Zusammenfassung Informations- und Kodierungstheorie

Henrik Tscherny

23. August 2021

Inhaltsverzeichnis

1 Grundlagen

2 Informationsquellen

Die verschiedenen Informationsquellen können wie folgt eingeordnet werden:

- diskrete Quellen
  - Einzelquellen
    - Quellen mit unabhängigen Ereignissen
    - Quellen mit abhängigen Ereignissen (Markow-Quellen)
  - Verbundquellen
- kontinuierliche Quellen

2.1 Quellen mit unabhängigen Ereignissen

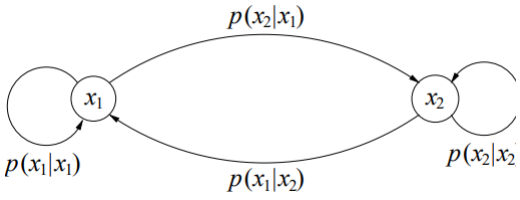
Entropie

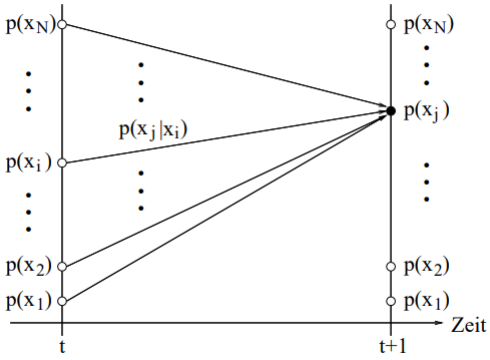
- Sei $X = \{x_1, ..., x_N\}$ das Quellenalphabet
- Sei p_i eine Menge von Auftrittswahrscheinlichkeiten mit $\sum_i^n p(x_i) = 1$

Entropie für das Zeichen x_i: $H_i = ld(\frac{1}{p(x_i)}) = -ld(p(x_i))$
Quellenentropie: $H_m = \sum_i^N p(x_i)\ H_i = -\sum_i^N p(x_i) ld(p(x_i))$
Sonderfall für Gleichverteilung der Zeichen: $H_0 = ld(N)$ gdw. $p(x_i) = \frac{1}{N}$

2.2 Quellen mit abhängigen Ereignissen (Markow-Quellen)

- Ein Ereignis tritt immer unter einer (Kette von) Vorbedingung ein (außer das erste Ereignis)
- Die Auswahl des nächsten Ereignisses erfolgt nach einer bedingten Wahrscheinlichkeit abhängig von den Ereignissen zuvor
- $p(x^{(m+1)}|x^m \cdots x^{(2)}\ x^{(1)})$ Wahrscheinlichkeit von Ereignis (m+1) wenn Kette aus Ereignissen (m...1) eingetreten sind
- Eine Markow-Quelle erster Ordnung berücksichtigt nur das letzte Ereignis
  - $p(x^{(m+1)}|x^{(x)})$ mit $p(x_j|x_i)$ (Wahrscheinlichkeit von x_j wenn x_i)



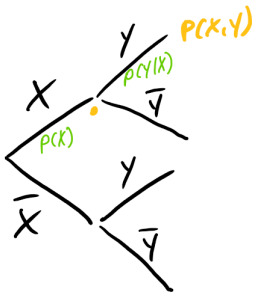


Entropie

- **Entropie für ein Zeichen:** $H_i = \sum_j^N p(x_j|x_i)\ ld(\frac{1}{p(x_j)|x_i)})$
- **Markow-Entropie:** $H_M = \sum_i^N \sum_j^N \overline{p(x_i)}\ p(x_j|x_i)\ ld(\frac{1}{p(x_j|x_i)}$

2.3 Verbundquellen

- Seien X und Y diskrete Quellen
- Sei $(p(x_i))$ und $(p(y_j)$ die zugehörigen Auftrittswahrscheinlichkeiten
- Die Ereignisse in X und Y separat sind **unabhängig** voneinander
- Ein Ereignis in X hat ein **bedingtes Ereignis** in Y $(p(y_j|x_i))$ zur Folge
- Das Auftreten zweier Ereignisse x_i und y_j heißt **Verbundereignis** (x_i, y_j)
- Die Wahrscheinlichkeit, dass x_i und x_j eintreten heißt **Verbundwahrscheinlichkeit** $p(x_i, y_j) = p(x_i) \cdot p(y_j|x_i)$ (Satz von Bayes)



Note: $p(x_i, y_j) = p(x_i \cap y_j)$

Entropie

- Sei Seien X und Y diskrete Quellen mit $p(x_i, y_j) \rightarrow$ **Verbundquelle (X,Y)**
- **Verbundentropie:** $H(X,Y) = \sum_i^N \sum_j^M\ p(x_i, y_j)\ ld(\frac{1}{p(x_i, y_j)})$
  $= H(X,Y) = H(X) + \sum_i^N \sum_j^M\ p(x_i)\ p(y_j|x_i)\ ld(\frac{1}{p(x_i, y_j)})$

- **Sonderfall X,Y unabhängig:** $H(Y|X) = H(Y)$ und $H(X,Y) = H(X) + H(Y)$
- **Sonderfall vollständig abhängig:** $H(Y|X) = 0$ und $H(X,Y) = H(X)$
- **Sonderfall identische Ereignismengen:** $H_M = H(X,X) - H(X)$ und $H(X) \leq H(X,X) \leq 2H(X)$

3 Kodierung

Kodwortlänge

- **gleichmäßiger Kode**: $l = \lceil ld\, N \rceil$
- **ungleichmäßiger Kode**: $l_m = \sum_i^N \; p(x_i) l_i$
- **optimaler Kode**: Sein $q, d_{min}, n \in \mathbb{N}$, mit n... Länge, q-närer Zeichenvorrat d_{min}... Mindestdistanz. Es gibt keinen Kode welcher unter diesen Parameter mehr Wörter kodiert

Dekodierbarkeit

- *Hinreichend*: Das Ende eines Kodewortes darf nicht gleich dem Anfang eines anderen Kodewortes sein (**Präfixfrei**)
- *Notwendig*: $\sum_i^N \; 2^{-l_i} \leq 1$ (l_i = Kodewortlänge)
  Auftrittswahrscheinlichkeiten der Kodewörter sind nur Zweierpotenzen
- $l_m \geq H_m$

Redundanz

- **redundanzarm**: $H_m \leq l_m < H_m + 1$
- **redundanzfrei**: $l_m = H_m$
- **Koderedundanz**: $R_K = l_m(\cdot H_K) - H_Q \geq 0$ (H_K meist 1 → fällt weg)
- **Mittlere Kodewortlänge** l_m Wahrscheinlichkeit des Zeichens * dessen Länge dessen Kodewortes

3.1 Shannonsche Kodierungstheoreme

3.1.1 1. Theorem

- Es ist eine redundanzfreie Kodierung auch für Auftrittswahrscheinlichkeiten welche keine Zweierpotenz sind möglich
  $p(x_i) \neq 2^{-l_i}$
- **Lösung:** m-Fache Erweiterung der Quelle
  → Kodieren der Quellzeichen als Blöcke von m Zeichen
- bildet Grundlage für die **Entropiekodierung**

3.1.2 2. Theorem

- Die **Restfehlerwahrscheinlichkeit** p_R kann beliebig klein gehalten werden
- Dazu muss aber die Koderate R den Wert der Maximalen Transinformation H_T nicht überschreiten

3.2 Verfahren

3.2.1 Shannon-Fano-Verfahren

1. Ordne Zeichen abfallend nach ihrer Auftrittswahrscheinlichkeit
2. Teile nun in zwei Teilmengen mit möglichst gleicher Wahrscheinlichkeitssumme
3. Weise den entstandenen Gruppen 1 (bzw. 0) zu (Reihenfolge egal)
4. GOTO 1 until Teilmengen nicht mehr teilbar

3.2.2 Huffman-Verfahren

1. Ordne Zeichen abfallend nach ihrer Auftrittswahrscheinlichkeit
2. Fasse Zeichen(menge) mit der kleinsten Auftrittswahrscheinlichkeit(-Summe) zusammen
3. Ordne neu erhaltene Zeichenmengen nach ihren Auftrittswahrscheinlichkeiten
4. GOTO 2 while SUM $\neq 1$
5. weise den entstandenen Ästen des Baumes 1 (bzw. 0) zu

3.3 Quellenkodierung

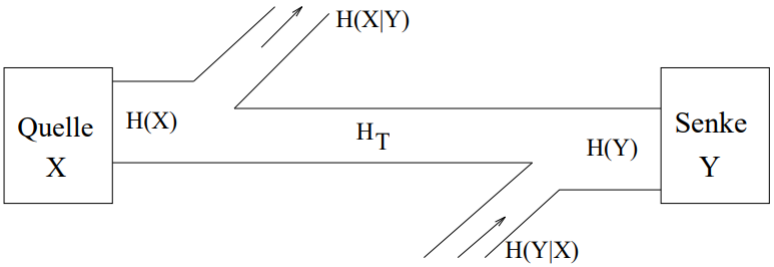
- Erster Schritt der Kodierung
- möglichst redundanzfrei/arm
- **verlustfrei:** Redundanzreduktion
- **verlustbehaftet:** Irrelevanzreduktion

3.4 Kanalkodierung

- erfolgt nach Quellenkodierung
- dient Schutz vor Störungen und Veränderungen
- gezieltes hinzufügen von Redundanz (Kontrollinformationen)

4 Kanäle

4.1 Bergersches Entropiemodell des Übertragsungskanal



- $H(X)$: Entropie Kanaleingang
- $H(Y)$: Entropie Kanalausgang
- H_T: Transinformation
- $H(X|Y)$: Äquivokation (Rückschlussentropie)
- $H(Y|X)$: Irrelevanz (Störentropie)

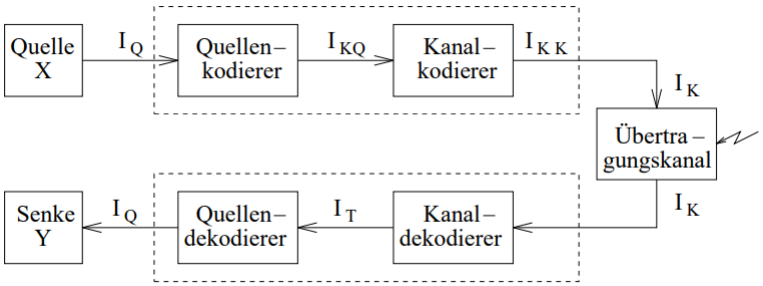
Note: Im Idealfall (ungestört) gilt $H(X) = H(Y) = H_T$

Transinformation
Ist die im Mittel durch ein Kanalzeichen übertragene Informationsmenge

- $H_T = H(X) + H(Y) - H(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$
- Transinformation = Eingang minus 'Alles was verloren geht'
- Transinformation = Ausgang minus 'Alles was dazu gekommen ist'

4.2 Kanalkapazität

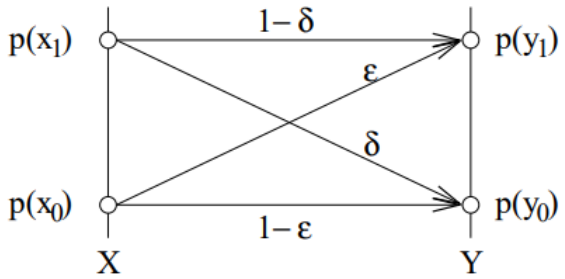
Die Kanalkapazität C ist der Maximalwert des Transinformationsflusses I_T ($\rightarrow C = max\{I_T\}$



- I_Q: Quelleninformationsfluss ($I_Q = f_Q H_Q$)
- I_{KQ}: Quellenkodeinformationsfluss ($f_Q l H_K$ mit $l = \lceil \frac{ld\ N}{H_K} \rceil$)
- I_{KK}: Kanalkodeinformationsfluss ($f_Q(l + \Delta l)H_K = f_Q n H_K$ mit Kanalkode $n = l + k, k \geq \rceil \Delta l \lceil$))
- I_K: Kanalinformationsfluss (= Übertragungsgeschwindigkeit $v_{ü}$) ($v_{ü} = I_K = v_s H_K$
- I_T: Transinformationsfluss ($I_T = v_s H_T$)
- f_Q: Quellensymbolfrequenz
- f_K: Kanalsymbolfrequenz

Note: Der Transinformationsfluss I_T eines gestörten Kanal ist immer kleiner als die Übertragungsgeschwindigkeit $v_{ü} = I_K$

4.3 diskrete Binärkanal



- δ: Wahrscheinlichkeit, dass das Zeichen x_1 in y_0 verfälscht wird
- ϵ: Wahrscheinlichkeit, dass das Zeichen x_0 in y_1 verfälscht wird

- $1-\delta$, $1-\epsilon$: Wahrscheinlichkeit, dass das gesendete Zeichen richtig Übertragen wird
- $(p(y_j|x_i)) = \begin{pmatrix} 1-\epsilon & \epsilon \\ \delta & 1-\delta \end{pmatrix}$

gesichert

- $f_Q = \frac{v_{ü}}{n}$ $(v_{ü} = v_s)$, $(n = l H_K^2)$, $(n = \frac{H_K}{H_T})$
- $n = \frac{l}{H_T}$
- $f_Q = \frac{v_{ü} H_t}{l}$

ungesichert

- $v_s = f_q l$
- $f_q = \frac{v_{ü}}{l}$

symmetrisch gestört

$\epsilon = \delta = p_s$ (Die Wahrscheinlichkeit, dass ein Zeichen in ein anderes verfälscht wird, ist für alle Zeichen gleich

- $H_T = H(Y) - ((1-p_s)\, ld\, \frac{1}{(1-p_s)} + p_s\, ld\, \frac{1}{p_s})$
- $H_{T_{max}} = 1 - ((1-p_s)\, ld\, \frac{1}{(1-p_s)} + p_s\, ld\, \frac{1}{p_s})$

einseitig gestört

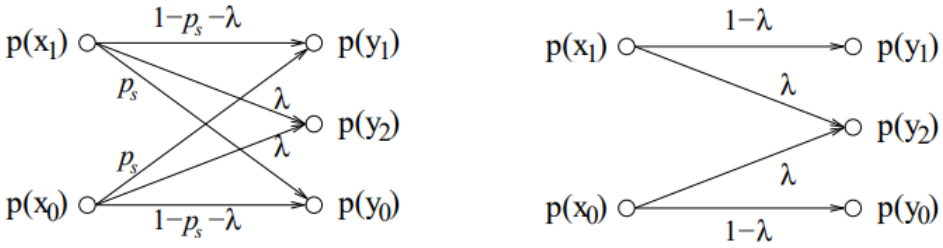
$\epsilon = p_s$ $\delta = 0$ Nur ein Zeichen kann mit einer Wahrscheinlichkeit in ein anderes umgewandelt werden, das Andere wird immer sicher übertragen

- $H_T = H(Y) - p(x_0)\, ((1-p_s)\, ld\, \frac{1}{(1-p_s)} + p_s\, ld\, \frac{1}{p_s})$
- $H_T = 1 + \frac{1}{2}\, ((1+p_s)\, ld\, \frac{1}{(1+p_s)} - p_s\, ld\, \frac{1}{p_s})$

Kanal mit Auslöschung

Es wird ein Zusätzliches **Auslöschungszeichen** eingeführt mit Übergangswahrscheinlichkeit λ

- $H_{T_{max}} = (1-\lambda) - p_s\, ld\, \frac{1}{p_s} + (1-\lambda)\, ld\, \frac{1}{(1-\lambda)} - (1-p_s\lambda)\, ld\, \frac{1}{(1-p_s-\lambda)}$
- Das Modell kann so umgeformt werden, dass Zeichen nicht in ein gültiges Kanalzeichen umgeformt werden können,
  sondern nur in ein das Auslöschungszeichen (hier y_2)
  → bessere Fehlerbehandlung
- $H_{T_{max}} = (1-\lambda)$



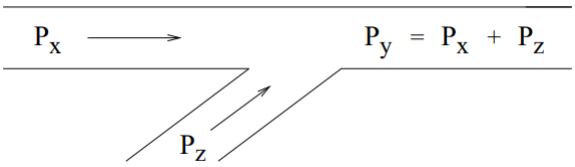
(a) Umwandlung in anderes Kodezeichen möglich

(b) Umwandlung nur in Auslöschungzeichen möglich

4.4 analoge Kanäle

Bei einem analogen Kanal ist es nicht möglich den Wert eines Zeichens klar zu bestimmen, vielmehr wird der Wert anhand einer Funktion beschrieben
Man kann dann die Wahrscheinlichkeit berechnen das der Wert sich in einem gewissen Intervall befindet
$p(x_i) = \int_{\Delta x} f(x)dx \approx f(x_i)\Delta x$

- $H_{diskr} = H_m \sum_i f(x_i)\, \Delta\, ld(\frac{1}{f(x_i)\Delta})$
- $H_{an} = \int_{-\infty}^{\infty} f(x)\, ld(\frac{1}{f(x)})dx - ld\ \Delta x$
- Da $ld\,\Delta x$ meist 0 nimmt man oft die **relative Entropie**
  $H_{rel} = \int_{-\infty}^{\infty} f(x)\, ld(\frac{1}{f(x)})dx$



- Signale und Störungen addieren sich → beide am Ausgang als Summe
- keine Störsignale welche vom Nutzsignal abhängig sind
  → Leistung des Empfangssignals ist die Summe aus Nutz- und Störsignalleistung

Transinformation

- **Entropie**: $H(X) = \frac{1}{2} ld(2\pi\epsilon P_x)$
- **Störentropie**: $H(Y|X) = \frac{1}{2} ld(2\pi\epsilon P_z)$
- **Entropie Kanalausgang**: $H(Y) = \frac{1}{2} ld(2\pi\epsilon(P_x + P_z))$
- **Transinformation**: $H_T = H(Y) - H(Y|X) = \frac{1}{2} ld(1 + \frac{P_x}{P_z})$
  wenn $\frac{P_x}{P_z} \gg 1$ dann $H_T \approx \frac{1}{2} ld \frac{P_x}{P_z}$
- **Rauschabstand**: $r = 10 lg(\frac{P_x}{P_z})$
- **Kanalkapazität**: $C_{an} = 2B\, H_T = 2B\frac{1}{2}(1 + \frac{P_x}{P_z})$

Quantisierung

- Zeitquantisierung
  - $f_A \geq 2f_g$
  - $t_A \leq \frac{1}{2f_g} = \frac{1}{f_A}$
    → bei Einhaltung kein Informationsverlust
- Amplitudenquantisierung
  - Informationsverlust

5 Hamming-Code

5.1 Hamming-Distanz

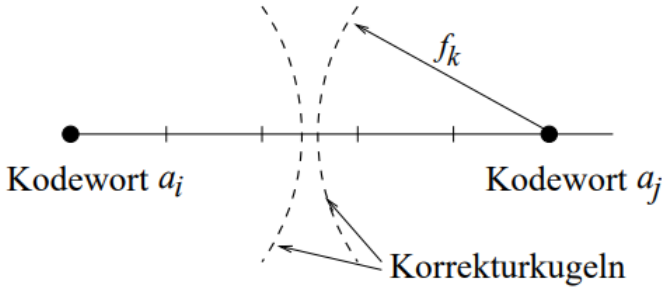
Die **Hamming Distanz** ($d(a_i, a_j)$) ist die Anzahl an Stellen um welche sich zwei Kodewörter unterscheiden. Für einen Binärkode gilt:

- **Hamming Distanz**: $d(a_i, a_j) = \sum_{g=1}^{n} (u_{ig} \oplus u_{jg})$ (Summe der XOR's aller Wörter)
- **Mindestdistanz**: $d_{min} = \min_{a_i, a_j \in A, a_i \neq a_j} d(a_i, a_j)$
- **Hamming Gewicht**: $w(a_i) = \sum_{g=1}^{n} u_{ig} = d(0, a_i)$

Fehlerkorrektur/erkennung

Durch die Hamming Distanz können Fehler in den Kodewörten erkannt werden. Indem Festgestellt wird, dass das erhaltene Kodewort nicht in der Gültigen Kodewortmenge liegt, kann erkannt werden, dass ein Fehler vorliegt (f_e). Liegt zudem das fehlerhafte Kodewort näher an einem anderen gültigen Kodewort (d.h. nicht exakt in der Mitte) so kann der Fehler sogar korrigiert werden (f_k)
Es gilt:

- $d_{min} = f_e + f_k + 1$
- $f_e = \lfloor \frac{d_{min}}{2} \rfloor$
- $f_k = \lfloor \frac{d_{min}-1}{2} \rfloor$



Hamming-Schranke
Anzahl der benötigten redundanten Stellen k

- $l = n - k$
- $r_k = \frac{n-l}{n} = \frac{k}{n}$ (relative Redundanz)
- $R = \frac{l}{n}$ (Koderate)

Ein Kode heißt **Perfekt** oder **dicht gepackt** wenn jedes Kodewort nur zu einem Wort einen geringsten Hammingabstand hat → jeder erkannte Fehler kann auch erkannt werden, da ein verfälschtes Kodewort immer einem Kodewort zugeordnet werden kann
$k \geq log_2(\sum_{i=0}^{f_k=1} \binom{n}{i})$ mit $n = l + k$

5.2 fehlerkorrigierender Hammingcode

- $d_{min} = 3$, somit gilt auch $f_k = 1, f_e = 1$
- $n = 2^k - 1$

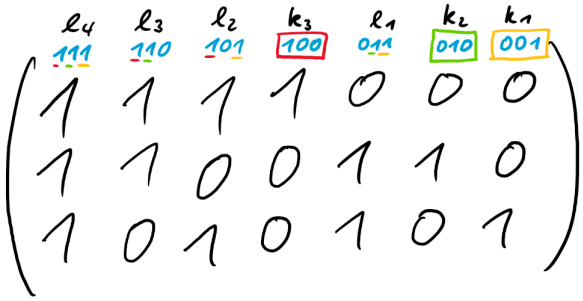


Abbildung 3: k... Kontrollstellen, l... Informationsstellen

Kontrollmatrix

- $k_3 = l_4 \oplus l_3 \oplus l_2$ (Stelle 3 := 1)
- $k_2 = l_4 \oplus l_3 \oplus l_1$ (Stelle 2 := 1)
- $k_1 = l_4 \oplus l_2 \oplus l_1$ (Stelle 1 := 1)

Fehlersyndrom How To machen:

1. Kontrollbit mit den dazugehörenden Informationsbits XORen
2. mach 1) für jede Zeile der Matrix
3. der dadurch erhaltene Vektor ist das Fehlersyndrom S
    - S enthält nur Nullen $S = \vec{0} \rightarrow$ kein Fehler gefunden
    - S enthält eine/mehrere Nullen $S \neq \vec{0} \rightarrow$ ließ S als Zahl und flippe das Bit an dieser Stelle
      z.B. $S = (1, 0, 0, 1) \rightarrow$ flippe Stelle 9

5.3 Erweiterter Hamming-Code

- Fügt jedem Kanalkodewort ein weiteres Kontrollbit k_0 hinzu
- k_0 wird dann hinten einfach angehangen
- $n_0 = k_0 = \sum_i^n n_i mod 2$ (gerades Paritätsbit)
- Erhöht d_{min} auf 4 (von 3)

6 Lineare Blockkodes (Linearkode)

Ein Kode ist einer linear Blockkode wenn die Kodierungfunktion eine Verknüpfungsoperation einer Gruppe ist
Eigenschaften:

- Abgeschlossenheit
- Assoziativität
- neutrales Element
- inverses Element
- (kommutativ)* $\rightarrow$ abelsche Gruppe

Durch die obrigen Eigenschaften ergibt sich demnach auch:

- eine Verknüpfung eines Kanalkodewortes ergibt wieder ein Kanalkodewort (Abgeschlossenheit)
- Nullwort ist immer ein Kanalkodewort (neutrales Element)

- Fehlererkennung/korrektur:
  - $f_e = d_{min} - 1$
  - $f_k = \lfloor \frac{d_{min}-1}{2} \rfloor$, $f_e = \lfloor \frac{d_{min}}{2} \rfloor$

Systematischer Kode

Ein Linearkode heißt systematisch, wenn man durch das Streichen der redundanten Stellen das Quellenwort erhält.
→ z.B. Hamming Code durch streichen der Stellen 2^i

6.1 Generatormatrix

How To machen:

1. jede Zeile von G muss ein Kodewort a sein
2. die Kodewörter müssen so ausgewählt werden das eine Einheitsmatrix der Größe l am Anfang entsteht (I_n)
3. Somit gilt dann $G_{l \times n} = [I_n C]$
4. Kodewörter können nun durch das Multiplizieren mit G kodiert werden $a_i = a_i^* \cdot G$

6.2 Kontrollmatrix

How To machen:

1. Sei C die Generatormatrix G ohne die Einheitsmatrix ($G_{l \times n} = [I_n C]$)
2. Transponiere C, dies bildet die ersten Spalten der Kontrollmatrix H
3. fülle den Rest (so das man auf n Spalten kommt) mit einer Einheitsmatrix (I_K) passender Größe auf
4. somit gilt dann: $H_{k,n} = [C^\top I_K]$
5. zum Prüfen eines Kodewortes multipliziere es mit der Kontrollmatrix

   - Das Ergebnis ist gleich $\vec{0}$ → kein Fehler (oder zumindest keiner erkannt)
   - Das Ergebnis ist ungleich $\vec{0}$ → Fehler

- Das Ergebnis gleicht einer Spalte n_i in H → flippe das Bit i
- Das Ergebnis gleich keiner Spalte in H → nicht korrigierbar (Mehrfachfehler ?)

7 Zyklische Kodes

Ein Kode in welchem durch zyklische Verschiebung der Elemente wieder ein Kanalkodewort entsteht
Ein zyklischer Kode ist ein spezieller Linearkode welcher Ring also auch Körperaxiome erfüllt
BCH-Codes können Bündelfehler f_b erkennen mit $f_b \leq k$

- $m_1(x) = M(X)$
- $m_0(x) = (x + 1)$ (Abramson)

7.1 Modularpolynome M(X)

- **irreduzibel**: Ist nicht in ein Produkt von Polynomen zerlegbar
  $M(X)$ bestimmt n mit $n \leq 2^{grad(M(X))} - 1$
- der tatsächliche Wert von n ergibt sich aus dem Zyklus der Polynomreste mit $n = p|2^{grad(M(X))} - 1$
- **primitiv**: gilt $n = p = 2^{grad(M(X))} - 1$ dann ist $M(X)$ auch primitiv
  (es existiert eine Nullstelle α in $GF(p^m)$, so dass
  $\{\alpha^n | \forall n \in \{0, ..., p^m - 2\}\} = GF(p^m)$
  Ein primitives Polynom ist wenn seine Nullstelle Ordnung $n = p^m - 1$ ist
  ist $f(x)$ irreduzibel dann reicht $ggt(k, n) \neq 1$
- Die Leistung des Kodes hängt von der Anzahl aufeinanderfolgender Nullstellen ab

Fundamentalsatz der Algebra
Jedes Polynom r-ten Grades hat mindestens eine Nullstelle (ggf. in einem andern Körper) und lässt sich in genau r Linearfaktoren zerteilen (mit Zuhilfenahme von Erweiterungselementen α_i
Nimmt man nun eine Nullstelle α zu $M(X)$ hinzu entsteht ein endlicher Erweiterungskörper $GF(2^{k_1})$ mit Nullstelle α von $M(X)$ ($k_1 = grad(M(X))$)

7.2 Generatorpolynom

- Produkt von Minimalpolynom $m_i(x)$
- beschreibt den Kode vollständig
- $n = 2^{grad(M(x))} - 1$ (da $M(X)$ primitiv)
- $k = grad(g(x))$
- $l = n - k$
- d_{min} tatsächlich aufeinanderfolgende Nullstellen + 1

7.3 Verfahren

7.3.1 Multiplikationsverfahren

How To machen:

1. Multipliziere das Kodepolynom mit dem Generatorpolynom ($a(x) = a^*(x)g(x)$)

7.3.2 Divisionsverfahren

How To machen:

1. Multipliziere das Kodepolynom mit x^k ($k = grad(M)$)
2. Teile das Ergebnis mittel Polynomdivision über GF(2) und ermittle den Rest $r(x)$
3. Das Fertige Ergebnis ist nun $a(x) = a^*(x) \cdot x^k + r(x)$

Fehlererkennung

- $a(x) mod g(x) = 0$ kein Fehler
- sonnst: Fehler

Kodes

- Zyklischer Hamming Code: $d_{min} = 3$
- Abramson Code: $d_{min} = 4$

Kodegeneration

- Entwurfsabstand d_E
- $g(x) = kgV\{m_\mu(x), ..., m_{\mu+d_E-2}(x)\}$ mit $\mu\{0, 1\}$