3 Kanalcodierung

3.1 Das Modell

$A = \{a_1, \ldots, a_m\}$ ist das Eingangsalphabet in den Kanal (aus dem Quellencodierer) und $B = \{b_1, \ldots, b_n\}$ das Ausgangsalphabet aus dem Kanal. Gegeben sind die W-Verteilung $\underline{p} = (p(a_1), \ldots, p(a_m))$ und die Kanalmatrix $P = (p_{ij})$ mit $p_{ij} = p(b_j|a_i)$.

$$\begin{array}{c} \\ a_1 \\ \vdots \\ a_i \\ \vdots \\ a_m \end{array} \begin{array}{c} \begin{array}{cccccc} b_1 & \ldots & b_j & \ldots & b_n \end{array} \\ \begin{pmatrix} & & & & \\ & & & & \\ & & p(b_j|a_i) & & \\ & & & & \\ & & & & \end{pmatrix} \end{array}$$

P ist eine $m \times n$-Matrix und die Einträge sind die bedingten W-keiten $p(b_j|a_i)$, dass b_j empfangen wurde, falls a_i in den Kanal eingegeben wurde.

Wir haben

$$p(a_i, b_j) = p(a_i)p(b_j|a_i) = p(b_j)p(a_i|b_j),$$

$$p(a_i) = \sum_{j=1}^{n} p(a_i, b_j), p(b_j) = \sum_{i=1}^{m} p(a_i, b_j)\,,$$

$$\sum_{j=1}^{n} p(b_j|a_i) = \sum_{j=1}^{n} \frac{p(a_i, b_j)}{p(a_i)} = 1\,.$$

Jede Zeile von P summiert also zu 1, und wir nehmen an, dass in jeder Spalte mindestens ein Element $\neq 0$ steht (da ansonsten dieses Element niemals empfangen wird).

Es sei $\underline{q}_j = (q_1, \ldots, q_n)$, $q_j = p(b_j)$ der Ausgangsvektor, dann gilt für alle j

$$q_j = p(b_j) = \sum_{i=1}^{m} p(a_i, b_j) = \sum_{i=1}^{m} p(a_i)p(b_j|a_i)\,,$$

also erhalten wir das Vektor-Matrixprodukt

$$\underline{q} = \underline{p}P\,.$$

Ferner ist

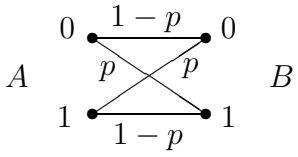
$$\sum_{j=1}^{n} q_j = \sum_{j=1}^{n}\sum_{i=1}^{m} p(a_i)p(b_j|a_i) = \sum_{j=1}^{n}\sum_{i=1}^{m} p(a_i,b_j) = \sum_{j=1}^{n} p(b_j) = 1\,,$$

das heißt $\underline{q}$ ist W-Verteilung.

Beispiele.
1. Binärer symmetrischer Kanal.

$$A = B = \{0,1\},\ P = \begin{array}{c} \\ 0 \\ 1 \end{array}\begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} \end{array} \quad \text{für } p < \tfrac{1}{2}.$$

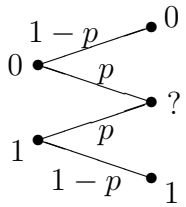
Eine übliche Darstellungsform ist das Kanaldiagramm:



2. Der q-äre symmetrische Kanal.
$A = B = \{a_1, \ldots, a_q\}$, $q \geq 2$, mit der Kanalmatrix

$$\begin{array}{c} \\ a_1 \\ a_2 \\ \vdots \\ a_q \end{array}\begin{array}{c} \begin{array}{cccc} a_1 & a_2 & \ldots & a_q \end{array} \\ \begin{pmatrix} 1-p & \frac{p}{q-1} & & \frac{p}{q-1} \\ \frac{p}{q-1} & 1-p & \ldots & \frac{p}{q-1} \\ & & & \\ \frac{p}{q-1} & \frac{p}{q-1} & \ldots & 1-p \end{pmatrix} \end{array}$$

3. Binärer auslöschender Kanal

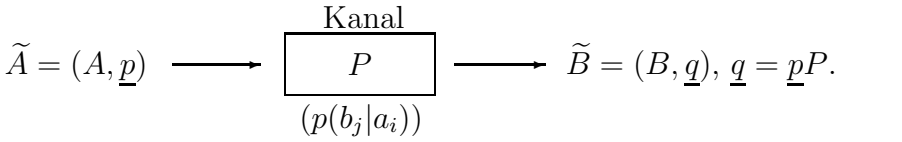


$$P = \begin{array}{c} \\ 0 \\ 1 \end{array}\begin{array}{c} \begin{array}{ccc} 0 & ? & 1 \end{array} \\ \begin{pmatrix} 1-p & p & 0 \\ 0 & p & 1-p \end{pmatrix} \end{array}$$

Die Elemente werden mit W-keit $1-p$ richtig übertragen und mit W-keit p "verwischt".

3.2 Bedingte Entropie, Informationsaustausch und Kapazität

Wir studieren nun das Modell genauer:



Die Idee ist, eine bedingte Entropie $H(\widetilde{A}|\widetilde{B})$ zu definieren mit der Interpretation: $H(\widetilde{A}|\widetilde{B})$ ist die Unsicherheit über $\widetilde{A}$ (das gesendete Symbol)unter der Kenntnis von $\widetilde{B}$ (dem empfangenen Symbol). Das Ziel ist, eine Codierung zu entwickeln, mit der wir $H(\widetilde{A}|\widetilde{B}) \to 0$ schließen können.

Es sei $\underline{p}_i = (p(b_1|a_i), \ldots, p(b_n|a_i))$ der i-te Zeilenvektor der Kanalmatrix P.

Definition. Wir setzen für alle i

$$H(\widetilde{B}|a_i) = H(\underline{p}_i) = -\sum_{j=1}^{n} p(b_j|a_i) \log p(b_j|a_i)$$

und

$$H(\widetilde{B}|\widetilde{A}) = \sum_{i=1}^{m} p(a_i) H(\underline{p}_i) = -\sum_{i=1}^{m} p(a_i) \sum_{j=1}^{n} p(b_j|a_i) \log p(b_j|a_i).$$

$H(\widetilde{B}|\widetilde{A})$ heißt die *bedingte Entropie* von $\widetilde{B}$ unter der Kenntnis von $\widetilde{A}$.
Es gilt

$$H(\widetilde{B}|\widetilde{A}) = -\sum_{i,j} p(a_i, b_j) \log p(b_j|a_i).$$

Satz 5. *Wir haben*

$$H(\widetilde{B}|\widetilde{A}) + H(\widetilde{A}) = H(\widetilde{A}, \widetilde{B})$$
$$H(\widetilde{A}|\widetilde{B}) + H(\widetilde{B}) = H(\widetilde{A}, \widetilde{B})\,.$$

Beweis. Es gilt

$$\begin{aligned} H(\widetilde{A},\widetilde{B}) &= -\sum_{i,j} p(a_i,b_j)\log p(a_i,b_j) \\ &= -\sum_{i,j} p(a_i,b_j)\log(p(a_i)\cdot p(b_j|a_i)) \\ &= -\sum_{i,j} p(a_i,b_j)\log p(a_i) - \sum_{i,j} p(a_i,b_j)\log p(b_j|a_i) \\ &= -\sum_{i=1}^{m} p(a_i)\log p(a_i) + H(\widetilde{B}|\widetilde{A}) \\ &= H(\widetilde{A}) + H(\widetilde{B}|\widetilde{A}). \end{aligned}$$

Die zweite Formel wird analog gezeigt.
Insbesondere folgt aus $H(\widetilde{A},\widetilde{B}) \le H(\widetilde{A}) + H(\widetilde{B})$,

$$H(\widetilde{B}|\widetilde{A}) \le H(\widetilde{B}), \quad H(\widetilde{A}|\widetilde{B}) \le H(\widetilde{A})$$

mit Gleichheit genau dann, wenn $\widetilde{A}$ und $\widetilde{B}$ unabhängig sind.

Definition. Der Ausdruck

$$I(\widetilde{A}\|\widetilde{B}) := H(\widetilde{A}) + H(\widetilde{B}) - H(\widetilde{A},\widetilde{B})$$

heißt der *Informationsaustausch* zwischen $\widetilde{A}$ und $\widetilde{B}$.

Es gilt also $I(\widetilde{A}\|\widetilde{B}) = I(\widetilde{B}\|\widetilde{A})$.

Satz 6. *Wir haben*

$$I(\widetilde{A}\|\widetilde{B}) = H(\widetilde{B}) - H(\widetilde{B}|\widetilde{A}) = H(\widetilde{A}) - H(\widetilde{A}|\widetilde{B}).$$

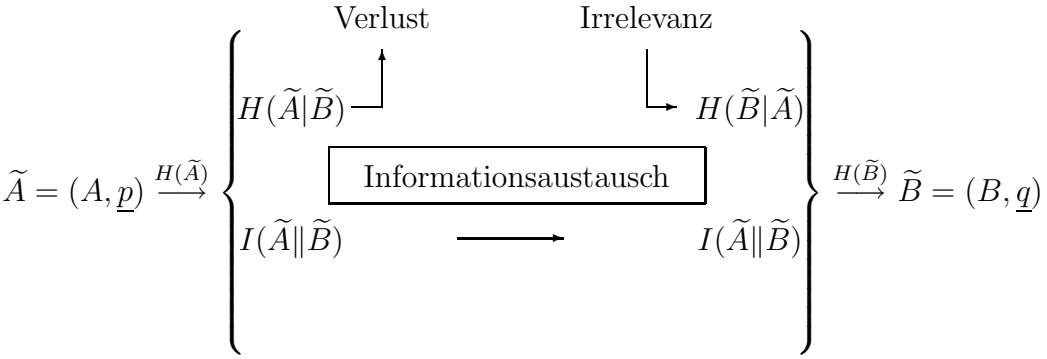
Beweis. Nach dem obigen Satz gilt

$$\begin{aligned} I(\widetilde{A}\|\widetilde{B}) =& H(\widetilde{A}) + H(\widetilde{B}) - H(\widetilde{A},\widetilde{B}) \\ =& H(\widetilde{A}) + H(\widetilde{B}) - H(\widetilde{A}) - H(\widetilde{B}|\widetilde{A}) \\ =& H(\widetilde{B}) - H(\widetilde{B}|\widetilde{A}). \qquad \square \end{aligned}$$

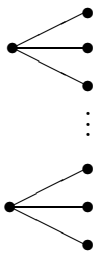
Wir haben somit

$$H(\widetilde{A}) = I(\widetilde{A}\|\widetilde{B}) + H(\widetilde{A}|\widetilde{B})$$
$$H(\widetilde{B}) = I(\widetilde{A}\|\widetilde{B}) + H(\widetilde{B}|\widetilde{A})$$

mit der folgenden Interpretation:

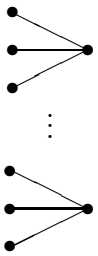


Beispiele.
1. Verlustloser Kanal



Hier ist $H(\widetilde{A}|\widetilde{B}) = 0$, da aus der Kenntnis von $b_j \in B$ sofort das gesendete Symbol $a_i \in A$ ermittelt werden kann.

2. Deterministischer Kanal



Hier ist $H(\widetilde{B}|\widetilde{A}) = 0$, es liegt keine Irrelevanz vor, die durch den Kanal verursacht wird. Das gesendete Symbol a_i determiniert das empfangene Symbol.

Der Informationsaustausch $I(\widetilde{A}\|\widetilde{B})$ hängt von der Eingangsverteilung $\underline{p}$ und

der Kanalmatrix P ab. Wir schreiben kurz $I(\underline{p})$.

Definition. $K = \max\limits_{\underline{p}} I(\underline{p})$ heißt die *Kapazität* des Kanals.

Bemerkung. Das Maximum existiert, da $I(\underline{p})$ eine stetige Funktion ist und der Definitionsbereich kompakt ist.

Beispiel. Für den symmetrischen q-ären Kanal ist die Kanalmatrix

$$P = \begin{pmatrix} 1-p & \frac{p}{q-1} & \dots & \frac{p}{q-1} \\ & & \dots & \\ \frac{p}{q-1} & & \dots & 1-p \end{pmatrix}, \; p < \frac{1}{2}\,.$$

Das ergibt

$$I(\underline{p}) = I(\widetilde{A}\|\widetilde{B}) = H(\widetilde{B}) - H(\widetilde{B}|\widetilde{A}) \le \log q - H(\widetilde{B}|\widetilde{A})\,.$$

Für $H(\widetilde{B}|\widetilde{A})$ erhalten wir

$$H(\widetilde{B}|\widetilde{A}) = \sum_{i=1}^{q} p(a_i) H(\underline{p}_i) = \sum_{i=1}^{q} p(a_i) \cdot H(\underline{p}_1) = H(\underline{p}_1),$$

da $H(\underline{p}_1) = H(\underline{p}_2) = \ldots = H(\underline{p}_q)$ ist. Somit ist

$$\begin{aligned} H(\widetilde{B}|\widetilde{A}) &= -(1-p)\log(1-p) - (q-1)\frac{p}{q-1}\log(\frac{p}{q-1}) \\ &= -(1-p)\log(1-p) - p\log(\frac{p}{q-1})\,, \end{aligned}$$

also

$$I(\underline{p}) \le \log q + (1-p)\log(1-p) + p\log(\frac{p}{q-1}).$$

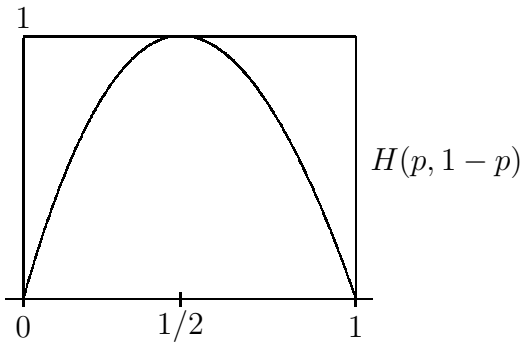
Nun gilt für $\underline{p} = (\frac{1}{q}, \ldots, \frac{1}{q})$, $\underline{q} = \underline{p}P = (\frac{1}{q}, \ldots, \frac{1}{q})$, somit

$$K = \log q + (1-p)\log(1-p) + p\log(\frac{p}{q-1}).$$

Insbesondere erhalten wir für $q = 2$ die Kapazität

$$K = K(p) = 1 + p\log p + (1-p)\log(1-p) = 1 - H(p, 1-p)\,.$$

Die Funktion $H(p, 1-p)$ ist symmetrisch um $p = \frac{1}{2}$



Die Kapazität $K = K(p)$ nimmt monoton für wachsendes $p < \frac{1}{2}$ ab mit $K(\frac{1}{2}) = 0$. Die Interpretation ist klar: Mit steigender Fehlerwahrscheinlichkeit nimmt die Kanalkapazität ab und wird 0 für $p = \frac{1}{2}$.

3.3 Blockcodes

Wir stellen uns nun das Hauptproblem. Gegeben der Kanal (A, P, B), wie können wir die Fehler im Kanal (Rauschen) ausgleichen? Wir senden k-Wörter aus A^k als Blöcke und *codieren* sie in Wörter der Länge n:

$$A^k \underset{\text{Codierung}}{\xrightarrow{\varphi}} C \subseteq A^n \quad (n \geq k)\,,$$

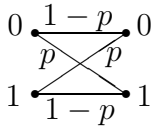
φ ist eine injektive Abbildung von A^k nach A^n. Das Bild $C = \varphi(A^k)$ heißt *Code* und die Wörter $v \in C$ Codewörter. Die Codewörter schicken wir durch den Kanal und decodieren das empfangene Wort in ein Codewort.

Wir haben somit folgendes Schema:

$$\underset{u=u_1\ldots u_k}{A^k} \underset{\text{Codierung}}{\xrightarrow{\varphi}} \underset{v=v_1\ldots v_n}{C \subseteq A^n} \longrightarrow \boxed{\text{Kanal}} \longrightarrow \underset{w=w_1\ldots w_n}{B^n} \underset{\text{Decodierung}}{\xrightarrow{\psi}} \underset{\widehat{v}=\widehat{v}_1\ldots\widehat{v}_n}{C} \xrightarrow{\varphi^{-1}} \underset{\widehat{u}=\widehat{u}_1\ldots\widehat{u}_k}{A^k}$$

φ ist die Codierungsregel, ψ die Decodierungsregel.

Beispiel. Binärer symmetrischer Kanal



Wir wählen den Wiederholungscode und wiederholen jedes Symbol dreimal. Die Decodierungsregel ψ ist die Majoritätsregel.

$$
\begin{array}{lll}
k=1 & 0 \xrightarrow{\varphi} 000 & \left.\begin{array}{l} 000 \\ 001 \\ 010 \\ 100 \end{array}\right\} \xrightarrow{\psi} 000 \longrightarrow 0 \\
n=3 & 1 \xrightarrow{\varphi} 111 & \left.\begin{array}{l} 011 \\ 101 \\ 110 \\ 111 \end{array}\right\} \xrightarrow{\psi} 111 \longrightarrow 1
\end{array}
$$

Angenommen 000 wird gesendet, dann ist die Decodierungsfehlerwahrscheinlichkeit

$$\sum_{w,\psi w \neq 000} p(w|000) = 3p^2(1-p) + p^3 = 3p^2 - 2p^3 < p.$$

Für $p = 0,1$ ergibt dies die Fehlerwahrscheinlichkeit $0,028$. Die Fehlerwahrscheinlichkeit ist klein, allerdings ist die Übertragungsrate $\frac{k}{n} = \frac{1}{3}$ gering.

Diese Ziele divergieren: Je höher die Rate, desto geringer die Sicherheit und umgekehrt.

Wir wollen nun das Modell für Blockcodierung, Rate und Fehlerwahrscheinlichkeit präzisieren. Dazu noch einmal das Schema, mit den einzelnen Symbolen als Zufallsvariable.

$$
\underset{U=(U_1,\ldots,U_k)}{A^k} \xrightarrow{\varphi} \underset{V=(V_1,\ldots,V_n)}{C \subseteq A^n} \longrightarrow \boxed{\text{Kanal}} \longrightarrow \underset{W=(W_1,\ldots,W_n)}{B^n} \xrightarrow{\psi} \underset{\widehat{V}=(\widehat{V}_1,\ldots,\widehat{V}_n)}{C} \longrightarrow \underset{\widehat{U}=(\widehat{U}_1,\ldots,\widehat{U}_k)}{A^k}
$$

1. $H(U_1, \ldots, U_k) = \sum_{i=1}^{k} H(U_i) = k \log q$
   (Quelle ohne Gedächtnis, Gleichverteilung auf A)

2. $H((W_1, \ldots, W_n)|(V_1, \ldots, V_n)) = \sum_{i=1}^{n} H(W_i|V_i)$

(Kanal ohne Gedächtnis, $p(w_1 \dots w_n | v_1 \dots v_n) = \prod_{i=1}^{n} p(w_i|v_i)$)

3. $$\begin{aligned} I(V\|W) &= I((V_1,\dots,V_n)\|(W_1,\dots,W_n)) \\ &= H(W_1,\dots,W_n) - H((W_1,\dots,W_n)|(V_1,\dots,V_n)) \\ &\le \sum_{i=1}^{n} H(W_i) - \sum_{i=1}^{n} H(W_i|V_i) \\ &= \sum_{i=1}^{n} [H(W_i) - H(W_i|V_i)] = \sum_{i=1}^{n} I(V_i\|W_i) \le nK \end{aligned}$$

4. $$\begin{aligned} I(V\|\widehat{V}) &\le I(V\|W,\widehat{V}) \qquad \text{(Übung)} \\ &= H(V) + H(W,\widehat{V}) - H(V,W,\widehat{V}) + (H(W) - H(W)) \\ &\quad + (H(V,W) - H(V,W)) \\ &= I(V\|W) + H(\widehat{V}|W) - H(\widehat{V}|V,W) \\ &= I(V\|W). \end{aligned}$$

(Da das Decodieren ψ deterministisch ist, gilt $H(\widehat{V}|W) = H(\widehat{V}|V,W) = 0$.)

Somit erhalten wir

$$I(V\|\widehat{V}) \le nK.$$

Definition. Die *Rate* R_C des Codes C ist $R_C = \frac{\log|C|}{n}$. Die Rate ist somit die Information von C pro Symbol bei Gleichverteilung der Codewörter.

Heuristische Überlegung: Der Kanal soll die gesamte Eingangsinformation $k \log q$ durch den Kanal transportieren, das heißt $k \log q \le nK$ oder mit $|C| = q^k$

$$R_C = \frac{\log|C|}{n} = \frac{k \log q}{n} \le K.$$

Dies legt nahe, dass die Rate R_C eines Codes mit beliebiger kleiner Fehlerwahrscheinlichkeit die Kapazität K nicht übersteigen kann.

Angenommen wir haben einen Code $C \subseteq A^n$ mit Verteilung $p(v)$, $v \in C$,

gegeben und eine Decodierungsregel ψ.

Definition. Für $v \in C$ ist

$$p_{E,\psi}(v) = \sum_{\substack{w \in B^n \\ \psi w \neq v}} p(w|v)$$

die Fehlerwahrscheinlichkeit.

$$P_{E,\psi} = \sum_{v \in C} p(v) p_{E,\psi}(v)$$

ist die (erwartete) *Block-Fehler-Wahrscheinlichkeit*,

$$\widehat{P}_{E,\psi} = \max_{v \in C} p_{E,\psi}(v) \geq P_{E,\psi}$$

die *maximale Block-Fehler-Wahrscheinlichkeit*.

Beispiel. Für das Beispiel von eben, Wiederholungscode mit $k = 1$, $n = 3$ und ψ Majoritätsdecodierung erhalten wir

$$p_{E,\psi}(000) = p_{E,\psi}(111) = 3p^2 - 2p^3 = P_{E,\psi} = \widehat{P}_{E,\psi}\,.$$

Frage: Wie sollen wir intelligent decodieren?

Beispiel 1. MED-Decodierung (Minimum Error Decoding).

Definiere $\psi_0 : B^n \to C$, $\psi_0 w = v_0$ so dass gilt:

$$p(v_0|w) \geq p(v|w) \quad \text{für alle} \quad v \in C.$$

Falls v_0 nicht eindeutig bestimmt ist, wählt man eines dieser Codewörter.

Beispiel 2. MLD-Decodierung (Maximum Likelihood Decoding).

Definiere $\psi_1 : B^n \longrightarrow C$, $\psi_1 w = v_1$ so dass gilt:

$$p(w|v_1) \geq p(w|v) \quad \text{für alle} \quad v \in C.$$

Falls v_1 nicht eindeutig bestimmt ist, wähle eines dieser Codewörter.

Satz 7. 1. MED *ist die beste Decodierungsregel, das heißt*

$$P_{E,\psi} \geq P_{E,\psi_0} \quad \textit{für alle Decodierungsregeln } \psi.$$

2. *Ist C gleichverteilt, das heißt $p(v) = \frac{1}{|C|}$, so ist* MLD *gleich gut wie* MED $P_{E,\psi_1} = P_{E,\psi_0}$.

Beweis. 1. Es sei ψ eine beliebige Regel, $\psi : B^n \to C$. Wir haben

$$\begin{aligned}
P_{E,\psi} &= \sum_{v\in C} p(v) p_{E,\psi}(v) = \sum_{v\in C} p(v) \sum_{\substack{w\in B^n \\ \psi w \neq v}} p(w|v) \\
&= \sum_{w\in B^n} \sum_{\substack{v\in C \\ v\neq \psi w}} p(v)p(w|v) = \sum_{w\in B^n} \sum_{\substack{v\in C \\ v\neq \psi w}} p(v,w) \\
&= \sum_{w\in B^n} \left(\sum_{\substack{v\in C \\ v\neq \psi w}} p(v|w) \right) p(w) \\
&= \sum_{w\in B^n} (1 - p(\psi w|w))p(w) \\
&\geq \sum_{w\in B^n} (1 - p(\psi_0 w|w))p(w) = P_{E,\psi_0} = P_{E,\mathrm{MED}}\,.
\end{aligned}$$

2. Angenommen, $w \in B^n$ wird empfangen, dann gilt

$$p(v,w) = p(v|w)p(w) = p(w|v)p(v) = \frac{p(w|v)}{|C|} \quad (v \in C)\,,$$

somit

$$p(v|w) = \frac{p(w|v)}{p(w)|C|}\,.$$

Dies bedeutet, dass $p(v|w)$ maximal wird genau dann, wenn $p(w|v)$ maximal wird, also können wir $\psi_1(w) = \psi_0(w)$ wählen. □

Die Minimum Error Decodierung ist im allgemeinen unpraktisch, da die W-keiten $p(v|w)$ schwierig zu berechnen sind. Da wir nach dem ersten Haupt-

satz Gleichverteilung auf A und damit auf dem Code $C \subseteq A^n$ voraussetzen können, ist die Maximum Likelihood Decodierung die naheliegende Methode.

Beispiel. Betrachten wir wieder den q-ären symmetrischen Kanal mit dem Code C und MLD-Codierung. Angenommen $w = w_1 \dots w_n \in A^n$ wird empfangen, dann gilt für $v = v_1 \dots v_n \in C$,

$$p(w|v) = \prod_{i=1}^{n} p(w_i|v_i) = (\frac{p}{q-1})^d (1-p)^{n-d},$$

für $d = \#\{i : w_i \neq v_i\}$.

Behauptung. $(\frac{p}{q-1})^d (1-p)^{n-d}$ ist monoton fallend in d. Wir haben

$$(\frac{p}{q-1})^d (1-p)^{n-d} \geq (\frac{p}{q-1})^{d+1} (1-p)^{n-d-1} \iff (q-1)(1-p) \geq p$$

und dies ist richtig wegen

$$(q-1)(1-p) \underset{q\geq 2}{\geq} 1-p \underset{p\leq \frac{1}{2}}{\geq} p\,.$$

Ergebnis: $p(w|v)$ ist ein Maximum $\iff$ d ist ein Minimum.

Definition. In A^n, $|A| = q$, definieren wir den *Hamming Abstand* $\Delta(u, v)$ zweier Codewörter $u, v \in C$ als

$$\Delta(u, v) = \#\{i : u_i \neq v_i\}.$$

$\Delta(u, v)$ ist eine Metrik auf A^n, erfüllt also die Dreiecksungleichung. Die Menge

$$S_t(u) = \{v \in A^n : \Delta(v, u) \leq t\}$$

heißt die *Kugel* um u vom *Radius* t.
Durch Abzählen der Wörter mit Abstand k erhält man

$$|S_t(u)| = \sum_{k=0}^{t} \binom{n}{k} (q-1)^k \quad (u \in A^n),$$

und insbesondere für $q = 2$

$$|S_t(u)| = \sum_{k=0}^{t} \binom{n}{k}.$$

Zusammenfassung. Für den q-ären symmetrischen Kanal mit Gleichverteilung auf dem Code $C \subseteq A^k$ ist die Maximum Likelihood Decodierung durch die *Hamming Decodierung* gegeben: Wird $w \in A^n$ empfangen, so decodiere zu einem Codewort $\widehat{v} \in A^n$ mit kleinstem Hamming Abstand zu w.

3.4 Der zweite Hauptsatz von Shannon (Hauptsatz der Kanalcodierung)

Wir beweisen den Satz nur für den binären symmetrischen Kanal, er gilt aber sinngemäß allgemein.

Voraussetzungen. $A = B = \{0,1\}$, $P = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$, $p < \frac{1}{2}$, $K = K(p) = 1 - H(p, 1-p) = 1\, p \log p + (1-p)\log(1-p)$. Rate eines Codes $C \subseteq A^n$ ist $R_C = \frac{\log |C|}{n}$. Wir nehmen Gleichverteilung auf C an und schreiben $p_{E,C}(v)$, $P_{E,C}$ für die Fehler-Wahrscheinlichkeiten, wobei wir die Decodierungsregel ψ noch festlegen werden:

$$P_{E,C} = \frac{1}{|C|} \sum_{v \in C} p_{E,C}(v),\ p_{E,C}(v) = \sum_{\substack{w \in A^n \\ \psi w \neq v}} \varphi(w|v)\,,$$

$$\widehat{P}_{E,C} = \max_{v \in C} p_{E,C}(v).$$

$S_t(w)$ ist die Kugel um $w \in A^n$ vom Radius t mit $|S_t(w)| = \sum_{k=0}^{t} \binom{n}{k}$.

Satz 8 (Zweiter Hauptsatz)**. A.** *Es sei $R < K$ und $\varepsilon > 0$. Dann existieren Codes $C_n \subseteq A^n$ mit $R_{C_n} \geq R$ und $\widehat{P}_{E,C_n} < \varepsilon$ für $n \geq n_0(R, \varepsilon)$. Insbesondere existiert eine Folge von Codes C_n mit $\lim\limits_{n\to\infty} \widehat{P}_{E,C_n} = 0$.*
B. *Es sei $R > K$. Dann gibt es keine Folge von Codes $C_n \subseteq A^n$ mit $R_{C_n} \geq R$ und $\lim\limits_{n\to\infty} P_{E,C_n} = 0$.*

Beweis von A. Wir betrachten Codes $C \subseteq A^n$ mit $|C| = 2 \cdot 2^{\lceil Rn \rceil}$, also $R_C = \frac{\lceil Rn \rceil + 1}{n}$. Dann ist $R_C > R$ und $R_C < \frac{Rn+2}{n} = R + \frac{2}{n}$, und wir wählen n

groß genug, dass $R_C < K$ gilt. Zur Abkürzung setzen wir $M = |C| = 2^{\lceil Rn \rceil + 1}$.

Wir betrachten alle Codes $C \subseteq A^n$ mit $|C| = M$ und decodieren alle mit der folgenden Regel ψ. Es werde $w \in A^n$ empfangen:

1) Wenn $S_t(w) \cap C = \{v\}$ mit $v \in C$, so sei $\varphi w = v$ (t wird noch festgelegt).

2) Ansonsten setze $\psi w = v_0$ für ein festes $v_0 \in C$.

Angenommen $v \in C$ wurde gesendet und $w \in A^n$ empfangen. Ein Fehler tritt demnach auf, wenn

1) $\Delta(v, w) > t$
oder
2) $\Delta(v, w) \leq t$ und $\Delta(v', w) \leq t$ für $v \neq v' \in C$.

Diese Fehler wollen wir getrennt behandeln, also

$$p_{E_1,C}(v) = \sum_{\substack{w \in A^n \\ \Delta(w,v) > t}} p(w|v) \quad (v \in C)\,.$$

Für den Fehler zweiter Art schätzen wir ab

$$\sum_{\substack{w \,\in\, A^n \\ \Delta(w,\,v) \,\leq\, t \\ \Delta(w,\,v') \,\leq\, t \\ \text{für ein } v' \,\neq\, v}} p(w|v) \leq \sum_{\substack{v' \in C \\ v' \neq v}} \sum_{\substack{w \in A^n \\ \Delta(w,v') \leq t}} p(w|v) =: p_{E_2,C}(v)\,.$$

Es gilt dann $p_{E,C}(v) \leq p_{E_1,C}(v) + p_{E_2,C}(v)$,

$$P_{E_1,C} = \frac{1}{|C|} \sum_{v \in C} \sum_{\substack{w \in A^n \\ \Delta(w,v) > t}} p(w|v)$$

$$P_{E_2,C} = \frac{1}{|C|} \sum_{\substack{v, v' \in C \\ v \neq v'}} \sum_{\substack{w \in A^n \\ \Delta(w,v') \leq t}} p(w|v)\,.$$

Unser Ziel ist der Nachweis, dass für n groß genug es einen Code $C \subseteq A^n$ gibt, für den $\widehat{P}_{E_1,C}$ und $\widehat{P}_{E_2,C}$ beliebig klein werden.

Dazu betrachten wir *alle* $\binom{2^n}{M}$ Codes C mit $|C| = M$ und verifizieren, dass

die *Mittelwerte* der Fehlergrößen $P_{E_1,C}$, $P_{E_2,C}$ beliebig klein werden. Dann muß es so einen Code geben.

Es seien $E_1 = \frac{1}{\binom{2^n}{M}} \sum_C P_{E_1,C}$, $E_2 = \frac{1}{\binom{2^n}{M}} \sum_C P_{E_2,C}$ diese Mittelwerte. Wir haben

$$\begin{aligned} E_1 =& \frac{1}{\binom{2^n}{M}} \sum_C \frac{1}{M} \sum_{v \in C} \sum_{\substack{w \in A^n \\ \Delta(w,v) > t}} p(w|v) \\ =& \frac{1}{\binom{2^n}{M}} \frac{1}{M} \sum_{v \in A^n} \sum_{C : v \in C} \sum_{\substack{w \in A^n \\ \Delta(w,v) > t}} p(w|v)\,. \end{aligned}$$

Da $\sum\limits_{\substack{w \in A^n \\ \Delta(w,v) > t}} p(w|v)$ nur von p abhängt, also konstant ist für alle $v \in C$ und es $\binom{2^n-1}{M-1}$ Codes C mit $v \in C$ gibt, folgt wegen $\binom{2^n}{M} = \frac{2^n}{M}\binom{2^n-1}{M-1}$,

$$E_1 = \frac{1}{2^n} \sum_{v \in A^n} \sum_{\substack{w \in A^n \\ \Delta(w,v) > t}} p(w|v)\,,$$

also

$$E_1 = \sum_{\substack{w \in A^n \\ \Delta(w,v) > t}} p(w|v) \quad \text{für ein festes } v \in C. \tag{1}$$

Für E_2 erhalten wir

$$\begin{aligned} E_2 &= \frac{1}{\binom{2^n}{M}} \sum_C \frac{1}{M} \sum_{v \neq v' \in C} \sum_{\substack{w \in A^n \\ \Delta(w,v') \leq t}} p(w|v) \\ &= \frac{1}{\binom{2^n}{M}} \frac{1}{M} \sum_{v \neq v' \in A^n} \sum_{C : v \neq v' \in C} \sum_{\substack{w \in A^n \\ \Delta(w,v') \leq t}} p(w|v)\,. \end{aligned}$$

Da die innere Summe nur von v, v' abhängt, aber nicht vom Code C, und es $\binom{2^n-2}{M-2}$ Codes C mit $v \neq v' \in C$ gibt, so erhalten wir wegen $\binom{2^n}{M} =$

$\frac{2^n(2^n-1)}{M(M-1)}\binom{2^n-2}{M-2}$,

$$
\begin{aligned}
E_2 &= \frac{M-1}{2^n(2^n-1)} \sum_{v \neq v' \in A^n} \sum_{\substack{w \in A^n \\ \Delta(w,v') \leq t}} p(w|v) \\
&\leq \frac{M-1}{2^n(2^n-1)} \sum_{v \in A^n} \sum_{w \in A^n} \sum_{\substack{v' \in A^n \\ \Delta(v',w) \leq t}} p(w|v) \\
&= \frac{M-1}{2^n(2^n-1)} \sum_{v \in A^n} \sum_{w \in A^n} |S_t(w)| p(w|v) \\
&= \frac{M-1}{2^n(2^n-1)} \sum_{k=0}^{t} \binom{n}{k} \underbrace{\sum_{v \in A^n} \underbrace{\sum_{w \in A^n} p(w|v)}_{1}}_{2^n}, \\
&= \frac{M-1}{2^n-1} \sum_{k=0}^{t} \binom{n}{k},
\end{aligned}
$$

und daraus mit $\frac{M-1}{2^n-1} \leq \frac{M}{2^n}$

$$
E_2 \leq \frac{M}{2^n} \sum_{k=0}^{t} \binom{n}{k}. \tag{2}
$$

Lemma. *Sei $t \leq \frac{n}{2}$, $\lambda = \frac{n}{t}$, dann gilt*

$$
\sum_{k=0}^{t} \binom{n}{k} \leq 2^{nH(\lambda, 1-\lambda)}.
$$

Beweis. Wir haben

$$
\begin{aligned}
1 = [\lambda + (1-\lambda)]^n &= \sum_{k=0}^{n} \binom{n}{k} \lambda^k (1-\lambda)^{n-k} \geq \sum_{k=0}^{t} \binom{n}{k} \lambda^k (1-\lambda)^{n-k} \\
&= (1-\lambda)^n \sum_{k=0}^{t} \binom{n}{k} (\frac{\lambda}{1-\lambda})^k.
\end{aligned}
$$

Da $(\frac{\lambda}{1-\lambda})^k \geq (\frac{\lambda}{1-\lambda})^{k+1}$ ist wegen $1-\lambda \geq \lambda \iff \lambda \leq \frac{1}{2}$, so können wir abschätzen

$$1 = (1-\lambda)^n \sum_{k=0}^{t} \binom{n}{k} (\frac{\lambda}{1-\lambda})^k \geq (1-\lambda)^n \sum_{k=0}^{t} \binom{n}{k} (\frac{\lambda}{1-\lambda})^t$$

$$= \lambda^t (1-\lambda)^{n-t} \sum_{k=0}^{t} \binom{n}{k},$$

also

$$\sum_{k=0}^{t} \binom{n}{k} \leq \lambda^{-t} (1-\lambda)^{-(n-t)}.$$

Setzen wir $t = \lambda n$ ein, so ergibt dies

$$\sum_{k=0}^{t} \binom{n}{k} \leq \lambda^{-\lambda n} (1-\lambda)^{-(1-\lambda)n} = 2^{\log \lambda^{-\lambda n}} \cdot 2^{\log (1-\lambda)^{-(1-\lambda)n}}$$

$$= 2^{-\lambda n \log \lambda - (1-\lambda) n \log (1-\lambda)}$$

$$= 2^{n(-\lambda \log \lambda - (1-\lambda)\log(1-\lambda))} = 2^{nH(\lambda, 1-\lambda)}. \qquad \square$$

Mit $M = 2^{R_C n}$, $R < R_C < K$ folgt

$$E_2 \leq \frac{2^{R_C n}}{2^n} 2^{nH(\lambda,1-\lambda)} = 2^{n(R_C + H(\lambda,1-\lambda) - 1)}$$

$$= 2^{-n(1-H(\lambda,1-\lambda)-R_C)}.$$

Wenn wir zeigen können, dass $\delta = 1 - H(\lambda, 1-\lambda) - R_C > 0$ ist, so geht $E_2 \to 0$ mit $n \to \infty$.

Jetzt erst wählen wir t. Es sei $\rho > 0$ klein genug, dass $p + \rho < \frac{1}{2}$ ist und $R_C < K(p+\rho) < K(p) = K$ gilt. Wir wählen $t = \lfloor (p+\rho)n \rfloor$, dann gilt

$$\lambda = \frac{t}{n} \leq \frac{(p+\rho)n}{n} = p + \rho,$$

$$H(\lambda, 1-\lambda) \leq H(p+\rho, 1-(p+\rho)),$$

also

$$\begin{aligned} 1 - H(\lambda, 1-\lambda) - R_C &\geq 1 - H(p+\rho, 1-(p+\rho)) - R_C \\ &= K(p+\rho) - R_C > 0. \end{aligned}$$

Mit diesem t geht dann $E_2 \to 0$ für $n \to \infty$.

Nun zu E_1. Nach (1) gilt mit $v = 0$

$$E_1 = \sum_{\substack{w \in A^n \\ \Delta(w,0) > t}} p(w|0), \quad t = \lfloor (p+\rho)n \rfloor .$$

Wir fassen A^n als W-Raum mit W-keiten $p(w|0)$ auf und definieren die Zufallsvariable $Z : A^n \longrightarrow \{0, 1, \ldots, n\}$, $Z(w) = \Delta(w, 0)$. Offenbar ist $E[Z] = np$, $\mathrm{Var}[Z] = np(1-p)$. Wir haben

$$E_1 = p(Z > t)$$

und erhalten mit der Tschebyschev Ungleichung

$$\begin{aligned} p(Z > t) &= p(Z > \lfloor n(p+\rho) \rfloor) = p(Z > n(p+\rho)) \ (Z \text{ ist ganzzahlig}) \\ &= p(Z - np > n\rho) \\ &\leq p(|Z - np| > n\rho) \leq \frac{\mathrm{Var}[Z]}{n^2\rho^2} = \frac{np(1-p)}{n^2\rho^2} \\ &= \frac{p(1-p)}{n\rho^2} . \end{aligned}$$

Also geht $E_1 = p(Z > t) \to 0$ für $n \to \infty$.

Es existiert also ein Code $C'_n \subseteq A^n$ mit $P_{E,C'_n} < \frac{\varepsilon}{2}$ wenn n groß genug ist. In C'_n erfüllen mindestens $\frac{M}{2} = 2^{\lceil Rn \rceil}$ Codewörter v, $p_{E,C'_n}(v) < \varepsilon$. Andernfalls wäre

$$P_{E,C'_n} \geq \frac{1}{M} \frac{M}{2} \varepsilon = \frac{\varepsilon}{2}$$

im Widerspruch zu $P_{E,C'_n} < \frac{\varepsilon}{2}$. $C_n \subseteq C'_n$ bestehe aus $\frac{M}{2}$ solchen Codewörtern, dann gilt für C_n

$$R_{C_n} = \frac{\lceil Rn \rceil}{n} \geq R, \ \widehat{P}_{E,C_n} < \varepsilon.$$

Beweis von B. Der folgende Beweis gilt für alle Kanäle. Es sei $R > K$, und wie üblich

$$v \in C \longrightarrow \boxed{\text{Kanal}} \longrightarrow w \in A^n \longrightarrow \widehat{v} \in C\,.$$

Zur Vorbereitung benötigne wir zwei Hilfssätze.

Lemma 1. *Es gilt*

$$H(X|Y) \leq H(X|Y,Z) + H(Z)\,.$$

Beweis. Wir haben

$$\begin{aligned} H(X|Y) &= H(X,Y) - H(Y) = H(X,Y,Z) - H(Z|X,Y) - H(Y) \\ &\leq H(X,Y,Z) - H(Y) + (H(Y,Z) - H(Y,Z)) \\ &\leq H(X|Y,Z) + H(Z) \quad (\text{wegen } H(Y,Z) \leq H(Y) + H(Z)). \end{aligned}$$ □

Fanos Lemma. *Es sei $C \subseteq A^n$ Code mit $P_E = P_{E,C}$, $M = |C|$, dann gilt*

$$H(V|\widehat{V}) \leq H(P_E, 1-P_E) + P_E \log(M-1).$$

Beweis. Wir definieren die Zufallsvariable Z mit Werten in $\{0,1\}$ durch

$$Z = \begin{cases} 1 & \text{für } V \neq \widehat{V} \text{ mit W-keit } P_E \\ 0 & \text{für } V \neq \widehat{V} \text{ mit W-keit } 1 - P_E\,. \end{cases}$$

Nach Lemma 1 gilt

$$H(V|\widehat{V}) \leq H(V|\widehat{V}, Z) + H(Z)\,.$$

Hier ist $H(Z) = H(P_E, 1-P_E)$ und für $H(V|\widehat{V}, Z)$ haben wir nach Definition der bedingten Entropie

$$\begin{aligned} H(V|\widehat{V}, Z) &= \sum_{\widehat{v},z} p(\widehat{V} = \widehat{v}, Z = z) H(V|\widehat{v}, z) \\ &= \sum_{\widehat{v}} p(\widehat{V} = \widehat{v}, Z = 1) H(V|\widehat{v}, 1) \\ &\quad + \sum_{\widehat{v}} p(\widehat{V} = \widehat{v}, Z = 0) H(V|\widehat{v}, 0)\,. \end{aligned}$$

Aus der Definition von Z folgt $H(V|\widehat{v},0)=0$, da $Z=0$ besagt, dass $V=\widehat{V}$ ist, und ferner $H(V|\widehat{v},1)\leq\log(M-1)$, da $V\neq\widehat{V}$ ist. Somit erhalten wir

$$\begin{aligned}H(V|\widehat{V},Z)&\leq\sum_{\widehat{v}}p(\widehat{V}=\widehat{v},Z=1)\log(M-1)\\&=p(Z=1)\log(M-1)\\&=P_E\log(M-1),\end{aligned}$$

und daraus das Resultat. $\square$

Nun zum eigentlichen Beweis. Es sei $C_n\subseteq A^n$, $R_{C_n}\geq R>K$ mit $\delta=R_{C_n}-K>0$, $|C_n|=2^{R_{C_n}n}$. Da C_n gleichverteilt ist, gilt $H(V)=R_{C_n}n$. Wir haben

$$I(V\|\widehat{V})=H(V)-H(V|\widehat{V})\leq nK$$

also

$$H(V|\widehat{V})\geq nR_{C_n}-nK=n\delta\,, \tag{3}$$

und andererseits nach Fanos Lemma

$$\begin{aligned}H(V|\widehat{V})&\leq H(P_E,1-P_E)+P_E\log M\\&=H(P_E,1-P_E)+P_ER_{C^n}n\,.\end{aligned} \tag{4}$$

Aus (3) und (4) ergibt sich

$$n\delta\leq nP_ER_{C_n}+H(P_E,1-P_E)\leq nP_ER_{C_n}+1\,,$$

also

$$P_E=P_{E,C_n}\geq\frac{n\delta-1}{nR_{C_n}}=\frac{\delta}{R_{C_n}}-\frac{1}{nR_{C_n}},\quad\frac{\delta}{R_C}>0\,.$$

Da $\frac{1}{nR_{C_n}}\to 0$ geht, ist keine Konvergenz $P_{E,C_n}\to 0$ mit $n\to\infty$ möglich, und der Beweis ist erbracht. $\square$

Bemerkung. Es gilt sogar, dass die maximale Fehlerwahrscheinlichkeit gegen 1 strebt.