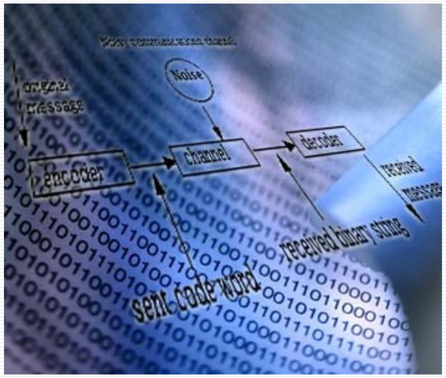
Kapitel 11: Binäre Kanäle



Ziele des Kapitels

ETH

- Binäre Kanäle und ihre Eigenschaften
- Gedächtnisfreie Kanäle
- Codierung als Schätzung und Schätzverfahren
- Kanalkapazität
- Shannon'sches Kanalcodierungstheorem

Grundlegendes

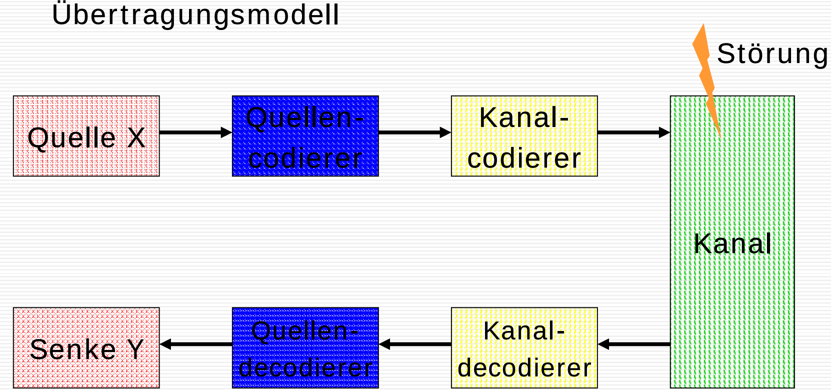
ETH

- Reale Übertragungskanäle sind meist fehlerbehaftet
- „Rohe“ Bits oft als analoge Signale (Spannungspegel) empfangen
- Störungen führen zu Fehlklassifikationen
- Redundante Codierung ermöglicht eine gewisse Fehlertoleranz
- Quellencodierer versuchen, einen fehlerfreien Input optimal zu codieren
- Kanalcodierer bringen gezielt Redundanz (Prüfbits) in den Code ein

Übertragungsmodell

ETH

- Wir betrachten wieder das folgende Übertragungsmodell

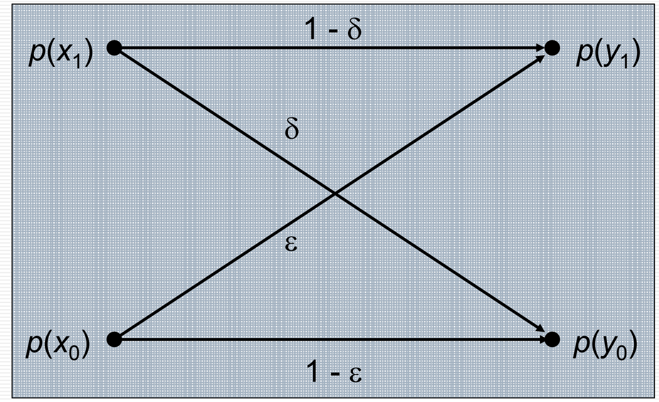


Grundlegendes

- Eine Variante ist, Fehler zu detektieren und das Zeichen nochmals zu übertragen
- Sinnvoll bei kleinen Fehlerwahrscheinlichkeiten
- Anspruchsvollere Variante: Fehler detektieren und beim Empfänger korrigieren
- Dazu müssen Fehlerkorrekturverfahren entwickelt werden
- Sehr grosse praktische Bedeutung (Speichermedien, Netzwerkübertragung etc.)
- Grundlegendes Modell: Allgemeiner Binärer Kanal (BK)

BK Modell

- Der binäre Kanal



Eigenschaften

- Seien $p(x_0)$ und $p(x_1)$ die Wahrscheinlichkeiten für die Symbole $\{0,1\}$ am Kanaleingang
- Ebenso seien $p(y_0)$ und $p(y_1)$ die Wahrscheinlichkeiten am Kanalausgang
- Dann gilt

$$\begin{pmatrix} p(y_0) \\ p(y_1) \end{pmatrix} = \begin{pmatrix} 1-\varepsilon & \delta \\ \varepsilon & 1-\delta \end{pmatrix} \begin{pmatrix} p(x_0) \\ p(x_1) \end{pmatrix}$$

wobei

$$p(y_j \mid x_i) \Leftrightarrow \begin{pmatrix} 1-\varepsilon & \delta \\ \varepsilon & 1-\delta \end{pmatrix}$$

die Matrix der Übergangswahrscheinlichkeiten ist

Eigenschaften

- Definition: Die Transinformation H_T ist die pro Kanalzeichen übertragene Information
- Für den binären Kanal ergibt sie sich wie folgt:

$$H_T = I(X;Y) = H(Y) - H(Y \mid X)$$

mit

$$H(Y) = -\sum_{j \in \{0,1\}} p(y_j) \log_2 p(y_j)$$

und

$$H(Y \mid X) = -\sum_{i \in \{0,1\}} p(x_i) \sum_{j \in \{0,1\}} p(y_j \mid x_i) \log_2 p(y_j \mid x_i)$$

Binärer Kanal

ETH

- Gegeben: $p(x_0)=0.2$, $p(x_1)=0.8$, $\delta=0.1$ und $\varepsilon=0.001$
- Gesucht H_T

$$p(y_j \mid x_i) = \begin{pmatrix} 0.999 & 0.1 \\ 0.001 & 0.9 \end{pmatrix} \qquad \begin{aligned} p(y_0) &= 0.999 \cdot 0.2 + 0.1 \cdot 0.8 = 0.280 \\ p(y_1) &= 0.001 \cdot 0.2 + 0.9 \cdot 0.8 = 0.720 \end{aligned}$$

$$H(Y) = 0.855 \text{ Bit/QZ}$$

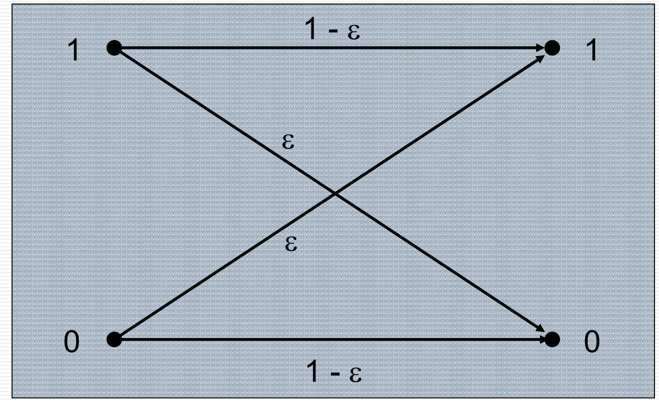
$$\begin{aligned} H(Y \mid X) &= -0.2 \cdot \left(0.999 \cdot \log_2 0.999 + 0.001 \cdot \log_2 0.001\right) \\ &\quad - 0.8 \cdot \left(0.9 \cdot \log_2 0.9 + 0.1 \cdot \log_2 0.1\right) \\ &= 0.377 \text{ Bit/QZ} \end{aligned}$$

$$H_T = H(Y) - H(Y \mid X) = 0.478 \text{ Bit/QZ}$$

Spezialfall 1

ETH

- Der binäre, symmetrische Kanal (BSK)



BSK Modell

ETH

- Jedes Bit wird mit einer Wahrscheinlichkeit ε bei der Übertragung invertiert (verfälscht)
- Bei $\varepsilon=0$ kann 1 bit Information pro Kanalnutzung zuverlässig übertragen werden
- Bei $\varepsilon=0.5$ wird die Ausgabe-Bitfolge gleichverteilt und statistisch unabhängig von der Eingabe
- Es wird keine Information übertragen
- Die Kapazität des Kanals ist 1 bit/Nutzung ($\varepsilon=0$) und 0 bit/Nutzung ($\varepsilon=0.5$)

$$H_T = H(Y) + (1-\varepsilon)\log_2(1-\varepsilon) + \varepsilon \log_2 \varepsilon$$

BSK Modell

ETH

- $\varepsilon=1$ ist gleichwertig zu $\varepsilon=0$
- Für $0< \varepsilon <0.5$ kann eine beliebig zuverlässige Übertragung erreicht werden, wenn jedes Bit genügend oft gesendet wird
- Mehrheitsentscheidung am Kanalausgang notwendig
- Mit zunehmender Redundanz nimmt hierbei jedoch die Übertragungsrate ab
- Durch geschickte Codierung kann die Fehlerwahrscheinlichkeit bei gleichbleibender Rate beliebig verkleinert werden

Shannon's Resultate

ETH

- Shannon zeigte, dass die Übertragungsrate in Grenzen unabhängig von der Übertragungskapazität ist
- Jeder Kanal besitzt eine Kapazität
- Diese ist die maximale Rate, mit der Information zuverlässig übertragbar ist
- Die dazu nötigen Codes können entsprechend komplex werden

Dies ist ein zweites, fundamentales Gesetz von Claude Shannon und in seinem (zweiten) Kanalcodierungstheorem zusammengefasst

Gedächtnisfreie Kanäle

ETH

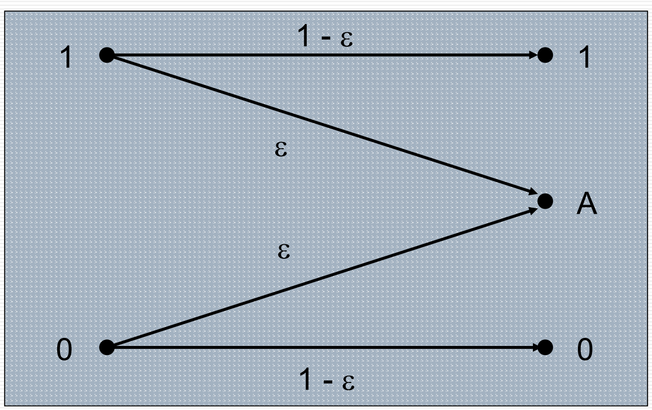
- Generell wird ein Kanal durch die Übergangsmatrix zwischen Eingang und Ausgang definiert
- Eine bedeutende Unterklasse sind sogenannte gedächtnisfreie Kanäle
- Hierbei ist der Output Y_i nur vom aktuellen Input X_i abhängig, nicht von seiner Vorgeschichte $X_{i-1} \dots X_1$
- Definition: Ein diskreter, gedächtnisfreier Kanal (DGK) für ein Inputalphabet χ und ein Outputalphabet γ ist eine bedingte Verteilung

$$P_{Y|X} : \gamma \times \chi \rightarrow R^+$$

Spezialfall 2

ETH

- Der binäre, Auslöschungskanal (BAK)
- Keine Bitinversion, nur Auslöschung



Blockcodes

ETH

- Definition: Ein Blockcode C mit Blocklänge N für einen Kanal mit Inputalphabet χ ist eine Teilmenge $C=\{c_1,\dots,c_M\}$ von χ^N der N-Tupel über χ. Die Rate R von C ist

$$R = \frac{\log_2 M}{N}.$$

- R ist die Anzahl der Bits, die pro Kanalnutzung gesendet werden können

Decodierung als Schätzung

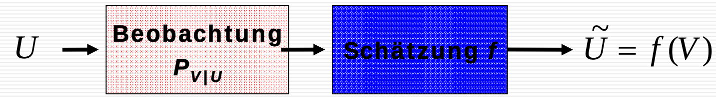
ETH

- Die Decodierung einer fehlerbehafteten Zeichenfolge, gegeben die Symbolfolge am Kanalausgang kann als (Parameter)-Schätzproblem betrachtet werden
- Wir bedienen uns hierzu allgemeiner, statistischer Schätzmethoden
- Es sei *U* dabei eine Zufallsvariable, die aufgrund einer Beobachtung *V* geschätzt werden soll
- Die Wahrscheinlichkeitsverteilung $P_{V|U}$ sei bekannt

Wir erinnern uns an das Informationstheorie-Lemma, welches besagt, dass wir durch Berechnung KEINE Information hinzufügen können.

Bild dazu

ETH



- Der Schätzer ist eine Funktion *f*, welche jedem Wert *v* der Beobachtung den entsprechenden Schätzwert *ũ* zuordnet

Decodierung als Schätzung

ETH

- Die Schätzung ist optimal, wenn die Wahrscheinlichkeit einer korrekten Schätzung $P(U=\tilde{U})$ maximiert wird

$$P(U=\tilde{U}) = \sum_v P(U=\tilde{U}, V=v) \to \max$$

- Wir schreiben

$$P(U=\tilde{U}) = \sum_v P_{UV}(f(v), v) = \sum_v P_{V|U}(v, f(v)) P_U(f(v))$$

- Dieser Ausdruck soll durch Wahl von *f* maximiert werden

Decodierung als Schätzung

ETH

- Fall 1: P_U bekannt (prior bekannt): In diesem Fall muss für jedes *v* dasjenige *ũ* für *f*(*v*) gewählt werden, damit

$$P_{V|U}(v, \tilde{u}) P_U(\tilde{u}) \to \max$$

- Dies wird auch als minimum-error estimation (ME) bezeichnet
- Fall 2: P_U nicht bekannt (uniform prior): Man nimmt an, dass alles Werte von *U* gleichwahrscheinlich sind
- Da $P_U(u)$ für alle *u* gleich ist, muss es bei der Maximierung nicht beachtet werden

Decodierung als Schätzung

ETH

- In diesem Fall muss für jedes v dasjenige $\tilde{u}$ für $f(v)$ gewählt werden, damit

$$P_{V|U}(v,\tilde{u}) \to \max$$

- Dies wird auch als maximum likelihood estimation (ML) bezeichnet

Diese beiden Schätzverfahren sind universell und werden in vielen Anwendungen der Natur- und Ingenieurwissenschaften ausgiebig eingesetzt.
In der Praxis dominiert oft die ML-Methode, da der Prior oft nicht bekannt ist.

Decodierung als Schätzung

ETH

- Wenn das Codewort $c_j=[c_{j1},\ldots,c_{jN}]$ über einen DGK mit Übergangsverteilung $P_{Y|X}$ gesendet wird, so ist der Kanaloutput eine Zufallsvariable $Y^N=[Y_1,\ldots,Y_N]$ mit Wertmenge γ^N und Verteilung

$$P_{Y^N|X^N}(\mathrm{y}^N,\mathrm{c}_j)=\prod_{i=1}^{N} P_{Y|X}(y_i,c_{ji})$$

- Im Decoder muss also für ein empfangenes Kanaloutputwort

$$y^N=[y_1,\ldots,y_N]$$

die beste Schätzung für das gesendete Codewort finden

Decodierung als Schätzung

ETH

- Dieser Schätzvorgang heisst Decodierung
- Mit den Entsprechungen

$$U=X^N \quad V=Y^N$$

erhalten wir die folgenden Theoreme
- Es sei $\tilde{U}$ die Schätzung des Coders

Minimum Error Decoder

ETH

- Ein Decoder, der für ein gegebenes Empfangswort y^N als Schätzung des gesendeten Codewortes eines derjenigen $c_j=[c_{j1},\ldots,c_{jN}]$ wählt, welches

$$P_{Y^N|X^N}(y^N,c_j)P_{X^N}(c_j) \to \max$$

erreicht die minimale Fehlerwahrscheinlichkeit

- Nachteil ist hierbei, dass die Verteilung der Codewörter bekannt sein muss

- In der Praxis ist die Quellenstatistik oft nicht bekannt

Maximum Likelihood Decoder

- Ein Decoder, der für ein gegebenes Empfangswort y^N als Schätzung des gesendeten Codewortes eines derjenigen $c_j=[c_{j1},\dots,c_{jN}]$ wählt, welches

$$P_{Y^N|X^N}(y^N, c_j) \to \max$$

erreicht die minimale Fehlerwahrscheinlichkeit, wenn alle Codewörter gleichwahrscheinlich sind

Kanalkapazität

- Ein DGK ist durch die bedingte Wahrscheinlichkeitsverteilung $P_{Y|X}$ eindeutig beschrieben
- Die Inputverteilung P_X ist jedoch frei
- Wir wählen sie so, dass maximal viel Information übertragen wird
- Definition: Die Kapazität eines durch $P_{Y|X}$ charakterisierten DGK ist das Maximum über Inputverteilungen P_X von $I(X;Y)$

$$C = \max_{P_X} I(X;Y) = \max_{P_X}[H(Y) - H(Y|X)]$$

Kanalkapazität

- Wir werden zeigen, dass die Kapazität eine obere Grenze für die Rate darstellt, mit der Information zuverlässig übertragen werden kann
- Im Allgemeinen sind Kapazitätsberechnungen eher schwierig
- Wir suchen eine Inputverteilung P_X, die $H(Y)$ maximiert und $H(Y|X)$ minimiert


Die Kanalkapazität ist also das Maximum der Transinformation des Kanals

Kanalkapazität

- Die Berechnung vereinfacht sich für folgende Bedingungen
  - A) $H(Y|X=x)$ ist für alle x gleich, also

$$H(Y \mid X = x) = t$$

  - B) Die folgende Summe ist für alle y gleich

$$\sum_X P_{Y|X}(y, x) = s$$

- Letzteres bewirkt, dass bei Gleichverteilung am Kanaleingang auch Gleichverteilung am Kanalausgang vorliegt

Kanalkapazität

ETH

- Theorem: Die Kapazität eines Kanals, welcher die Bedingungen A) und B) erfüllt, ist

$$C = \log|\gamma| - t$$

- Wir betrachten den BSK als Beispiel

$$H_T = H(Y) + (1-\varepsilon)\log_2(1-\varepsilon) + \varepsilon\log_2\varepsilon$$

Kanalkapazität des BSK

ETH

- Übertragungsmatrix:

$$\begin{pmatrix} p(y_0) \\ p(y_1) \end{pmatrix} = \begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix} \begin{pmatrix} p(x_0) \\ p(x_1) \end{pmatrix}$$

- Bedingung A:

$$\begin{aligned} H(Y \mid X = x_0) &= -p(y_0 \mid x_0) \cdot \log_2 p(y_0 \mid x_0) \\ &\quad - p(y_1 \mid x_0) \cdot \log_2 p(y_1 \mid x_0) \\ &= -(1-\varepsilon) \cdot \log_2(1-\varepsilon) - \varepsilon \cdot \log_2 \varepsilon \\ H(Y \mid X = x_1) &= -p(y_0 \mid x_1) \cdot \log_2 p(y_0 \mid x_1) \\ &\quad - p(y_1 \mid x_1) \cdot \log_2 p(y_1 \mid x_1) \\ &= -\varepsilon \cdot \log_2 \varepsilon - (1-\varepsilon) \cdot \log_2(1-\varepsilon) \end{aligned}$$

Kanalkapazität des BSK

ETH

- Bedingung B:

$$\sum_X P_{Y|X}(y,x) \text{ ist gleich für alle y}$$

$$\rightarrow \text{Zeilensumme der Übergangsmatrix}$$

- Berechnung der Kapazität:

$$\begin{aligned} &\text{Es gilt}: p(y_0) = 1 - p(y_1) \\ &\text{Transinformation}: H_T = H(Y) + \varepsilon \cdot \log_2 \varepsilon + (1-\varepsilon) \cdot \log_2(1-\varepsilon) \\ &\text{Kapazität}: C = \max H_T \\ &\text{gemäss Formel } = \log_2|\gamma| - t \text{ mit } t = \varepsilon \cdot \log_2 \varepsilon + (1-\varepsilon) \cdot \log_2(1-\varepsilon) \\ &\rightarrow \text{eingesetzt} \quad = 1 + \varepsilon \cdot \log_2 \varepsilon + (1-\varepsilon) \cdot \log_2(1-\varepsilon) \end{aligned}$$

Kanalkapazität des BSK

ETH

- Inputverteilung:

$$\max H(Y) = \max(p(y_0) \cdot \log_2 p(y_0) + p(y_1) \cdot \log_2 p(y_1))$$

$$\text{erreicht bei } p(y_0) = p(y_1) = \frac{1}{2} \quad \text{(Gleichverteilung)}$$

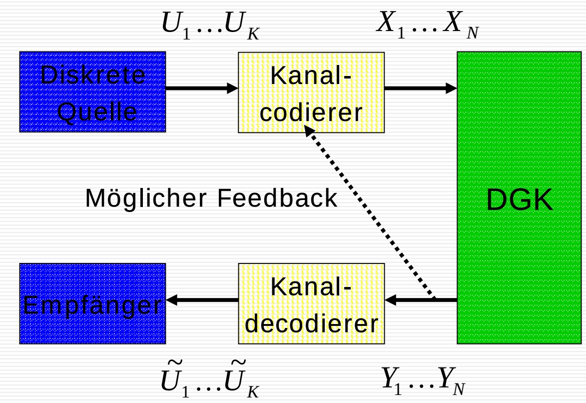
$$\begin{aligned} p(y_0) &= (1-\varepsilon) \cdot p(x_0) + \varepsilon \cdot p(x_1) \\ p(y_1) &= \varepsilon \cdot p(x_0) + (1-\varepsilon) \cdot p(x_1) \end{aligned}$$

$$\rightarrow p(x_0) = p(x_1) = \frac{1}{2}$$

Kapazität und Rate

- Die Kapazität ist eine Obergrenze für die Rate, mit der Information zuverlässig übertragen werden kann
- Je höher die Rate über der Kapazität, umso grösser die Fehlerwahrscheinlichkeit
- Wir betrachten das Modell eines DGK
- Es sollen K Informationsbits $U^K=[U_1,\ldots,U_K]$ durch N Benutzungen übertragen werden
- Die Kapazität sei C
- Der Codierer übersetzt die Informationsbits in ein Codewort $X^N=[X_1,\ldots,X_N]$, welches vom Kanal in $Y^N=[Y_1,\ldots,Y_N]$ verfälscht wird

DGK Modell



Kapazität und Rate

- Die Rate R ist demnach

$$R=\frac{K}{N} \quad \text{bits} \quad \text{pro} \quad \text{Nutzung}$$

- Der Decoder schätzt nun $[\tilde{U}_1,\ldots,\tilde{U}_K]$
- Ein möglicher Feedback kann Information an den Codierer zurückliefern
- Man kann zeigen (Skript), dass

$$H(U^K|\tilde{U}^K)\geq H(U^K)-NC$$

- Die Kanalübertragung kann die Unsicherheit beim Empfänger nicht um mehr als NC reduzieren

Kapazität und Rate

- Damit U^K durch $\tilde{U}^K$ bestimmt ist, muss die Anzahl der Kanalbenutzungen N mindestens sein:

$$N=\frac{H(U^K)}{C}$$