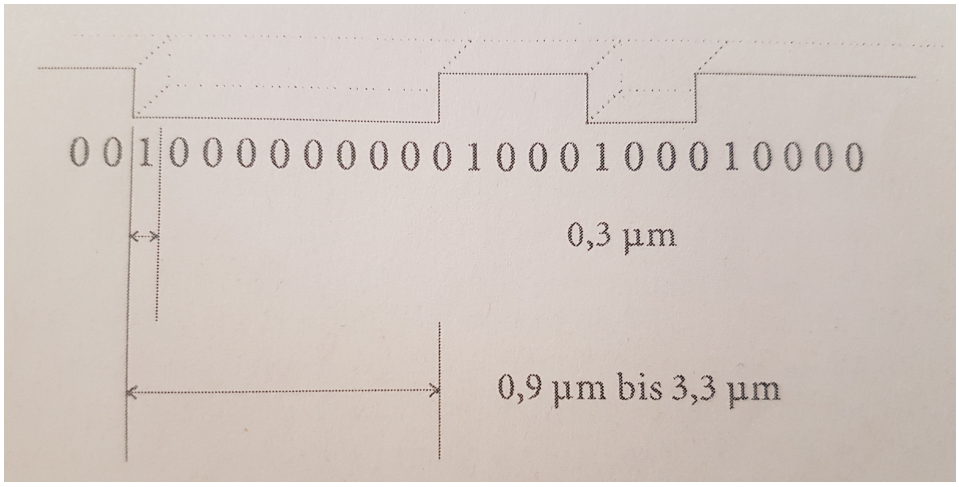
Vorlesung vom 16.06.2020

1 Die Compact Disc (CD)

Auf einer CD ist die digitale Information auf einer spiralförmigen nach innen laufenden Spur in Form von Vertiefungen (Pits) und Nicht-Vertiefungen (Lands) gespeichert.



Im Unterschied zu vielen anderen Anwendungen treten Fehler auf einer CD meist nicht vereinzelt auf, sondern, verursacht durch Kratzer, Staub oder auch Fingerabdrücke, in gehäufter Form, also in sogenannten **Bündeln**, d.h. vielfach sind mehrere hundert Bits in Folge zerstört.

Definition. *Ein Vektor* $v \in K^n$ *heißt* Bündel der Länge b *falls*

$$v = (0, \ldots, 0, v_i, \ldots, v_{i+b-1}, 0, \ldots, 0)$$

mit $v_i \neq 0 \neq v_{i+b-1}$.

Mit anderen Worten ist ein Fehlerbündel der Länge b ein Fehlervektor, deren von 0 verschiedene Einträge sich in b aufeinanderfolgenden Koordinaten befinden. Zur Korrektur derartiger Fehler eigent sich das sogenannte **Interleaving**, welches Bündelfehler verteilt, sodass sie beim Decodierer als zufällige Einzelfehler auftreten.

Beispiel. *Die gesendete Nachricht lautet* ***WARMER APFELKUCHEN*** *und die emfangene und verfälschte Nachricht ist* ***WARMER*** $\star\star\star\star\star$***KUCHEN***. *Letzteres könnte beispielsweise* ***WARMER PFANNKUCHEN*** *bedeuten. Werden nun die Buchstaben um* 5 *versetzt geschrieben, lautet das gesendete Wort* ***WPHEKA FERURENACML***. *Geht der mittlere Teil wieder verloren, so lautet die empfangene Nachricht* ***WPHEKA*** $\star\star\star\star\star$***ENACML***. *Dies kann man nun als* ***WA*** $\star$ ***ME*** $\star$ ***AP*** $\star$ ***ELK*** $\star$ ***CH*** $\star$ ***N*** *lesen. D.h. der Informationsverlust tritt nicht mehr konzentriert auf, sondern weitgestreut, sodass die Nachricht einfacher rekonstruierbar ist.*

1.1 Interleaving

Sei C ein $[n, k, d]$-Code und $t \in \mathbb{N}$. Dann nennt man den $[tn, tk, d]$-Code

$$C(t) := \{(c_{11}, \ldots, c_{t1}, \ldots, c_{1n}, \ldots, c_{tn}) \mid (c_{i1}, \ldots, c_{in}) \in C \text{ für } i = 1, \ldots, t\}$$

Interleaving von C zur Tiefe t.
Die Codewörter aus $C(t)$ sind also die spaltenweise gelesenen Einträge der Matrizen

$$\begin{pmatrix} c_{11} & \dots & c_{1n} \\ \vdots & & \vdots \\ c_{t1} & \dots & c_{tn} \end{pmatrix} \tag{1}$$

mit Zeilen aus C.

Beispiel. *Sei* $C = \mathbb{F}_2^4$ *und* $t = 3$. *Es gilt*

$$(0,1,1,0), (1,0,1,0), (0,1,1,1) \in C$$

und somit

$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix} \hat{=} (0,1,0,1,0,1,1,1,1,0,0,1) \in C(3).$$

Lemma. *Der Code* $C(t)$ *hat die Parameter* $[tn, tk, d]$, *wobei* C *ein* $[n,k,d]$*-Code ist.*

Beweis. Die Blocklänge ist offensichtlich tn. Sei $\{c_1, \dots, c_k\}$ eine Basis von C, so ist

$$\left\{ \begin{pmatrix} & c_i & \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{pmatrix}, \begin{pmatrix} 0 & \dots & 0 \\ & c_i & \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \\ & c_i & \end{pmatrix} \mid 1 \le i \le k \right\}$$

eine Basis von $C(t)$. Sei

$$c = \begin{pmatrix} c_1 \\ \vdots \\ c_t \end{pmatrix} = \begin{pmatrix} c_{11} & \dots & c_{1n} \\ \vdots & & \vdots \\ c_{t1} & \dots & c_{tn} \end{pmatrix} \in C(t)$$

mit $c_i \in C$ für $1 \le i \le t$, so gilt

$$\mathrm{wt}(c) = \sum_{i=1}^{t} \mathrm{wt}(c_i).$$

Dementsprechend ist die Minimaldistanz $d(C(t)) = \min_{c \in C(t) \setminus \{0\}} \{\mathrm{wt}(c)\} = d(C)$. □

Man beachte, dass durch das Interleaving nicht die Korrektur zufälliger Fehler verbessert wird, da C und $C(t)$ die gleiche Minimaldistanz haben. Allerdings ist bemerkenswert, dass einer Verbesserung bei den korrigierbaren Fehlerbündel auftritt. Wenn C Fehlerbündel bis zur Länge b korrigieren kann, so sind mit $C(t)$ Fehlerbündel bis zu einer Länge tb korrigierbar, denn diese beeinflussen höchstens b aufeinanderfolgende Einträge in jeder Zeile der Matrix. Anschaulich bedeutet dies: Sei für $m \le bt$ ein Fehlerbündel für $C(t)$ aufgetreten, so sind in den Zeilen der Matrix (1) maximal b Bündelfehler aufgetreten. Diese können aber zeilenweise korrigiert werden, da C Fehlerbündel der Länge b erkennt.

1.2 Verzögerten Interleaving

In der Praxis werden in der Regel beim Interleaving Codewörter in einer Matrix angeordnet, wobei sie die Zeilen bilden. Die Matrix wird dann spaltenweise verschickt. Dadurch muss, wegen des Aufbaus der Matrizen, eine zeitliche Verzögerung in Kauf genommen werden, da sie erst verschickt wird, wenn sie gefüllt ist. Einen derartigen Nachteil gibt es beim sogenannten (einfach-) Verzögerten Interleaving nicht. Die Codewöter werden wie folgt notieren

$$\begin{array}{ccccccccc} c_{i1} & c_{i+1,1} & c_{i+2,1} & \dots & & & & & \\ & c_{i2} & c_{i+1,3} & c_{i+2,2} & \dots & & & & \\ & & c_{i3} & c_{i+1,3} & c_{i+1,4} & \dots & & & \\ & & & \ddots & \ddots & & & & \\ & & & c_{in} & c_{i+1,n} & c_{i+2,n} & \dots, & & \end{array}$$

wobei der Rest mit 0 aufgefüllt wird. Der Vorteil ist, dass in jeder Spalte ein Codewort endet, wodurch der Datenfluss gleichmäßiger wird.
Die ersten zwei Zeilen des zweifach verzögerten Interleaving sind

$$\begin{array}{cccccc} c_{i1} & c_{i+1,1} & c_{i+2,1} & \dots & & \\ & & c_{i2} & c_{i+1,3} & c_{i+2,2} & \dots \end{array}$$

Der Vorteil von mehrfach verzögerten Interleaving ist, dass die Fehler im Bündel auf mehrere Codewörter verteilt werden. Bei der CD wird 4-fach verzögertes Interleaving benutzt.

1.3 Cross-Interleaving

Um Fehler mit einer sehr großen Wahrscheinlichkeit entdecken und verbessern zu können, wird das Cross-Interleaving eingesetzt. Hierbei werden zwei Codes C_1 und C_2 verschachtelt. Die Codewörter von C_1 werden als Zeilen einer Matrix geschrieben und dessen Spalte mit C_2 codiert.

Beispiel. *Sei C_1 der erweiterte binäre Hammingcode $[8,4,4]$ mit Erzeugermatrix*

$$G_1 = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Schreibe drei Codewörter von C_1 als Zeilen einer Matrix M

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

Wähle nun einen zweiten Code C_2 vom Typ $[7,3,4]$ mit Erzeugermatrix

$$G_2 = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

Die Spalten von M werden mit C_2 wie folgt codiert:

$$\begin{aligned} (1,0,0) &\longmapsto (1,0,1,1,1,0,0) \\ (1,1,0) &\longmapsto (1,1,1,0,0,1,0) \\ (1,0,1) &\longmapsto (1,0,0,1,0,1,1) \\ &\vdots \end{aligned}$$

Aus M erhalten wir dann die folgenden acht Codewörter, die wir als die Zeilen der folgenden Matrix M' schreiben:

$$M' = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

1.4 Der zugrundeliegende Körper

Auf der CD wird mit Codes über $K = \mathbb{F}_{2^8}$ codiert, wobei K der eindeutige Körper mit 2^8 Elementen ist. Er kann wie folgt konstruiert werden. Sei $f \in \mathbb{F}_2[X]$ ein irreduzibles Polynom vom Grad 8, so gilt

$$K \cong \mathbb{F}_2[X]/(f) \cong \mathbb{F}_2[\alpha] = \{g(\alpha) \mid g \in \mathbb{F}_2[X]\},$$

wobei α eine Nullstelle des Polynoms f in einem Erweiterungskörper von $\mathbb{F}_2$ ist. Z.B. ist das Polynom $f = x^8 + x^4 + x^3 + x + 1$ irreduzibel über $\mathbb{F}_2$ und es gilt

$$x^8 = x^4 + x^3 + x + 1 \mod (f),$$

d.h. x^8 entspricht in K dem Polynom x^4+x^3+x+1. Das Polynom x^9 entspricht somit nach folgender Rechnung $x^5 + x^4 + x^2 + x$ in K

$$x^9 = x \cdot x^8 = x \cdot (x^4 + x^3 + x + 1) = x^5 + x^4 + x^2 + x.$$

Also kann man jedes Element $g \in K$ durch genau einem Polynom $g = \sum_{i=0}^{7} a_i X^i$ mit $a_i \in \mathbb{F}_2$ darstellen und somit mit dem 8-Tupel $(a_0, \ldots, a_7)$ identifizieren.
Auf der CD werden die Elemente aus $K = \mathbb{F}_{2^8}$ als 8-Tupel aus $\mathbb{F}_2^8$ geschrieben.

1.5 Die CD

Wie bereits beschrieben wird die digitale Information spiralförmig nach innen laufenden Spur in Form von Vertiefungen (Pits) und Nicht-Vertiefungen (Lands) gespeichert. Dabei entspricht ein Bit der ungefähren Länge $0,3\mu m = 3 \cdot 10^{-7} m$. Eine CD hat eine Spurlänge von $\approx 5km$, sodass damit ungefähr 17 Milliarden Bits abgespeichert werden können. Die korrekte Lesbarkeit der Folge mittels des Lasers erfordert, dass zwischen zwei Einsen mindestens $r = 2$ Nullen stehen müssen, die Synchronisation verlangt, dass höchstens $s = 10$ Nullen stehen dürfen.
Mit einer Frequenz von 44,1kHz wird die Amplitude des Musiksignals abgetastet und mittels eines Analog-Digital-Wandlers über die binäre Darstellung in einen 16 Bit Vektor umgewandelt.

Beispiel. *Ist zur Zeit t_0 die Amplitude*

$$53261 = 2^0 + 2^2 + 2^3 + 2^{12} + 2^{14} + 2^{15},$$

so ist der zugehörige 16 *Bit Vektor*

$$(1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1).$$

Die hohe Abtastfrequenz folgt aus dem **Satz von Nyquist-Shannon**. Dieser besagt, dass ein analoges Signal mit maximaler Frequenz f mindestens $2f$-mal abgetastet werden muss, damit man aus dem zeitdiskreten Signal das Ursprungssignal ohne Informationsverlust rekonstruieren kann. Die menschliche Hörfrequenz von etwa 22kHz verlangt also eine Abtastrate von mindestens 44kHz.

1.6 Das Audiowort

Einmaliges Abtasten liefert zwei 16-Tupel, also ein 32-Tupel. Die 8-Tupel werden als Elemente in $\mathbb{F}_{2^8}$ aufgefasst, sodass ein 32-Tupel einem 4-Tupel über $\mathbb{F}_{2^8}$ entspricht. Je 6 Abtastungen werden zu einem Audiowort zusammengefasst. Dies ist also ein $32 \cdot 6 = 192$-Wort über $\mathbb{F}_2$ oder $4 \cdot 6 = 24$-Wort über $\mathbb{F}_{2^8}$. D.h. ein Audiowort ist ein 24-Tupel über $\mathbb{F}_{2^8}$.

1.7 Cross Interleaved Reed-Solomon-Code (CIRC)

Es wird mit CIRC codiert. In der Übung werden die dazugehörigen Codes C_1 und C_2 konstruiert. Der Code C_1 hat die Parameter $[32, 28, 5]$ und C_2 die Parameter $[28, 24, 5]$ über dem Körper $\mathbb{F}_{2^8}$, die verkürzte Reed-Solomon Codes sind.
Jedes Audiowort, also ein Wort der Länge 24 über $\mathbb{F}_{2^8}$ wird mit C_2 zu einem Wort der Länge 28 codiert

$$C_2 : \mathbb{F}_{2^8}^{24} \longrightarrow \mathbb{F}_{2^8}^{28}, \ v \longmapsto v \cdot G,$$

wobei G die Erzeugermatrix von C_2 ist. Also

$$(a_1, \ldots, a_{24}) \longmapsto (a_1, \ldots, a_{24}) \cdot G = (v_1, \ldots, v_{28}).$$

Anschließend wird ein 4-fach verzögertes Interleaving der Tiefe 28 durchgeführt. Es werden 28 Codewörter $(v_{1,1}, \ldots, v_{1,28}), \ldots, (v_{28,1} \ldots, v_{28,28})$ gesammelt und dann 4-fach verzögert notiert

$$\begin{array}{ccccccccccc} v_{1,1} & v_{2,1} & v_{3,1} & v_{4,1} & v_{5,1} & \ldots & & & & & \\ & & & & v_{1,2} & v_{2,2} & v_{3,2} & v_{4,2} & v_{5,2} & \ldots \\ & & & & & \vdots & & & & & \end{array}$$

Nun werden die Spalten dieser Matrix, die Länge 28 haben, mit C_1 codiert. Also erhalten wir für jede Spalte ein Wort der Länge 32. An jedes so erhaltende Codewort wird nun ein Element aus $\mathbb{F}_{2^8}$ angehängt, welches Display-Informationen erhält. Insgesamt führen also 6 Abtastungen zu einem Vektor der Länge 33 über $\mathbb{F}_{2^8}$.
Ein solcher Vektor kann jedoch nicht unmittelbar auf eine CD übertragen werden, da die Bedingungen $r = 2$ und $s = 10$ zum korrekten Lesen durch den Laser nicht erfüllt sind. Um diese Problematik zu umgehen, benutzt man die Eight-to-Fourteen Methode. Es werden 8 aufeinanderfolgende Bits in 14 umgewandelt, die $r = 2$ und $s = 10$ erfüllen. Außerdem werden zwischen diesen nochmals 3 Bits geeignet eingefügt, sodass die Übergänge zwischen der 14 Bits ebenfalls die Bedingungen erfüllt. Das Einfügen von zwei Bits hätte gereicht, aber durch die drei zusätzlichen Bits ist die Gesamtlänge von Pits und die Gesamtlänge von Lands von Beginn der Folge bis zum zweiten 14 Bit Vektor etwa gleich.
Bei 6 Abtastungen erhalten wir $33 \cdot (14 + 3) = 561$ Bits. An diese werden noch 24 Bits für die Synchronisation und weitere 3 Pufferbits angehängt. Also haben wir bei 6 Abtastungen $561 + 24 + 3 = 588$ Bits. Diese werden Kanalbits genannt. Es werden pro Sekunde also $\frac{44100 \cdot 588}{6} = 4321800$ Kanalbits erzeugt.

1.8 Die Decodierung

Lemma. *Ein $[n, k, d]$-Code über einem beliebigen Körper K kann bis zu $d - 1$ Auslöschungen korrigieren.*

Beweis. Sei $c = (c_1, \ldots, c_n) \in C$ und seien o.B.d.A. die ersten $a \leq (d-1)$-Stellen ausgelöscht worden, etwa $c = (\star, \ldots, \star, c_{a+1}, \ldots, c_n)$. Ist $c' \in C$ mit $c' = (\ldots, c_{a+1}, \ldots, c_n)$, so gilt offenbar $d(c, c') \leq a \leq d - 1$, also $c = c'$. □

Beim Abspielen der CD werden zuerst Puffer und Synchronisationsbits entfernt und das 14-Tupel wieder in ein 8-Tupel umgewandelt. Dann wird der C_1-Decodierer auf die Wörter der Länge 32 angesetzt. Der Code C_1 ist 2-fehlerkorrigierend. Aber er korrigiert auf der CD nur einen Fehler, aber erkennt dafür mit hoher Wahrscheinlichkeit mehr als drei Fehler und mit Sicherheit 2 oder 3 Fehler.
Wenn der C_1-Decodierer feststellt, dass ein 32-Tupel Abstand 1 zu einem Codewort hat, dann wird der Fehler korrigiert. Ansonsten werden alle 28 Symbole, die in das Codewort übergegangen waren, als Auslöschung, d.h. die entsprechenden Stellen können nicht korrekt gelesen werden, markiert. Dann wird das Interleaving zur Tiefe 4 wieder rückgängig gemacht und die so erhaltenen 28-Tupel werden von dem C_2-Decodierer bearbeitet. Dieser kümmert sich nur um Auslöschungen (bis zu 4 pro C_2-Codewort).
Die 4-fach Verzögerung bewirkt, dass bis zu 16 aufeinanderfolgende Audioworte wieder hergestellt werden können, was einer Spurlänge von $\approx 2,5mm$ auf der CD entspricht.

2 Duale Codes

Sei K ein Körper. Wir versehen den K-Vektorraum K^n mit der folgenden Paarung

$$\langle -,- \rangle : K^n \times K^n \longrightarrow K,$$

wobei $\langle (v_1,\dots,v_n),(w_1,\dots,w_n)\rangle = \sum_{i=1}^n v_i w_i$ für $(v_1,\dots,v_n),(w_1,\dots,w_n) \in K^n$. Die Paarung ist nicht-ausgeartet, d.h. für $u \in K^n$ folgt aus $\langle u, v\rangle = 0$ für alle $v \in K^n$ gerade $u = 0$, und symmetrisch.

Bemerkung. *Für $K = \mathbb{F}_2$ und $n = 2$ gilt $\langle (1,1),(1,1)\rangle = 1 + 1 = 0$.*

Definition. *Sei K ein beliebiger Körper, $n \in \mathbb{N}$ und $C \subseteq K^n$.*

(a) $C^\perp := \{v \in K^n \mid \langle c, v\rangle = 0$ *für alle* $c \in C\}$ *heißt der zu C* duale Code.

(b) Wenn $C = C^\perp$ gilt, so nennt man C selbstdual

Bemerkung. *Die Menge $C^\perp$ ist stets ein Vektorraum und somit ein linearer Code. Ist $\dim_K(\langle C\rangle_K) = k$, so gilt $\dim_K(C^\perp) = n - k$ und $(C^\perp)^\perp = \langle C\rangle_K$.*

Theorem 2.1. *Sei C ein $[n,k]$-Code über dem Körper K.*

(a) Eine Matrix H ist genau dann eine Kontrollmatrix für C, wenn H eine Erzeugermatrix für $C^\perp$ ist.

(b) Ist $(E_k \mid A)$ eine Erzeugermatrix für C (bis auf Äquivalenz hat jeder Code eine Erzeugermatrix dieser Form), so ist $\left(-A^t \mid E_{n-k}\right)$ eine Erzeugermatrix für $C^\perp$.

Beweis. (a) Sei $H \in K^{(n-k)\times n}$ eine Kontrollmatrix für C vom Rang $n-k$. Da die Zeilen von H in $C^\perp$ liegen und

$$\dim_K(C^\perp) = n - k = \mathrm{rk}(H)$$

sind die Zeilen von H eine Basis von $C^\perp$, d.h. H ist eine Erzeugermatrix für $C^\perp$.

Ist $H \in K^{(n-k)\times n}$ eine Erzeugermatrix für $C^\perp$, so gilt $H \cdot C = 0$, d.h. $C \subseteq \ker(H)$. Wegen $\mathrm{rk}(H) = n - k$ ist

$$\dim_K(\ker(H)) = n - (n-k) = k = \dim_K(C)$$

und deshalb $\ker(H) = C$.

(b) Ist $(E_k \mid A)$ eine Erzeugermatrix für $C = (C^\perp)^\perp$, so ist nach (a) $(E_k \mid A)$ eine Kontrollmatrix von $C^\perp$. Nun folgt aus dem Lemma von der Lösung des 5. Übungsblatts, dass $\left(-A^t \mid E_{n-k}\right)$ eine Erzeugermatrix von $C^\perp$ ist.

□

Beispiel. *Der duale Code zu dem Hammingcode $Ham_q(k)$ ist ein* Simplex-Code *$Sim_q(k)$.*

Lemma. *(a) Ist $c \in Sim_q(k) \setminus \{0\}$, so gilt $wt(c) = q^{k-1}$. Es gibt also nur die Gewichte 0 und q^{k-1}.*

(b) $Sim_q(k)$ hat die Parameter $\left[n = \frac{q^k-1}{q-1}, k, q^{k-1}\right]$.

Beweis. (a) Sei H eine Kontrollmatrix für $\mathrm{Ham}_q(k)$, d.h. eine Erzeugermatrix für $\mathrm{Sim}_q(k)$. Seien $z_1,\dots,z_k$ die Zeilen von H und $c \in \mathrm{Sim}_q(k) \setminus \{0\}$ beliebig mit

$$0 \neq c = \sum_{i=1}^k a_i z_i.$$

Seien $z_i = (z_{i1},\dots,z_{in})$, so sind $h_j = \begin{pmatrix} z_{1j} \\ \vdots \\ z_{kj} \end{pmatrix}$ die Spalten von H und somit ein Repräsentantensystem der 1-dim Unterräume von K^k. Also

$$c = (c_1,\dots,c_n) = \sum_{i=1}^k a_i z_i = (\sum_i^k a_i z_{i1},\dots,\sum_i^k a_i z_{in}) = (a_1,\dots,a_k)\cdot(h_1,\dots,h_n).$$

Nun gilt genau dann $c_j = 0$, wenn $(a_1, \ldots, a_n)h_j = 0$. Der Vektorraum $U = (a_1, \ldots, a_k)^\perp$ in K^k ist $(k-1)$-dimensional und es gilt genau dann $c_j = 0$, wenn $h_j \in U$. Der Raum U besitzt $\frac{q^{k-1}-1}{q-1}$ 1-dimensionale Unterräume, also liegen genauso viele $h_j's$ in U. Somit gilt

$$\mathrm{wt}(c) = n - \frac{q^{k-1}}{q-1} = \frac{q^k - 1}{q-1} - \frac{q^{k-1}}{q-1} = q^{k-1}.$$

(b) Die Länge ist offensichtlich $n = \frac{q^k-1}{q-1}$. Es gilt

$$\dim_K(\mathrm{Sim}_q(k)) = \dim_K(\mathrm{Ham}_q(k)^\perp) = n - (n-k) = k.$$

Nach (a) ist die Minimaldistanz $d(\mathrm{Sim}_q(k)) = q^{k-1}$.

□

Um die Minimaldistanz von linearen Codes zu berechnen, ist es hilfreich zu wissen, welche Gewichte höchstens auftreten können.

Definition. *Ein Code C heißt r-*dividierbar *($r \in \mathbb{N}$), falls für alle $c \in C$ gilt $r \mid wt(c)$. Ist $r = 2$ bzw. $r = 4$ so heißt C* gerader *bzw.* doppelt gerader Code.

Lemma (Dividierbarkeitslemma)**.** *Sei C ein selbstdualer binärer Code der Länge n. Dann gilt*

(a) C ist gerade.

(b) Ist $4 \mid wt(c)$ für alle c aus einer Basis von C. Dann ist C doppelt gerade.

Beweis. (a) Sei $c = (c_1, \ldots, c_n) \in C = C^\perp$. Wir erhalten

$$0 = \langle c, c\rangle = \sum_{i=1}^n c_i^2 = \sum_{i=1}^n c_i = \sum_{i=1,\ c_i \neq 0}^n 1 = \mathrm{wt}(c)1,$$

also $2 \mid \mathrm{wt}(c)$.

(b) Seien $c = (c_1, \ldots, c_n), c' = (c_1', \ldots, c_n') \in C$ mit $4 \mid \mathrm{wt}(c), \mathrm{wt}(c')$. Für die Träger $\mathrm{Tr}(c) = \{i \mid c_i \neq 0\}$ und $\mathrm{Tr}(c') = \{i \mid c_i' \neq 0\}$ gilt

$$\mathrm{wt}(c + c') = \mathrm{wt}(c) + \mathrm{wt}(c') - 2 \mid \mathrm{Tr}(c) \cap \mathrm{Tr}(c') \mid .$$

Es gilt

$$0 = \langle c, c'\rangle = \sum_{i=1,\ c_i = c_i' = 1}^n 1 = \mid \mathrm{Tr}(c) \cap \mathrm{Tr}(c') \mid 1,$$

also $2 \mid \mid \mathrm{Tr}(c) \cap \mathrm{Tr}(c') \mid$. Damit gilt $4 \mid \mathrm{wt}(c + c')$.

□

Beispiel. *Wir betrachen zunächst den Hamming-Code $C = Ham_2(3)$ mit Parameter $\left[\frac{2^3-1}{2-1}, \frac{2^3-1}{2-1} - 3, 3\right] = [7, 4, 3]$. Die Matrix*

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

ist eine Erzeugermatrix von $Ham_2(3)$. Fügen wir nun an jedes Codewort ein Kontrollbit hinzu, so erhalten wir den erweiterten Hamming-Code

$$\widehat{C} = \{(c_1, \ldots, c_8) \mid (c_1, \ldots, c_7) \in C,\ \sum_{i=1}^7 c_i = c_8\}$$

mit Erzeugermatrix

$$\widehat{G} = \left(\begin{array}{c|c} & 1 \\ G & \vdots \\ & 1 \end{array} \right) = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

Seien $z_1, \ldots, z_4$ die Zeilen von $\widehat{G}$, so gilt $\langle z_i, z_j \rangle = 0$ für $1 \leq i, j \leq 4$ und somit $\widehat{C} \subseteq \widehat{C}^\perp$. Wegen

$$dim_{\mathbb{F}_2}(\widehat{C}) = 4 = 8 - dim_{\mathbb{F}_2}(\widehat{C}) = dim_{\mathbb{F}_2}(\widehat{C}^\perp).$$

Also ist $C = \widehat{C}$ ein selbstdualer Code. Weil die Zeilen von $\widehat{G}$ Gewicht 4 haben, ist $\widehat{C}$ ein doppelt gerader Code mit Parameter $[8, 4, 4]$.

3 Der binäre Golay-Code

Sei $C_1 = \widehat{C}$ der erweiterte Hamming-Code mit Parameter $[8, 4, 4]$, den wir bereits betrachtet haben. Sei $C_2 \subseteq \mathbb{F}_2^8$ erzeugt von

$$G_2 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Die Matrix G_2 entsteht durch die Permutation $(c_1, \ldots, c_8) \longmapsto (c_7, c_6, \ldots, c_1, c_8)$. Also sind C_1 und C_2 äquivalent und somit ist C_2 ein selbstdualer $[8, 4, 4]$-Code. Weiter gilt

$$C_1 \cap C_2 = \{(0, \ldots, 0), (1, \ldots, 1)\}.$$

Setze

$$\widetilde{C} = \{(c_1 + c_2, c_1' + c_2, c_1 + c_1' + c_2) \mid c_1, c_1' \in C_1, c_2 \in C_2\} \subseteq K^{24}.$$

Man nennt $\widetilde{C}$ den binären erweiterten Golay-Code.
Es gilt $\dim_{\mathbb{F}_2}(\widetilde{C}) = 12$ (Übung).
Bemerke, dass die Vektoren $(c_1, 0, c_1), (0, c_1', c_1'), (c_2, c_2, c_2)$ eine Basis enthalten und da diese paarweise orthogonal zueinander sind, ist $\widetilde{C} \subseteq \widetilde{C}^\perp$. Also gilt wegen

$$\dim_{\mathbb{F}_2}(\widetilde{C}^\perp) = 3 \cdot 8 - \dim_{\mathbb{F}_2}(\widetilde{C}) = 24 - 12 = 12$$

auch $\widetilde{C} = \widetilde{C}^\perp$. Der Code $\widetilde{C}$ ist ein selbstdualer doppelt gerader $[24, 12]$-Code.
Welche Minimaldistanz d besitzt $\widetilde{C}$? Da $\widetilde{C}$ doppelt gerade ist, ist $d \in 4\mathbb{N}$. Das Wort $c_1 \in C_1$ hat Gewicht 4 und $(c_1, 0, c_1) \in \widetilde{C}$ hat Gewicht 8. Also ist $d \in \{4, 8\}$. Im Folgenden zeigen wir, dass $d = 8$.
Es gilt für $x \in C_1$ und $y \in C_2$

$$\mathrm{wt}(x + y) = \mathrm{wt}(x) + \mathrm{wt}(y) - 2 \mid \mathrm{Tr}(x) \cap \mathrm{Tr}(y) \mid .$$

Aus $4 \mid \mathrm{wt}(x)$ und $4 \mid \mathrm{wt}(y)$ folgt daher $2 \mid \mathrm{wt}(x + y)$. Daher ist das Gewicht jeder Komponente von

$$0 \neq c = (c_1 + c_2, c_1' + c_2, c_1 + c_1' + c_2)$$

gerade. Sind alle Komponenten ungleich 0, gilt $\mathrm{wt}(c) \geq 6$ und wegen $4 \mid \mathrm{wt}(c)$ ist $\mathrm{wt}(c) = 8$. Angenommen eine ist 0, so ist $c_2 \in C_1$ und daher $c_2 \in \{(1, \ldots, 1), (0, \ldots, 0)\}$. In beiden Fällen ist $\mathrm{wt}(c) = 8$.
Letzteres kann man elementar nachrechnen. Sei z.B. $c_2 = (1, 1, 1, 1, 1, 1, 1, 1)$ und $c_1 + c_2 = 0$, so folgt $c_1 = c_2$ und damit

$$c = (c_1 + c_2, c_1' + c_2, c_1 + c_1' + c_2) = (0, c_1' + c_2, c_1').$$

Nun gilt $\mathrm{wt}(c) = \mathrm{wt}(c_1' + (1, 1, 1, 1, 1, 1, 1, 1)) + \mathrm{wt}(c_1') = (8 - \mathrm{wt}(c_1')) + \mathrm{wt}(c_1') = 8$. Die anderen Fälle berechnet man analog.
Zusammengefasst: der binäre erweiterte Golay-Code ist ein selbstdualer doppelt gerade Code mit Parameter $[24, 12, 8]$.

3.1 Der binäre Golay Code

Durch Streichung der letzten Zeile von $\widetilde{C}$ erhalten wir den sogenannten binären Golay-Code C mit Parameter $[23, 12, 7]$.
Es gilt $(0, \ldots, 0, 1, \ldots, 1) \in \widetilde{C}$ (8 Einsen). So folgt durch Streichen direkt $d(C) = 7$. Nach Satz 10.4 folgt aus $d(C) \geq 2$

$$12 = \dim_{\mathbb{F}_2}(C) = \dim_{\mathbb{F}_2}(\widetilde{C}).$$

Wegen

$$2^{23} = |\mathbb{F}_2|^{23} = |C| \sum_{j=0}^{3} \binom{23}{j} = 2^{12} 2^{11}$$

ist C perfekt.

Theorem 3.1. *Sei C ein binärer $(24, 2^{12}, 8)$-Code, der die 0 enthält, dann ist C ein doppelt gerader selbstdualer linearer $[24, 12, 8]$-Code, der äquivalent ist zum erweiterten binären Golay-Code.*