleton-Schranke

$$A_q(n,d) \leq q^{n-d+1}$$

folgt.

Die Schranke ergibt sich aus der folgenden Beobachtung. Ist C ein $(n, M, d)_q$ Code, so ist die Projektion $\pi : C \to K^{n-d+1}$ auf die ersten $n-d+1$ Koordinaten, also $\pi(c_1, \ldots, c_n) = (c_1, \ldots, c_{n-d+1})$, injektiv.

Codes, die wie die Reed-Solomon-Codes die Singleton-Schranke erreichen, für die also $d = n-k+1$ gilt, heißen MDS-Codes (<u>m</u>aximal <u>d</u>istance <u>s</u>eparable).

Sie spielen in den Anwendungen eine wichtige Rolle, da bei gegebenem n und k wegen der Singleton-Schranke die Minimaldistanz nicht größer als $n-k+1$ werden kann. Anders gesagt: Mit MDS-Codes kann man maximal viele Fehler korrigieren. Ein Nachteil liegt jedoch auf der Hand. Die Länge n eines Reed-Solomon-Codes ist beschränkt durch die Körpergröße q (bzw. $q+1$, wenn man mit der projektiven Geraden startet und L etwas anders wählt). Dies kann man leicht umgehen, indem man für $\mathcal{X}$ statt der Geraden eine algebraische Kurve wählt. Der Einfachheit halber wollen wir $\mathcal{X}$ als eine irreduzible planare projektive nichtsinguläre Kurve über $\overline{K}$ (algebraischer Abschluß von K) annehmen, die über K definiert ist. Die Punkte auf der Kurve sind dann die Klassen $(x:y:z) \in \mathbb{P}^2(\overline{K})$ mit $f(x,y,z)=0$, wobei das homogene Polynom $f \in K[x,y,z]$ die Kurve $\mathcal{X}$ definiert. Für $\mathcal{P}$ nimmt man dann eine Menge von K-rationalen Punkten auf der Kurve, d.h. von Punkten mit Koordinaten in K. Bezeichnen wir mit $N_q(\mathcal{X})$ die Anzahl aller K-rationalen Punkte auf $\mathcal{X}$, so gilt

Satz 5. (Hasse-Weil-Schranke) $\quad |N_q(\mathcal{X})-(q+1)| \leq 2g\sqrt{q}$, *wobei g das Geschlecht der Kurve ist.*

Man kann also hoffen, dass man die Länge $n=|\mathcal{P}|$ unabhängig von der Körpergröße $q=|K|$ beliebig groß machen kann - jedenfalls deutet der Satz von Hasse-Weil darauf hin -, indem man eine geeignete Kurve von großem Geschlecht wählt. Ist die projektive Kurve $\mathcal{X}$ das Nullstellengebilde eines homogenen Polynoms vom Grad d, so besagt die Nichtsingularität von $\mathcal{X}$, dass $g=(d-1)(d-2)/2$ ist.

Sind die Kurve $\mathcal{X}$ und die K-rationalen Punkte $\mathcal{P}=\{P_1,\ldots,P_n\}$ gewählt, so haben wir weiterhin einen Vektorraum L von Funktionen von $\mathcal{P}$ in K anzugeben. Dies ist ein endlichdimensionaler K-Vektorraum von rationalen Funktionen auf $\mathcal{X}$ und hängt von einem rationalen Divisor G ab. Divisoren G sind formale endliche $\mathbb{Z}$-Linearkombinationen von Punkten P auf der Kurve. Dabei heißt $G=\sum_P n_P P$ mit $n_P \in \mathbb{Z}$ rational, falls $n_{P^q}=n_P$ für alle P gilt, wobei der Punkt $P^q=(x^q:y^q:z^q)$ der zu $P=(x:y:z)$ Frobenius-konjugierte Punkt ist. Man beachte, dass mit P auch P^q auf der Kurve liegt, da diese durch ein Polynom mit Koeffizienten in $K=\mathbb{F}_q$ definiert ist.

Ist $\mathcal{X}$ definiert über das homogene absolut irreduzible Polynom f, so ist $K[\mathcal{X}]=K[x,y,z]/(f)$ ein Integritätsring. Somit existiert der Funktionenkörper $K(\mathcal{X})=\mathrm{Quot}(K[\mathcal{X}])$. Offenbar sind die Elemente aus $K(\mathcal{X})$ Funktionen, sogenannte rationale Funktionen, von $\mathcal{X}$ in $\overline{K}\cup\{\infty\}$. Der Wert ∞ tritt dabei an Polstellen auf. Man kann nun zeigen, dass jedes $0 \neq f \in K(\mathcal{X})$ nur endlich viele Null- und Polstellen hat. Ist P Null- beziehungsweise Polstelle von f von der Ordnung $a \in \mathbb{N}_0$, so setzen wir $v_P(f)=a$ beziehungsweise $v_P(f)=-a$. Somit definiert jedes $0 \neq f$ einen Divisor $(f)=\sum_P v_P(f)P$, einen sogenannten Hauptdivisor. Setzen wir noch $G=\sum_P n_P P \geq G'=\sum_P m_P P$, falls $n_P \geq m_P$ für alle P ist, so können wir für einen Divisor G definieren

$$L(G)=\{0 \neq f \in K(\mathcal{X}) \mid (f) \geq -G\} \cup \{0\}.$$

Der Divisor G sagt also etwas aus über die Ordnungen der Null- und Polstellen von $f \in L(G)$. Der Raum $L(G)$ ist ein endlichdimensionaler K-Vektorraum.

Weiterhin nehmen wir an, dass in G kein $P_i \in \mathcal{P}$ mit einem Koeffizienten ungleich Null vorkommt, d.h. die Punkte in $\mathcal{P}$ sind sicher keine Polstellen für $f \in L(G)$. Da die Koeffizienten der P_i alle in K liegen, gilt ferner $f(P_i) \in K$ für $f \in L(G)$. Wie im Fall der Reed-Solomon-Codes können wir nun Goppa-Codes definieren als

$$C(\mathcal{P}, G) = \{(f(P_1), \ldots, f(P_n)) \mid f \in L(G)\}.$$

Für einen Divisor $G = \sum_P n_P P$ sei schließlich der Grad festgelegt durch $\operatorname{grad}\ G = \sum_P n_P$. Ist nun

$$2g - 2 < \operatorname{grad}\ G < n,$$

wobei g das Geschlecht der Kurve ist, so erhalten wir unter Verwendung des Satzes von Riemann-Roch für die Dimension k und die Minimaldistanz d des Goppa-Codes $C(\mathcal{P}, G)$ die Aussagen

Satz 6. $k = \operatorname{grad}\ G - g + 1$ *und* $d \geq n - \operatorname{grad}\ G$.

Beachten wir noch die Singleton-Schranke $d \leq n - k + 1 = n - \operatorname{grad}\ G + g$, so besagt Satz 6, dass die Minimaldistanz eines Goppa-Codes höchstens um das Geschlecht g vom maximalen Wert entfernt bleibt. Geschlecht 0 Kurven liefern also MDS-Codes, Geschlecht 1 Kurven fast MDS-Codes usw.

Mit dieser Konstruktion konnte nicht nur die asymptotische Gilbert-Varshamov-Schranke verbessert werden, sondern sie ist ganz allgemein eine wesentliche Hilfe beim Auffinden optimaler Codes, d.h. von $[n, k]_q$-Codes mit maximaler Minimaldistanz d.

Wir wollen die Goppa-Konstruktion an einem Beispiel verdeutlichen. Sei dazu $K = \mathbb{F}_5$ ein Körper mit 5 Elementen und $\mathcal{X}$ die projektive elliptische Kurve, die durch $f = y^2 z - 4x^3 - 2xz^2 \in K[x, y, z]$ definiert ist. Das Geschlecht g der Kurve ist gleich 1. Für die Menge $\mathcal{P}$ der K-rationalen Punkte auf $\mathcal{X}$ berechnet man

$$\mathcal{P} = \{(0:0:1), (1:\pm 1:1), (2:\pm 1:1), (3:\pm 2:1), (4:\pm 2:1), (0:1:0)\}.$$

Ist h das über K irreduzible Polynom $1 + x - x^2 + x^4 + x^5$, so gilt $K[x]/(h) = \mathbb{F}_{5^5}$. Man bestätigt, dass

$$(2x^6)^2 = -x^3 + 2x - (1 + x)(2 + x)(1 + x^2 + x^3 + x^4)h \equiv -x^3 + 2x \bmod h.$$

Somit hat $Q = (x : \pm 2x^6 : 1)$ Koordinaten in $\mathbb{F}_{5^5}$ und Q liegt auf der Kurve. Sei G der rationale Divisor der aus den 5 Frobenius-Konjugierten von Q besteht. Dann gilt $0 = 2g - 2 < \operatorname{grad}\ G = 5 < n = |\mathcal{P}| = 10$. Anwendung von Satz 6 liefert $\dim C(\mathcal{P}, G) = 5$ und $d \geq n - \operatorname{grad}\ G = 5$. Wegen der Singleton-Schranke ist $d \leq n - k + 1 = 6$. Die sogenannte Griesmer-Schranke, auf die wir hier nicht eingehen, besagt, dass für einen MDS-Code stets $d \leq q$ gilt, sofern $k \geq 2$ ist. Wir erhalten somit $d = 5$. Also ist der Goppa-Code $C(\mathcal{P}, G)$ ein $[10, 5, 5]_5$-Code, und man findet durch Berechnung von $L(G)$ eine *Erzeugermatrix* (d.h. die Zeilen bilden eine Basis von $C(\mathcal{P}, G)$) der Gestalt

$$E = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 3 \\ 0 & 0 & 1 & 0 & 0 & 1 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 1 & 0 & 1 & 4 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & 1 & 4 & 4 & 1 & 0 \end{pmatrix}.$$

Versucht man ohne eine schlagkräftige mathematische Theorie einen $[10, 5, 5]$-Code über einem Körper mit 5 Elementen zu konstruieren, so kann man die ersten 5 Spalten von E als gegeben annehmen. Die letzten 5 sind dann so zu finden, dass je 4 Spalten von E stets linear unabhügig sind. Man versuche dies einmal.

Im Jahr 2005 konnte Stichtenoth [35] unter Verwendung einer geschickt gewählten Folge algebraischer Kurven zeigen, dass die asymptotische Abschätzung von Tsfasman-Vlăduţ-Zink-Schranke (Satz 4) sogar für die Klasse der transitiven Codes gilt. Dabei heißt ein linearer Code C *transitiv*, falls die Gruppe der Permutationen π auf $\{1, \ldots, n\}$ mit

$$(a_1, \ldots, a_n) \mapsto (a_{\pi(1)}, \ldots, a_{\pi(n)}),$$

die C invariant lassen, transitiv operiert, d.h., zu beliebig vorgebenen Indizes $i, j \in \{1, \ldots, n\}$ existiert eine Permutation π mit $\pi(i) = j$.

Wie bereits im vorigen Abschnitt erwähnt, kann man eine analoge Aussage für zyklische Codes bis heute nicht beweisen, obwohl der zyklische Shift transitiv operiert. Ändert man in der Stichtenoth-Konstruktion die Kurven so ab, dass die Galoisgruppen der zugehörenden Funktionenkörper nicht nur transitiv sondern zyklisch operieren (d.h. die assozierten Goppa-Codes sind zyklisch), so strebt die Folge der relativen Distanzen gegen 0. Die zentrale Frage, ob zyklische Codes asymptotisch gut sind, bleibt somit weiterhin offen. Neben diesem schwierigen Problem ist weiterhin die tiefliegende Frage, ob auch binäre Codes die Gilbert-Varshamov-Schranke übertreffen, offen.

Den zu Codes und algebraische Kurven interessierten Leser verweisen wir auf das Buch [34] von Stichtenoth, dass eine Fülle von interessanten Resultaten enthält.

V Obere Schranken und Lineare Optimierung

Die besten allgemeinen oberen Schranken für $A_q(n, d)$ liefern Methoden der Linearen Optimierung. Die Ideen gehen auf die wegweisende Arbeit von Delsarte [8] aus dem Jahr 1973 zurück und haben in der letzten Zeit durch Verfeinerungen von Schrijver [31] zu neuen und überraschenden Abschätzungen geführt.

Wir beschränken uns im folgenden auf $K = \mathbb{F}_2$. Sei C ein nicht notwendig linearer Code in K^n. Für $k = 0, \ldots, n$ definieren wir die Matrix M_k vom Typ

$(2^n, 2^n)$ durch die Festsetzung

$$(M_k)_{a,b} = \begin{cases} 1, & \text{falls} \quad \mathrm{d}(a,b) = k, \\ 0, & \text{sonst,} \end{cases}$$

wobei $a, b \in K^n$ ist. Ferner sei Γ die Gruppe aller Abbildungen, die die Koordinaten der Vektoren $a = (a_1, \dots, a_n) \in K^n$ permutiert, gefolgt von einem möglichen Switch $0 \leftrightarrow 1$ an den n Positionen. Man beachte, dass

$$\mathrm{d}(\gamma a, \gamma b) = \mathrm{d}(a, b)$$

für alle $a, b \in K^n$ und alle $\gamma \in \Gamma$ gilt. (Γ ist die Isometriegruppe des binären Hammingraums, wenn wir die Vektorraumstruktur vergessen.) Ferner operiert Γ transitiv auf der Menge der Paare (a, b) mit festem Abstand k. Ist χ^C die charakteristische Funktion von C auf K^n, so definieren wir

$$X^C = (\chi^C(a))_{a \in K^n},$$

wobei $\chi^C(a)$ als Spaltenvektor zu lesen ist. X^C ist ebenfalls eine 0-1 Matrix vom Typ $(2^n, 2^n)$. Schließlich setzen wir

$$(*) \qquad M^C = \frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} X^{\gamma C} (X^{\gamma C})^T.$$

Offensichtlich ist M^C eine positiv semidefinite Matrix und man bestätigt, dass

$$M^C = \sum_{k=0}^{n} x_k M_k,$$

wobei

$$x_k = \frac{|\{(c, c') \in C \times C,\ \mathrm{d}(c, c') = k\}|}{|\{(a, a') \in K^n \times K^n,\ \mathrm{d}(a, a') = k\}|}.$$

Die Matrix M^C liegt also in der sogenannten *Bose-Mesner* Algebra

$$\mathcal{BM} = \{\sum_{k=0}^{n} x_k M_k \mid x_k \in \mathbb{C}\}.$$

Die Menge $\mathcal{BM}$ ist in der Tat eine $\mathbb{C}$-Algebra, sogar eine kommutative Algebra. Da die Matrizen M_k symmetrisch sind und paarweise vertauschen, sind sie simultan diagonalisierbar. Die Eigenwerte und deren Vielfachheiten lassen sich wie folgt bestimmen. Für $k = 0, \dots, n$ definieren wir die *Krawtchouk-Polynome* $K_k(x)$ durch

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

Diese stehen in enger Beziehung zu den im Abschnitt III definierten Lloyd-Polynomen, denn es gilt

$$L_e(x) = \sum_{k=0}^{e} K_k(x).$$

Eine direkte Rechnung liefert, dass M_k für $l = 0, \ldots, n$ den Eigenwert $K_k(l)$ mit der Vielfachheit $\binom{n}{l}$ hat. Hierbei numeriert l die gemeinsamen Eigenräume der Matrizen M_k. Da $M^C = \sum_{k=0}^{n} x_k M_k$ positiv semidefinit ist, erhalten wir somit

$$\sum_{k=0}^{n} x_k K_k(l) \geq 0 \quad \text{für} \quad l = 0, \ldots, n.$$

Dies beinhaltet die folgende obere Schranke für die Mächtigkeit eines binären Codes der Länge n und der Minimaldistanz mindestens d.

Satz 7. (Delsarte-Schranke)
Es gilt

$$A_2(n,d) \leq \max \{\sum_{k=0}^{n} \binom{n}{k} x_k\}$$

unter den Nebenbedingungen

(i) $x_0 = 1$ *und* $x_1 = \ldots = x_{d-1} = 0.$

(ii) $x_d, \ldots, x_n \geq 0.$

(iii) $\sum_{k=0}^{n} x_k K_k(l) \geq 0$ *für* $l = 0, \ldots, n.$

Die Behauptung des Satzes ergibt sich, indem man x_k durch $\frac{2^n}{|C|} x_k$ ersetzt. Dann sind die Nebenbedingungen weiterhin erfüllt und es gilt

$$\frac{2^n}{|C|} \sum_{k=0}^{n} \binom{n}{k} x_k = |C|.$$

Schrijver hat nun 2005 in [31] gezeigt, dass eine Verfeinerung der Bose-Mesner Algebra die von Delsarte gefundene Schranke verbessert. Diese Verfeinerung wird durch die sogenannte *Terwilliger Algebra* $\mathcal{T}$ geliefert, die wie folgt definiert ist:

$$\mathcal{T} = \{ \sum_{k,i,j=0}^{n} x_{k,i,j} M_{k,i,j} \mid x_{k,i,j} \in \mathbb{C}\},$$

wobei die 0-1 Matrizen $M_{k,i,j}$ vom Typ $(2^n, 2^n)$ gegeben sind durch

$$(M_{k,i,j})_{a,b} = \begin{cases} 1, & \text{falls} \quad \mathrm{d}(a,b) = k,\ \mathrm{wt}(a) = i,\ \mathrm{wt}(b) = j, \\ 0, & \text{sonst.} \end{cases}$$

Nun gilt, wenn wir in $(*)$ für die Gruppe Γ die Untergruppe nehmen, die nur aus den Permutationen besteht, dass

$$M^C = \sum_{k,i,j=0}^{n} x_{k,i,j} M_{k,i,j},$$

wobei

$$x_{k,i,j} = \frac{2^n |\{(c,c',c'') \in C^3 \mid \mathrm{d}(c,c') = k,\ \mathrm{d}(c,c'') = i,\ \mathrm{d}(c',c'') = j\}|}{|C||\{(a,a',a'') \in K^{n^3} \mid \mathrm{d}(a,a') = k,\ \mathrm{d}(a,a'') = i,\ \mathrm{d}(a',a'') = j\}|}.$$

Man bestätigt wieder, dass $\mathcal{T}$ eine Algebra ist, jedoch im Gegensatz zur Bose-Mesner Algebra nicht mehr kommutativ. Somit können die symmetrischen Matrizen $M_{k,i,j}$ nicht mehr simultan diagonalisiert werden, jedoch noch *blockweise*. Mittels einer expliziten Beschreibung der Blöcke konnte Schrijver die Berechnung einer oberen Schranke für $A_2(n,d)$ auf ein Problem der Semidefiniten Optimierung zurückführen, für welches polynomiale Algorithmen bekannt sind. Benutzt man diese auf dem Rechner, so erhält man in vielen Fällen überraschende Resultate.

Für $n = 26$ und $d = 10$ liefert Delsarte's Methode als Schranke 1040. Während die beste bekannte Schranke vor Schrijver's Arbeit bei 989 lag, erhält man mit seiner Methode nun 886. Konstruieren kann man bislang jedoch nur einen Code mit 384 Worten.

Beide in diesem Abschnitt beschriebenen Methoden liefern auch obere Abschätzungen für die Anzahl der Worte in Codes, in denen alle Codeworte konstantes Gewicht w haben. Für $n = 23$, $d = 8$ und $w = 11$ findet man mit dem Verfahren von Delsarte 1417 als Schranke; andere Verfahren liefern 1328. Erstaunlich ist nun, dass Schrijver's Methode die bestmögliche Schranke 1288 findet, denn es gibt in der Tat einen $(23, 1288, 8)_2$-Code von konstantem Gewicht $w = 11$, nämlich die Menge der Codeworte vom Gewicht 11 im perfekten $[23, 12, 7]$-Golay-Code.

Asymptotisch liefert die Methode von Delsarte die sogenannte McEliece-Rodemich-Rumsey-Welch-Schranke [11], die zur Zeit beste bekannte obere asymptotische Schranke. Etwas abgeschwächt besagt sie

$$\alpha_2(\delta) \leq H_2((1 - \sqrt{(1-(1-2\delta)^2)}/2) \quad \text{für } 0 < \delta < 1/2.$$

Was hingegen Schrijver's Methode asympotisch liefert, scheint bislang offen zu sein.

VI Dualität, Gitter und Invariantentheorie

Dualitätsargumente sind in vielen Bereichen der Mathematik hilfreiche Werkzeuge, so auch in der Codierungstheorie. Analog zum Skalarprodukt in reellen Hilberträumen definieren wir hier eine nicht-ausgeartete symmetrische Bilinearform auf K^n durch die Festsetzung

$$< a, b > = \sum_{i=1}^{n} a_i b_i.$$

Man beachte, dass im Gegensatz zu Hilberträumen $< a, a > = 0$ werden kann, obwohl $a \neq 0$ ist; etwa $< (1,1), (1,1) > = 2 = 0$, falls die Charakteristik von K

gerade ist, d.h. $1+1=0$ in K gilt. Ist $C \subseteq K^n$, so setzen wir

$$C^\perp = \{a \mid a \in K^n, < a, c >= 0 \text{ für alle } c \in C\}$$

und nennen $C^\perp$ den zu C *orthogonalen Code*. Offenbar ist $C^\perp$ stets ein linearer Unterraum von K^n, auch wenn C keiner ist. Ist $C = C^\perp$, so heißt C *selbstdual*. Die Klasse der selbstdualen Codes ist von besonderem Interesse, da sie einerseits viele bestbekannte Codes enthält, andererseits tiefe Verbindungen zu anderen Gebieten der Mathematik wie Kombinatorik, Invariantentheorie, Gruppentheorie und Gitter aufweist.

Wir wollen in diesem Abschnitt stets annehmen, dass C ein linearer Code ist. Neben den Parametern n, k, d spielt dann insbesondere bei der Berechnung von Decodierfehlerwahrscheinlichkeiten [12], die eine Aussage über die Güte von Codes in den Anwendungen machen, die *Gewichtsverteilung*

$$A_i = |\{c \mid c \in C, \operatorname{wt}(c) = i\}| \quad (i = 0, \ldots, n)$$

eine zentrale Rolle. Das zugehörige erzeugende Polynom

$$A_C(x, y) = \sum_{i=0}^{n} A_i x^{n-i} y^i$$

nennt man *homogenes Gewichtspolynom*. Erstaunlich ist nun die von MacWilliams entdeckte Tatsache, dass sich die Gewichtsverteilung von $C^\perp$ vollständig aus der von C bestimmen läßt; genauer, es gilt:

Satz 8. (MacWilliams-Dualität) $A_{C^\perp}(x, y) = A_C(x + (q-1)y, x - y)$.

Was hat man damit gewonnen? Etwa folgendes: Die Berechnung der recht komplexen Gewichtsverteilung des $[n, n-k, 3]_q$-Hamming-Codes erfordert einiges an Talent. Der zu ihm duale Simplex-Code hat jedoch nur ein einziges Gewicht ungleich 0, nämlich q^{k-1}. Mittels Satz 8 können wir dann die Gewichtsverteilung des Hamming-Codes leicht angeben. Im kleinsten Fall $K = \mathbb{F}_2$ erhält man so

$$A_i = \begin{cases} \frac{1}{n+1}[\binom{n}{i} + n(-1)^{\frac{i}{2}}\binom{\frac{n-1}{2}}{\frac{i}{2}}] & \text{für } i \text{ gerade,} \\ \frac{1}{n+1}[\binom{n}{i} + n(-1)^{\frac{i+1}{2}}\binom{\frac{n-1}{2}}{\frac{i+1}{2}}] & \text{für } i \text{ ungerade.} \end{cases}$$

Dualitätsargumente sind häufig besonders schlagkräftig in gerader Charakteristik. Dementsprechend wollen wir im folgenden immer einen Körper gerader Charakteristik, besser noch $K = \mathbb{F}_2$ annehmen. In diesem Fall haben selbstduale Codes nur gerade Gewichte, denn für $c \in C = C^\perp$ gilt

$$0 =< c, c >= \sum_{c_i \neq 0} 1 \equiv \operatorname{wt}(c) \bmod 2.$$

Es kann sogar vorkommen, dass alle Gewichte in einem selbstdualen Code durch 4 teilbar sind. Solche Codes heißen *doppelt-gerade*. Erweitert man die Codeworte des binären $[7,4,3]$-Hamming-Codes um ein Kontrollbit (d.h. eine 1, falls das Gewicht ungerade ist, 0 sonst), so erhält man einen selbstdualen doppelt-geraden $[8,4,4]$-Code. Das gleiche Verfahren macht aus dem perfekten $[23,12,7]$-Golay-Code einen selbstdualen doppelt-geraden $[24,12,8]$-Code.

Binäre Codes haben nun die schöne Eigenschaft, dass man sie aus interessanten Gittern gewinnen kann. Dazu betrachten wir die modulo 2 Reduktion

$$\rho : \mathbb{Z}^n \to (\mathbb{Z}/2\mathbb{Z})^n = K^n.$$

Ist C ein linearer Code in K^n, so ist

$$\Gamma_C = \frac{1}{\sqrt{2}}\rho^{-1}(C)$$

ein Gitter in $\mathbb{R}^n$.

Eigenschaften des Codes C spiegeln sich nun in Eigenschaften des Gitters Γ_C wider. Dazu benötigen wir die folgenden Definitionen.

Ein Gitter $\Gamma \subseteq \mathbb{R}^n$ nennen wir *ganz*, falls das euklidische Skalarprodukt $x \cdot y = \sum_{i=1}^n x_i y_i$ für $x, y \in \Gamma$ nur Werte in $\mathbb{Z}$ annimmt; es heißt *gerade*, falls $x \cdot x \in 2\mathbb{Z}$ für alle $x \in \Gamma$ gilt und *unimodular*, falls

$$\Gamma = \{x \in \mathbb{R}^n \mid x \cdot y \in \mathbb{Z} \text{ für alle } y \in \Gamma\}.$$

Es gelten nun die folgenden Zusammenhänge:

C ist genau dann selbstorthogonal, d.h. $C \subseteq C^\perp$, wenn Γ_C ganz ist.

C ist genau dann selbstdual, wenn Γ_C unimodular und ganz ist.

C ist genau dann doppelt-gerade, wenn Γ_C gerade ist.

Dies legt nahe, dass man binäre selbstduale Codes über die zugehörigen Gitter studieren kann. Wie die Theorie der algebraischen Kurven über die Goppa-Codes zu tiefen Ergebnissen in der Codierungstheorie führt, so liefert die reichhaltige Theorie der Gitter, insbesondere die zugehörigen Thetareihen und modularen Formen, interessante codierungstheoretische Resultate. Dazu eine kleine Kostprobe.

Da ganze gerade unimodulare Gitter im $\mathbb{R}^n$ höchstens für $n \equiv 0 \bmod 8$ existieren, erhalten wir als direkte Anwendung das folgende Resultat.

Satz 9. (Gleason) *Ein binärer selbstdualer doppelt-gerader Code hat eine durch 8 teilbare Länge.*

Auf die vielen Zusammenhänge zwischen Gittern und Codes können wir hier nicht eingehen. Wir empfehlen dazu das interessante Buch [10] von Ebeling.

Als selbstduale doppelt-gerade Codes haben wir bereits den $[8, 4, 4]$-Hamming- und den $[24, 12, 8]$-Golay-Code kennengelernt. Diese Codes sind *extremal*, d.h., die Minimaldistanz d nimmt den größtmöglichen Wert an. Dies folgt aus einem Satz von Mallows und Sloane [26], dessen Beweis auf Methoden der Invariantentheorie beruht und folgendes besagt.

Satz 10. (Mallows-Sloane)
a) Ist C ein binärer selbstdualer doppelt-gerader $[n, k, d]$-Code, so gilt

$$d \leq 4 \left\lfloor \frac{n}{24} \right\rfloor + 4.$$

b) Für großes n gilt in a) nie die Gleichheit, d.h., für großes n gibt es keine extremalen doppelt-geraden Codes.

Mallows und Sloane haben in b) keine explizite Schranke für n angegeben. Das Beste, was man heute weiß, ist die Schranke von Zhang [40] aus dem Jahr 1999, nach welcher es für $n > 3928$ keine extremalen Codes gibt. Enttäuschend im Hinblick auf diese Schranke ist nun die Tatsache, dass man nur für ganz kleine n extremale Codes kennt. Verlangt man noch, dass n durch 24 teilbar ist - diese Codes sind aus geometrischer Sicht von besonderem Interesse, da die Codeworte von einem festen Gewicht ungleich 0 nach Assmus und Mattson [1] ein 5-Design bilden -, so kennt man nur zwei Beispiele, den $[24, 12, 8]$-Golay-Code und einen weiteren $[48, 24, 12]$-Code, der zu den sogenannten quadratischen Restklassen Codes gehört.

Bereits 1973 hat Sloane die Frage nach der Existenz eines binären selbstdualen doppelt-geraden $[72, 36, 16]$-Codes gestellt. Die Antwort ist immer noch offen, trotz enormer Anstrengungen vieler Mathematiker. Wir kommen hierauf im nächsten Abschnitt noch einmal zurück.

Was aber hat Satz 10 mit Invariantentheorie zu tun? Die entscheidenden Ideen, die wir hier kurz für den binären Körper $K = \mathbb{F}_2$ skizzieren, gehen auf die berühmte Arbeit [14] von Gleason zurück. Für einen binären selbstdualen Code C liefert der Dualitätssatz von MacWilliams, dass

$$A_C(x, y) = A_C(x + y, x - y).$$

Das Polynom $A_C(x, y)$ ist somit invariant unter der Involution

$$S = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Ist C zudem doppelt-gerade, also $A_i = 0$ für $4 \nmid i$, so ist $A_C(x, y)$ auch invariant unter

$$R = \begin{pmatrix} 1 & 0 \\ 1 & i \end{pmatrix}$$

mit $i = \sqrt{-1} \in \mathbb{C}$. Somit ist das Gewichtspolynom invariant unter der Gruppe $G = \langle R, S \rangle$, die von unitären Spiegelungen erzeugt wird. Mit etwas elementarer Gruppentheorie zeigt man, dass G genau 192 Elemente hat, von denen 30 unitäre

Spiegelungen sind. Aus Sätzen von Shephard-Todd und Chevalley aus den fünfziger Jahren des letzten Jahrhunderts über Invarianten von Spiegelungsgruppen weiß man, dass die Invariantenalgebra zu G in $\mathbb{C}[x, y]$ eine Polynomalgebra in zwei algebraisch unabhängigen homogenen Polynomen vom Grad d_1 und d_2 ist. Dabei gilt $|G| = d_1 d_2$ und $d_1 + d_2 - 2$ ist die Anzahl der unitären Spiegelungen in G. Somit sind die d_i die beiden Lösungen 8 und 24 der Gleichung

$$t^2 - 32t + 192 = 0.$$

Invarianten mit diesen Graden kennen wir bereits, nämlich die homogenen Gewichtspolynome zum $[8, 4, 4]$-Hamming- und $[24, 12, 8]$-Golay-Code. Diese sind

$$p = x^8 + 14x^4y^4 + y^4 \text{ und } q = x^{24} + 759x^{16}y^8 + 2576x^{12}y^{12} + 759x^8y^{16} + y^{24}.$$

Somit ist das homogene Gewichtspolyom eines binären selbstdualen doppelt-geraden Codes stets ein Polynom in p und q. Insbesondere zeigt dies wieder, dass die Länge solcher Codes durch 8 teilbar ist (dies ist der ursprüngliche Beweis von Gleason), welches wir bereits aus der Gittertheorie gefolgert hatten. Statt der Invarianten q kann man auch das Polynom

$$r = \frac{p^3 - q}{48} = x^4y^4(x^4 - y^4)^4$$

nehmen, mit dem sich etwas einfacher rechnen läßt. Das Gewichtspolynom

$$A_C(x, y) = \sum_{j=0}^{n/4} A_{4j} x^{n-4j} y^{4j}$$

eines binären selbstdualen doppelt-geraden Codes der Länge n liegt also in der homogenen Komponente vom Grad n der Polynomalgebra

$$\mathbb{C}[p, r] = \mathbb{C}[x^8 + 14x^4y^4 + y^4, x^4y^4(x^4 - y^4)^4].$$

Setzen wir noch $D = \lfloor \frac{n}{24} \rfloor$, so folgt

$$A_C(1, y) = \sum_{j=0}^{D} a_j p(1, y)^{n-24j} r(1, y)^j = \sum_{j=0}^{n/4} A_{4j} y^{4j}.$$

Wir wählen nun $a_0, a_1, \ldots, a_D \in \mathbb{Z}$, so dass

$$A_0 = 1, A_1 = \ldots = A_{4r} = 0$$

ist mit r so groß wie möglich, egal ob dazu ein Code existiert oder nicht. Das zugehörige Polynom $A^*(y)$ mit den Koeffizienten A_i^* nennen wir *extremales Gewichtspolynom*. Falls ein Code existiert mit diesem Gewichtspolynom, so hat es die größtmögliche Minimaldistanz unter den selbstdualen doppelt-geraden Codes. Der Koeffizient a_0 ist wegen $A_0 = 1$ als $a_0 = 1$ festgelegt. Die anderen Koeffizienten $a_1, \ldots, a_D$ wählen wir so, dass die kleinstmögliche Potenz von y, die in $A^*(y)$ mit einem Koeffizienten ungleich 0 auftreten kann, gleich $4(D+1)$ ist. Mallows und Sloane haben nun gezeigt, dass der Koeffizient bei $y^{4(D+1)}$ in der Tat ungleich 0 ist (dies liefert Teil a) des Satzes 10) und der Koeffizient bei y^{4D+8} für große n negativ wird (dies ist Teil b) des Satzes 10).

Wie bereits erwähnt liefert die Invariantentheorie $n = 3928$ als obere Schranke für die Länge eines binären extremalen selbstdualen doppelt-geraden Codes. Auch der kürzlich von Duursma [9] gefundene invariantenfreie Zugang zu derartigen Schranken liefert das gleiche n. Beachtet man nun, dass $n = 136$ die größte Länge eines bisher gefundenen Codes ist, so bleibt eine nicht unerhebliche Lücke zu schließen.

Die Invariantentheorie liefert deutlich mehr für die Codierungstheorie, als wir hier gezeigt haben. Das kürzlich erschienene Buch [29] von Nebe, Rains und Sloane beinhaltet eine sehr allgemeine Theorie invariantentheoretischer Methoden, die viele Resultate früherer Autoren, die im Wesentlichen auf Gleason's Ideen aufbauen, vereinheitlicht.

VII Darstellungstheorie und Automorphismen von Codes

Die Automorphismengruppe $\mathrm{Aut}(C)$ eines linearen Codes C im K^n ist definiert als die Gruppe der monomialen Matrizen $M \in \mathrm{GL}(n, K)$, die C invariant lassen, für die also $CM = C$ gilt. Dabei heißt M *monomial*, falls M in jeder Zeile und jeder Spalte genau einen Eintrag ungleich 0 aus K hat. Für $|K| = 2$ sind die monomialen Matrizen also gerade die Permutationsmatrizen.

Wählt man zufällig einen binären $[n, k]$-Code, so ist die Automorphismengruppe für großes n mit hoher Wahrscheinlichkeit trivial [21], sofern k relativ zu n nicht zu klein oder zu groß ist (dies sind die interessanten Fälle in den Anwendungen). Viele der uns bekannten Codes haben jedoch eine nicht-triviale Automorphismengruppe. So enthält die Automorphismengruppe eines zyklischen Codes C der Länge n einen n-Zykel. Die Hamming-Codes der Dimension $n-k$ haben die volle lineare Gruppe $\mathrm{GL}(k, K)$ als Automorphismengruppen und der $[24, 12, 8]$-Golay-Code die einfache Mathieu-Gruppe M_{24}.

Was hat man gewonnen, wenn ein linearer Code C eine nicht-triviale Automorphismengruppe G hat? Offensichtlich ist dann C nicht nur ein Vektorraum sondern ein Modul für die Gruppenalgebra KG, d.h. man kann darstellungstheoretische Methoden zur Lösung codierungstheoretischer Fragen einsetzen. Die Schlagkraft der Darstellungstheorie wollen wir an folgendem Resultat von Sloane und Thompson [33] verdeutlichen.

Satz 11. (Sloane-Thompson) *Sei $K = \mathbb{F}_2$ und G eine endliche Gruppe mit 2-Sylowgruppe T. Ist T zyklisch, so ist ein selbstduales Ideal in KG niemals doppelt-gerade.*
Insbesondere gibt es keine binären Codes, die zyklisch selbstdual und doppelt-gerade sind.

Der ursprüngliche von Sloane und Thompson gegebene Beweis ist zwar elementar, gibt jedoch wenig Einblick in den wirklichen Sachverhalt. Darstellungstheoretisch ist die Aussage einsichtig und der Beweis überraschend kurz. Da $T = \langle x \rangle$ zyklisch ist, hat T nach einem berühmten Satz von Burnside ein normales Komplement H in G. Das Idempotent $e = \sum_{h \in H} h$ liefert dann die direkten Zerlegungen

$$KG = eKG \perp (1-e)KG \quad \text{und} \quad C = eC \perp (1-e)C.$$

Dabei ist eC ein selbstdualer Untermodul von $eKG = eKT$. Der einzige selbstduale Untermodul in der zyklischen Gruppenalgebra KT wird von $x^{|T|/2} - 1$ erzeugt, hat also das Minimalgewicht 2. Somit hat $eC \subseteq C$ das Minimalgewicht

$$2\mathrm{wt}(e) = 2|H| \equiv 2 \bmod 4,$$

welches den Satz beweist.

In der Codierungstheorie betrachtet man häufig binäre erweiterte zyklische Codes, etwa quadratische Restklassen Codes der Länge $p+1$, wobei p eine Primzahl ist. Dies sind spezielle zyklische Codes erweitert um ein Kontrollbit. Der $[24, 12, 8]$-Golay-Code gehört zum Beispiel zu dieser Klasse, aber auch der extremale $[48, 24, 12]$-Code, der uns bereits im letzten Abschnitt begegnet ist. Derartige Codes C kann man als Untermoduln von $KG \oplus K$ auffassen, wobei die Gruppe G trivial auf der Komponente K, die das Kontrollbit angibt, operiert. Die codierungstheoretische Dualität $C \mapsto C^\perp$ läßt sich nun gegen die darstellungstheoretische Dualität

$$C \mapsto C^* = Hom_k(C, K)$$

ausspielen, denn es gilt der KG-Isomorphismus

$$(*) \quad (KG \oplus K)/C^\perp \cong C^*.$$

Ist C ferner selbstdual, also $C = C^\perp$, so hat jeder einfache KG-Modul $V \cong V^*$ wegen $(*)$ eine gerade Vielfachheit in $KG \oplus K$. Wegen $C = C^\perp$ muß $|G|$ gerade sein. Ist K nun von gerader Charakteristik, so ist die Gruppenalgebra KG nach einem berühmten Satz von Maschke ([18], Kap. I, Satz 17.6) halbeinfach und alle einfachen KG-Moduln haben ungerade Vielfachheit in KG. Da die Vielfachheit für einfache Moduln $V \cong V^*$ in $KG \oplus K$ jedoch gerade sein muß, wie wir oben festgestellt haben, ist der triviale Modul der einzige einfache KG-Modul mit $V \cong V^*$. Nutzt man weitere tiefergehende darstellungstheoretische Argumente, so erhält man mit der Festsetzung, dass, für $n \in \mathbb{N}$ ungerade, $s(n)$ die kleinste natürliche Zahl ist mit $n \mid 2^{s(n)} - 1$, die folgende Äquivalenz [28].

Satz 12. *Ist G eine endliche Gruppe mit $|G|$ ungerade und $K = \mathbb{F}_2$, so sind gleichwertig:*

a) Es gibt einen G-invarianten selbstdualen Code $C \leq KG \oplus K$.

b) Der triviale Modul ist der einzige einfache KG-Modul V mit $V \cong V^*$.

c) Alle einfachen KG-Moduln haben eine ungerade Dimension.

d) $s(p)$ ist ungerade für alle Primzahlen p mit $p \mid |G|$.

Die in der Regel schwer zu beantwortende Frage in a), ob ein erweiterter selbstdualer Gruppencode existiert, wird über die Darstellungstheorie in b) und c) zu einer zahlentheoretischen Bedingung in d) reduziert, die leicht nachprüfbar ist. Der Beweis liefert zudem ein Konstruktionsverfahren für den Code C. Dies kann man nun ausnutzen um zu zeigen, dass binäre selbstduale erweiterte Gruppencodes, deren Länge durch 8 teilbar ist, doppelt-gerade sind, d.h. es gilt die Umkehrung des Satzes 9 von Gleason für diese Klasse von Codes.

Kommen wir nochmals auf die Frage von Sloane nach der Existenz eines binären selbstdualen doppelt-geraden $[72, 36, 16]$-Codes zurück. Gäbe es einen solchen Code C und wäre seine Automorphismengruppe G nicht-trivial, so könnten wir G zur Konstruktion von C gewinnbringend nutzen. Hier stellt sich also die Frage, was G, insbesondere die Ordnung $|G|$ von G sein kann. Das Beste, was man heute dazu nach Vorarbeiten vieler Autoren weiß, ist das folgende Resultat [4], indem auch noch Gruppen der Ordung 56 und 72 vorkommen, die aber inzwischen ebenfalls ausgeschlossen wurden.

Satz 13. *Ist C ein binärer selbstdualer doppelt-gerader* $[72, 36, 16]$*-Code, so ist*

$$|\mathrm{Aut}(C)| = 5, 7, 10, 14 \text{ oder ein echter Teiler von } 72.$$

Insbesondere ist die Automorphismengruppe auflösbar.

Die beiden bekannten Beispiele mit einer durch 24 teilbaren Länge, also der $[24, 12, 8]$-Golay-Code und der $[48, 24, 12]$ quadratische Restklassen Code haben hingegen einfache Gruppen als Automorphismengruppen, nämlich M_{24} und $\mathrm{PSL}(2, 47)$. Neben elementarer Gruppentheorie verlangt der Beweis von Satz 14 tiefere darstellungstheoretische Methoden. Auch wird die *Meataxe* (Software zum Auffinden von Untermoduln) verwendet, etwa um die nach theoretischen Überlegungen übrig gebliebenen einfachen Gruppen $\mathrm{A}_5, \mathrm{A}_6$ und $\mathrm{SL}(2, 8)$ auszuschließen.

Der Spezialist wird nun einwenden: Diese Vorgehensweise kennen wir doch aus der Zeit, als man versuchte, eine projektive Ebene der Ordnung 10 zu finden. Nach Anstrengungen mehrerer Autoren konnte man schließlich zeigen, dass die Kollineationsgruppe einer solchen Ebene, falls sie existiert, trivial ist [19]. Ähnliches kann natürlich mit der Automorphismengruppe eines hypothetischen selbstdualen doppelt-geraden $[72, 36, 16]$-Codes passieren.

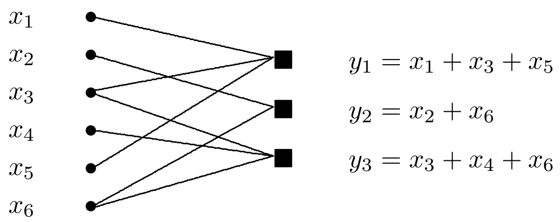
VIII LDPC-Codes und Graphentheorie

LDPC-Codes (*low density parity check codes*) haben in den letzten Jahren ein enormes Interesse in der Forschung gefunden. Dies sind binäre lineare Codes, deren Kontroll-Matrizen wenig Einsen besitzen, also dünn besetzt sind. Sie wurden bereits 1963 von Gallager [13] definiert, haben allerdings in den ersten Jahren nach ihrer Entdeckung wenig Beachtung gefunden.

Jeden binären $[n, n-k]$-Code C kann man bekanntlich über eine Kontroll-Matrix $H = (h_{ij})$ auch mittels eines bipartiten Graphen $\Gamma(C)$ (häufig auch *Tanner-Graph* genannt) beschreiben. Dieser hat die (linke) Eckenmenge $V = \{x_1, \ldots, x_n\}$ (*die Variablen*), die (rechte) Eckenmenge $W = \{y_1, \ldots, y_k\}$ (*die Nebenbedingungen*) und Kanten $(x_i, y_j) \in V \times W$, falls $h_{ji} = 1$ ist. Die Kontroll-Matrix

$$H = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$

bestimmt also den Tanner-Graphen



und umgekehrt. Genau dann liegt $(x_1, \ldots, x_n)$ im Code C, wenn alle Nebenbedingungen Null sind.

Wir nennen einen Code (l, r)-*regulär*, falls jede Ecke in V den Grad l und jede Ecke in W den Grad r hat. Für einen regulären $[n, n-k]$-Code gilt also $nl = rk$. Die zugehörige Kontroll-Matrix hat also genau r Einsen in jeder Zeile und l Einsen in jeder Spalte.

Betrachten wir nun eine Folge (C_n) von (l, r)-regulären Codes C_n der Länge n mit r und l fest, jedoch wachsendem n, so werden die Kontroll-Matrizen offenbar dünn besetzt. Bereits Gallager zeigte, dass, sofern man eine solche Folge zufällig wählt, diese asymptotisch gut ist. Mehr noch und dies ist aus der Sicht der Anwendungen entscheidend: Für diese Codes gibt es Codier- und Decodieralgorithmen mit einer Laufzeit, die linear in der Länge sind. Im allgemeinen hat die Implementation eines linearen Codes der Länge n die Komplexität $O(n^2)$ für die Codierung und $O(n^3)$ für die Decodierung.

Was fehlte waren explizite Konstruktionsverfahren für LDPC-Codes. Den ersten entscheidenden Schritt in dieser Richtung lieferte Margulis [27]. Die Konstruktion beruht auf guten Expander Graphen.

Ein (l, r)-regulärer bipartiter Graph mit Eckenmengen V, W und $|V| = n$ heißt ein $(n, l, r, \epsilon, \delta)$-*Expander* mit $\epsilon, \delta > 0$, falls für jede linke Eckenmenge $L \subseteq V$ mit $|L| \leq \epsilon|V|$ gilt

$$|\partial L| > \delta|L|,$$

wobei ∂L die Nachbarschaft von L ist, also die Menge aller Ecken (in W), die mit mindestens einer Ecke in L verbunden sind. δ nennt man auch den *Ausdehnungsfaktor* des Expanders. Jedem derartigen Expander E und einem beliebigen binären linearen Code B der Länge r können wir einen binären linearen *Expander-Code* $C(E, B)$ der Länge $n = |V|$ auf folgende Weise zuordnen: Ist

$j \in W$ eine rechte Ecke, so ist diese wegen der Regularität des Graphen mit genau r linken Ecken, etwa $i_{1,j}, \ldots, i_{r,j} \in V$, verbunden. Nun ist

$$(c_1, \ldots, c_n) \in C(E, B)$$

genau, falls

$$(c_{i_{1,j}}, \ldots, c_{i_{r,j}}) \in B$$

für alle rechten Ecken $j \in W$. Nach einem Resultat [30] von Sipser und Spielman aus dem Jahr 1996 gilt:

Satz 14. (Sipser-Spielman) *Sei E ein $(n, l, r, \alpha, \frac{l}{r\epsilon})$-Expander und sei B ein binärer linearer Code der Länge r, der Rate $R > \frac{l-1}{l}$ und der relativen Minimaldistanz ϵ. Dann hat der Expander-Code $C(E, B)$ mindestens die Rate $lR-(l-1)$ und mindestens die relative Minimaldistanz α.*

Der Beweis der beiden Aussagen ist erstaunlich einfach. Die angegebene Rate ergibt sich wie folgt: Jede Nebenbedingung erfordert $r - \dim\, B = r(1-R)$ Gleichungen. Somit wird $C(E, B)$ durch $\frac{nl}{r}(r(1-R))$ Gleichungen beschrieben, die aber nicht unbedingt unabhängig sein müssen. Es folgt

$$\dim\, C(E, B) \geq n - \frac{nl}{r}(r(1-R)) = n(lR - (l-1)),$$

woraus die Abschätzung für die Rate folgt.

Sei nun $v \in C(E, B)$ mit $\mathrm{wt}(v) \leq \alpha n$. Sei L die Menge der Variablen, die in v mit dem Wert 1 vorkommen. In diesen Variablen enden $l|L|$ Kanten. Der Ausdehnungsfaktor besagt, dass diese Kanten in mehr als $\frac{l}{r\epsilon}|L|$ Nebenbedingungen enden. Somit existiert mindestens eine Nebenbedingung, in der weniger als $r\epsilon$ Kanten enden. Dies besagt aber, dass B eine Rate echt kleiner als ϵ hat, ein Widerspruch zur Voraussetzung im Satz.

Wählen wir für E einen Expander mit dem Ausdehnungsfaktor $r/2$ und für B den $[r, r-1, 2]$-Code B_0, der aus allen Worten von geradem Gewicht besteht, so hat der Expander-Code $C(E, B_0)$ wegen $R = (r-1)/r$ nach dem obigen Satz mindestens die Rate $lR - (l-1) = 1 - \frac{l}{r}$ und die relative Minimaldistanz α. Hat der Expander nicht nur den Ausdehnungsfaktor $r/2$ sondern sogar $3r/4$, so kann man mittels eines einfachen iterativen Verfahrens sogar $\alpha n/2$ Fehler in linearer Zeit korrigieren (siehe [30]), wobei n die Länge von $C(E, B_0)$ ist. Das Decodierverfahren ist in der Tat recht einfach. Ist $v = (x_1, \ldots, x_n)$ ein empfangenes Wort mit höchstens $\alpha n/2$ Fehlern, so verfährt man wie folgt:

1. Der (binäre) Wert einer Variablen, die in den Nebenbedingungen mit dem Wert 1 öfter vorkommt als in solchen mit dem Wert Null, wird geflippt.

2. Der Schritt 1 wird solange wiederholt, bis alle Nebenbedingungen den Wert Null haben.

Dass das Verfahren terminiert, liegt an dem relativ großen Ausdehnungsfaktor.

Wie aber findet man für $n \to \infty$ eine Folge von $(n, l, r, \epsilon, \delta)$-Expander mit großem δ? Um solche zu konstruieren, startet man mit r-regulären nicht notwendig bipartiten Graphen, bei denen ebenfalls kleine Eckenmengen große Nachbarschaften haben. Sei dazu Γ ein r-regulärer Graph mit n Ecken. Dann ist der zugehörige Kanten-Ecken-Graph ein bipartiter $(2, r)$-Graph mit den Kanten als Ecken auf der linken Seite und den Ecken auf der rechten, also $|V| = rn/2$ und $|W| = n$. Dabei ist eine Kante mit einer Ecke verbunden, falls diese auf der Kante liegt.

Sei nun Γ ein zusammenhängender r-regulärer Graph. Da die zugehörige Adjazenz-Matrix symmetrisch ist, sind ihre Eigenwerte alle reell, und $\lambda_0 = r$ ist bekanntlich der größte Eigenwert und hat die Vielfachheit 1. Sei $\lambda = \lambda(\Gamma)$ der zweitgrößte Eigenwert. Man kann nun zeigen, dass der zu Γ gehörende Kanten-Ecken-Graph ein $(rn/2, 2, r, \alpha, \delta(\alpha))$-Expander mit

$$\delta(\alpha) = \frac{4}{2\alpha r + (r + \lambda)(1 - \alpha)}$$

ist. Setzen wir in Satz 14 nun $\delta(\alpha) = 2/(r\epsilon)$, so hat der zugehörige Expander-Code $C(E, B)$ mindestens die relative Minimaldistanz

$$\frac{2r\epsilon - r - \lambda}{r - \lambda}.$$

Um viele Fehler korrigieren zu können, sollte dieser Wert möglichst groß sein. Offenbar ist er um so größer, je kleiner λ ist. Der zweitgrößte Eigenwert von Γ sollte also möglichst großen Abstand vom größten Eigenwert haben. Er kann jedoch für Graphen mit einer großen Anzahl von Ecken nicht beliebig klein werden, denn es gilt das folgende Resultat, dessen Beweis man in [24] findet.

Satz 15. (Boppona-Alon) *Sei $r \in \mathbb{N}$ fest und sei (Γ_n) eine Folge von zusammenhängenden r-regulären Graphen, wobei n die Anzahl der Ecken in Γ_n ist. Dann ist*

$$\liminf_{n\to\infty} \lambda(\Gamma_n) \geq 2\sqrt{r-1}.$$

Die in Satz 15 angegebene untere Schranke wird asymptotisch von den sogenannten *Ramanuyan-Graphen* erreicht. Dies sind endliche zusammenhängende r-reguläre Graphen Γ mit $\lambda(\Gamma) \leq 2\sqrt{r-1}$, deren explizite Konstruktion auf Margulis [27] und Lubotzky, Phillips und Sarnak [24] zurückgeht. Genauer:

Satz 16. (Lubotzki-Phillips-Sarnak, Margulis) *Seien p, q zwei Primzahlen, wobei $q \equiv 1 \bmod 4$ und p ein quadratischer Rest modulo q ist. Dann gibt es einen $(p+1)$-regulären Cayley-Graphen Γ in $\mathrm{PSL}(2, \mathbb{Z}/q\mathbb{Z})$ mit $q(q^2-1)/2$ Ecken und zweitgrößtem Eigenwert $\lambda(\Gamma) \leq 2\sqrt{p}$.*

Unendliche Folgen von r-regulären Ramanuyan-Graphen sind bis heute nur für $r = q + 1$ bekannt, wobei q eine Primzahlpotenz ist. Der erste offene Fall ist $r = 7$. Die Ramanujan-Graphen treten auf, wenn man einen großen Ausdehnungsfaktor über den zweitgrößten Eigenwert der Adjazenz-Matrix abschätzen will. Das asymptotische Verhalten der Ausdehnungsfaktoren kann aber wesentlich verschieden sein vom Verhalten der zweitgrößten Eigenwerte [6]. Auch basiert die Abschätzung in Satz 14 von Sipser und Spielman auf der Unabhängigkeit der Nebenbedingungen. Kann man etwa durch geeignete Wahl von B Abhängigkeiten erzeugen, die eine größere relative Minimaldistanz des Expander-Codes zur Folge hat? Dies sind nur einige wenige der vielen offenen Fragen in diesem Bereich der Codierungstheorie.

Abschließend sei vermerkt, dass Expander auch außerhalb der Graphentheorie und der hier gezeigten codierungstheoretischen Anwendung von großer Bedeutung sind. Insbesondere in der Informatik spielen sie beim Design optimaler Netzwerke eine zentrale Rolle.

Eine schöne Darstellung der Ramanujan-Graphen findet man in [7]. Für weitergehende Resultate zwischen Expander und Codes verweisen wir auf [36].

IX Codes und die projektive Ebene der Ordnung 10

Wir haben bisher gesehen, dass vielfältige Methoden aus sehr unterschiedlichen Gebieten der Mathematik in der Codierungstheorie zu interessanten Resultaten führen. Dass dies keine Einbahnstraßen sind, sondern die Codierungstheorie auch der Mathematik zurück gibt, zeigt beispielsweise der 1989 zu Ende gebrachte Beweis, dass keine projektive Ebene der Ordnung $n = 10$ existiert. Es war Ed Assmus, der 1970 in einem Vortrag [2] in Oberwolfach die entscheidenden Ideen aus einer gemeinsamen Arbeit mit Mattson lieferte. Er bewies dort sehr elementar und codierungstheoretisch den folgenden Satz, der einen Spezialfall des bekannten Bruck-Ryser Theorems [5] darstellt.

Satz 17. *Es gibt keine projektive Ebene der Ordnung* $n \equiv 6 \bmod 8$.

Assmus' Beweis verläuft wie folgt. Sei A die Inzidenzmatrix einer projektiven Ebene der Ordnung $n \equiv 2 \bmod 4$, also eine $0,1$-Matrix mit $n^2 + n + 1$ Zeilen und Spalten, wobei jede Zeile das Gewicht $n + 1$ hat und je zwei verschiedene Zeilen an genau einer Stelle eine 1 gemeinsam haben. Da A über $\mathbb{F}_2$ vom Rang $(n^2+n+2)/2$ ist, wie man zeigen kann, erzeugen die Zeilen von A einen binären $[n^2+n+1, (n^2+n+2)/2]$-Code. Fügen wir zu den Codeworten ein Kontrollbit hinzu, so erhalten wir einen binären selbstdualen Code der Länge $n^2 + n + 2$. Wegen $n \equiv 2 \bmod 4$ ist der erweiterte Code dann doppelt-gerade und nach Gleason's Satz 10 gilt somit $8 \mid n^2 + n + 2$. Insbesondere ist $n \not\equiv 6 \bmod 8$ und der Satz ist bewiesen.

Mit diesen Argumenten allein konnte man natürlich den Fall $n = 10$ nicht ausschließen wegen $n^2 + n + 2 = 112 \equiv 0 \bmod 8$. Assmus und Mattson zeigten in [3] jedoch, dass das Gewichtspolynom eindeutig durch die drei Koeffizienten A_{12}, A_{15} und A_{16} festgelegt ist. Mittels Computer konnte man dann zeigen, dass diese Koeffizienten alle gleich Null sind, was zu $A_{19} = 24675$ führte. Langwierige Computerrechnungen haben schließlich erbracht, das keine Konfiguration der 24675 Codeworte vom Gewicht 19 zu einem binären Code gehört, der über die Inzidenzmatrix einer projektiven Ebene der Ordung 10 definiert ist. Der interessierte Leser sei für Details auf den Übersichtsartikel [20] verwiesen.

X Abschließende Bemerkungen

Die Codierungstheorie hat mittlerweile zu Recht ihren eigenständigen Platz in der Mathematik gefunden. Sie ist nicht mehr nur eine Domäne der Elektroingenieure und Jessi MacWilliams müsste heute sicher ihre Aussage von 1968 revidieren. Die Mathematiker haben die Herausforderung, von der Golay sprach, angenommen. Was die Codierungstheorie mathematisch so interessant macht - die Studierenden wissen dies auch zu schätzen -, sind die engen Verbindungen zu vielen anderen Gebieten der Mathematik. Gute Ergebnisse in der Codierungstheorie haben häufig gute Resultate aus der Kombinatorik, der endlichen Geometrie, der Graphentheorie, der Gittertheorie, der Optimierung, der Theorie algebraischer Kurven oder der Darstellungstheorie zum Ausgangspunkt. Wir haben in dieser Note versucht, dies, wenn auch nur an kleinen Beispielen, zu verdeutlichen. Die Codierungstheorie ist ein sehr lebendiges Gebiet. Viele tiefliegende Fragen sind noch offen; man denke nur an die im Text angesprochenen zentralen Probleme.

Literatur

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

[2] E.F. Assmus Jr.. The projective plane of order ten? Oberwolfach Vortragsbuch, April 1970.

[3] E.F. Assmus Jr. and H.F. Mattson Jr.. On the possibility of a projective plane of order 10. AF Cambridge Research Labs., AFCRL-71-0013 (1970).

[4] S. Bouyuklieva, E. O'Brien and W. Willems. The automorphism group of a self-dual doubly-even $[72, 36, 12]$ code is solvable. To appear *IEEE Trans. Inform. Theory* (2006).

[5] R.H. Bruck and H.J. Ryser. The nonexistence of certain finite projective planes. *Can. J. Math.* **1**, 88-93 (1949).

[6] R. Brooks and A. Zuk. On the asymptotic isoperimetic constants for Riemann surfaces and graphs. *J. Differential Geometry* **62**, 49-78 (2002).

[7] G. Davidoff, P. Sarnak and A. Valette. Elementary Number Theory, Group Theory, and Ramanujan Graphs. London Math. Soc., Student Texts 55, Cambridge 2003.

[8] P. Delsarte. An algebraic approach to the association schemes. *Philips Res. Repts. Suppl.* **10**, (1973).

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

[10] W. Ebeling. Lattices and Codes. Vieweg, second edition 2002.

[11] R.J. McEliece, E.R. Rodemich, H. Rumsey, Jr., and L. Welch. New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities. *IEEE Trans. Inform. Theory* **23**, 157-166 (1977).

[12] A. Faldum, J. Lafuente, G. Ochoa and W. Willems. Error probabilities for bounded distance decoding. To appear *Designs, Codes and Cryptography* (2006).

[13] R.G. Gallager. Low-density parity-check codes. Cambridge, MA: MIT Press, 1963.

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

[15] M.J.E. Golay. Notes on digital coding. *Proc. IRE* **37**, 657 (1949).

[16] V.D. Goppa. Codes associated with divisors. *Probl. Peredachi Inf.* **13**, 33-39 (1977). Englische Übersetzung in *Probl. Inf. Transm.* **13**, 22-27 (1977).

[17] 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).

[18] B. Huppert. Endliche Gruppen I. Springer Verlag 1967.

[19] Z. Janko and Tran van Trung. Projective planes of order 10 do not have a collineation of order 3. *J. Reine Angew. Math.* **325**, 189-209 (1981).

[20] C.W.H. Lam. The search for the projective plane of order 10. *Am. Math. Monthly* **98**, 305-318 (1991).

[21] H. Lefmann, K.T. Phelps and V. Rödl. Rigid linear binary codes. *J. Comb. Theory A* **63**, 110-128 (1993).

[22] J.H. van Lint. The mathematics of the Compact Disc. In *DMV Mitteilungen* Heft **4**, 25-29 (1998).

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

[24] A. Lubotzki, R. Phillips and P. Sarnak. Ramanujan graphs. *Combinatorica* **8**, 261-277 (1988).

[25] F.J. MacWilliams. A historical survey. In: *Error correcting codes,* Ed. H.B. Mann, John Wiley, 3-13 (1968).

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

[27] G.A. Margulis. Explicit constructions of concentrators. *Probl. Inform. Transm.* **9**, 325-332 (1975).

[28] C. Martínez-Pérez and W. Willems. Self-dual codes and modules for finite groups in characteristic two. *IEEE Trans. Inform. Theory* **50**, 1798-1803 (2005).

[29] G. Nebe, E.M. Rains and N.J.A. Sloane. Self-dual codes and invariant theory. Springer Verlag, 2006.

[30] M. Sipser and D.A. Spielman. Expander codes. *IEEE Trans. Inform. Theory* **42**, 1710-1722 (1996).

[31] A. Schrijver. New code upper bounds from the Terwilliger algebra and semidefinite programming. *IEEE Trans. Inform. Theory* **51**, 2859-2866 (2005).

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

[33] N.J.A. Sloane and J.G. Thompson. Cyclic self-dual codes. *IEEE Trans. Inform. Theory* **29**, 364-366 (1983).

[34] H. Stichtenoth. Algebraic function fields and codes. Springer Verlag 1993.

[35] H. Stichtenoth. Transitive and self-dual codes attaining the Tsfasman-Vladut-Zink bound. *IEEE Trans. Inform. Theory* **52**, 2218-2224 (2006).

[36] K.W. Shum and I.F. Blake. Expander graphs and codes. Dimacs Series in Discrete Mathematics and Theoretical Computer Science, Vol. **68**, 57-68 (2005).

[37] A. Tietäväinen. On the nonexistence of perfect codes over finite fields. *SIAM J. Appl. Math.* **24**, 88-96 (1973).

[38] 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).

[39] J.L.Vasil'ev. On nongroup close-packed codes. *Probl. Kibernet* **8**, 337-339 (1962).

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

[41] V.A. Zinov'ev and V.K. Leont'ev. The nonexistence of perfect codes over Galois fields. *Problems of Control and Inform. Theory* **2**, 123-132 (1973).