Kapitel 2
Grundzüge der Informationstheorie

Der Beweggrund für dieses Kapitel ist, die Kanalkapazität $\frac{C}{\mathrm{Bit/s}} = \frac{1}{3}\frac{f_g}{\mathrm{Hz}}\frac{SN}{\mathrm{db}}$ (vgl. Seite 36) wahrscheinlichkeitstheoretisch herzuleiten, um sie auf diesem Weg verständlich zu machen. Zu diesem Zweck werden die dafür notwendigen Erkenntnisse der Informationstheorie behandelt. Es wird sich zeigen, dass sich die Kanalkapazität als Ableitung der Shannonschen Formel erklären lässt. Claude Elwood Shannon (amerikanischer Mathematiker und Elektrotechniker, 1916 – 2001) ist der Begründer der Informationstheorie, der seine Arbeit „A mathematical theory of communication" 1948 veröffentlichte.

2.1 Der Informationsgehalt (die Formel von Hartley)

Jede Station eines Multicomputersystems kann sowohl Nachrichten senden als auch empfangen. Treten die Stationen miteinander in Kommunikation, dann betätigt sich im Idealfall eine der Stationen als **Nachrichtenquelle**, während eine andere als Nachrichtensenke arbeitet. Die Nachrichtenquelle wählt aus der Menge ihrer Nachrichten eine aus und sendet sie zur Nachrichtenquelle. Die auf diese Weise von der Nachrichtenquelle immer wieder erzeugten Nachrichten gelangen dann über den **Nachrichtenkanal** (Netzwerk) zum Bestimmungsort.
Im folgenden Bild ist das Modell eines Kommunikationssystems nach Shannon dargestellt.

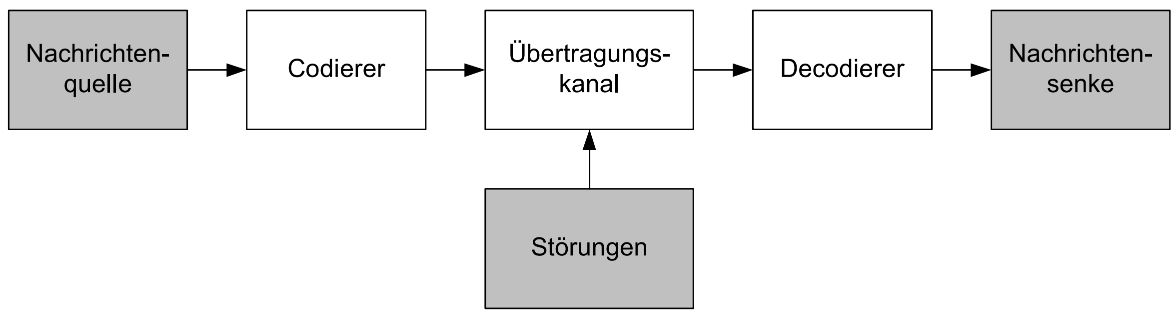

Bild 33: Modell eines Kommunikationssystems

Die übertragenen Nachrichten können beliebige Zeichen sein (z.B. Buchstaben, Zahlen, Symbole) oder diskrete und stetige Zeitfunktionen. Deshalb wird die von der Nachrichtenquelle erzeugte Nachricht von einem Codierer in ein übertragbares Signal umgewandelt. Auf der Empfängerseite wird dann umgekehrt das empfangene Signal decodiert und der Nachrichtensenke zugeführt. Doch auf dem Weg durch den Übertragungskanal kann das Signal (die Nachricht) beschädigt werden. Deshalb ist eine Aussage darüber, ob es sich bei der empfangenen Nachricht auch wirklich um die gesendete Nachricht handelt, nur mit einer gewissen Wahrscheinlichkeit möglich.
Wir nehmen nun an, dass die Nachrichtenquelle einen endlichen Zeichenvorrat $X = \{x_1, x_2, \cdots\cdots, x_n\}$ besitzt. Des weiteren nehmen wir an, dass die Nachrichtenerzeugung darin besteht, dass die Auswahl eines dieser Zeichen ein zufälliger Prozess ist. Mit anderen Worten, die Nachrichtenquelle wählt das zu sendende Zeichen mit der Wahrscheinlichkeit $p_i = P(X = x_i)$; i = 1, 2, ..., n aus.

Somit lässt sich die Nachrichtenquelle X durch das folgende Wahrscheinlichkeitsfeld darstellen.

$$X = \begin{pmatrix} x_1 & x_2 & \cdots\cdots\cdots & x_n \\ P(X = x_1) & P(X = x_2) & & P(X = x_n) \end{pmatrix}$$

Da die Quelle mit Sicherheit eines der Zeichen x_i aus dem Zeichenvorrat von X auswählt, gilt $\sum_{i=1}^{n} P(X = x_i) = 1$. Wir stellen uns nun vor, dass die Nachrichtenquelle im Zeitabstand von Δt immer wieder ein Zeichen auswählt und sendet. Dabei interessiert uns die Frage, ob die Zeitpunkte t_x, zu denen die Wahl getroffen wurde, Einfluss auf die Wahrscheinlichkeit $P(X = x_i)$ hat.

Betrachten wir dazu einen Würfel. Sein Zeichenvorrat lautet $X = \{x_1 = 1, x_2 = 2, \cdots\cdots, x_6 = 6\}$ mit |X| = 6. Werfen wir ihn zum Zeitpunkt t_0 dann liefert der Würfel mit der Wahrscheinlichkeit $p_i = P(X = x_i) = \frac{1}{|X|} = \frac{1}{6}$ eine der Augenzahlen aus $\{1, 2, ..., 6\}$. Mit der gleichen Wahrscheinlichkeit geschieht dies auch zum Zeitpunkt $t_0 + \Delta t$ und zu jedem anderen Zeitpunkt. Es spielt also keine Rolle, zu welchem Zeitpunkt der Würfel sein Ergebnis liefert. Dies bedeutet, die Auswahlwahrscheinlichkeit $p_i = P(X = x_i)$ ist **zeitunabhängig**.

Fassen wir den Würfel als eine Nachrichtenquelle auf, dann erzeugt sie im Zeitabstand Δt eine Serie von Augenzahlen (Zeichen x_i), die alle mit der gleichen Wahrscheinlichkeit $p_i = P(X = x_i) = \frac{1}{|X|} = \frac{1}{6}$ auftreten.

Nachrichtenquellen, bei denen die Auswahlwahrscheinlichkeiten der zu sendenden Zeichen völlig unabhängig davon sind, zu welchen Zeitpunkten die Auswahl getroffen wird, heißen **stationäre** Nachrichtenquellen.

Wir betrachten nun zwei Nachrichtenquellen X_1 und X_2, deren Zeichen x_{1_i} und x_{2_i} Binärzahlen zugeordnet sind. Wir nennen diese Zuordnungen Zeichencodierung.

<u>Quelle 1</u>: $X_1 = \{x_{1_1}, x_{1_2}\}$ mit $|X_1| = 2$ und den Codierungen

x_{1_1} : 0

x_{1_2} : 1

<u>Quelle 2</u>: $X_2 = \{x_{2_1}, x_{2_2}, x_{2_3}, x_{2_4}\}$ mit $|X_2| = 4$ und den Codierungen

x_{2_1} : 00

x_{2_2} : 01

x_{2_3} : 10

x_{2_4} : 11

Jedes Zeichen aus dem Vorrat $X_1 = \{x_{1_1}, x_{1_2}\}$ der Quelle 1 tritt mit der Wahrscheinlichkeit

$p_{1_i} = P(X_1 = x_{1_i}) = \frac{1}{|X_1|} = \frac{1}{2}$; i = 1, 2 in Erscheinung.

und jedes Zeichen aus dem Vorrat $X_2 = \{x_{2_1}, x_{2_2}, x_{2_3}, x_{2_4}\}$ der Quelle 2 mit der

Wahrscheinlichkeit $p_{2_i} = P(X_2 = x_{2_i}) = \frac{1}{|X_2|} = \frac{1}{4}$; i = 1, 2, 3, 4.

Wenden wir uns nun einer allgemeinen Nachrichtenquelle X mit **n** gleichwahrscheinlichen Zeichen zu. Sie hat den Zeichenvorrat $X = \{x_1, x_2, \cdots\cdots, x_n\}$ und das Wahrscheinlichkeitsfeld

$$X = \begin{pmatrix} x_1 & x_2 & \cdots\cdots\cdots & x_n \\ P(X = x_1) & P(X = x_2) & & P(X = x_n) \\ \frac{1}{n} & \frac{1}{n} & & \frac{1}{n} \end{pmatrix}$$

Unser Interesse gehört jetzt dem **Informationsgehalt**, den jedes Zeichen x_i aus diesem Zeichenvorrat besitzt.

Nach Hartley ist der Informationsgehalt $I(X = x_i) = I(x_i)$ eines Zeichens x_i gegeben durch

$$\boxed{I(X = x_i) = I(x_i) = \log_2 \frac{1}{P(X = x_i)} = \log_2 \frac{1}{P(x_i)}}$$

oder wegen $P(x_i) = \frac{1}{n}$ durch

$$\boxed{I(x_i) = \log_2 n}$$

Nach dem Vorschlag von A. Rényi wird diese Formel **Hartleysche Formel** genannt, weil sie Hartley (amerikanischer Elektroingenieuer, 1888 – 1970) als erstes 1928 veröffentlichte. Sie charakterisiert die n **gleichwahrscheinlichen** Zeichen eines Zeichenvorrats und ist eine mathematische Definition der Information. Sie braucht deshalb nicht bewiesen zu werden. Durch Umformung ergibt sich aus

$I(x_i) = \log_2 \frac{1}{P(x_i)}$ die Beziehung $I(x_i) = \log_2 1 - \log_2 P(x_i) = 0 - \log_2 P(x_i) = -\log_2 P(x_i)$.

Somit ist der Informationsgehalt einer homogen Menge mit gleichwahrscheinlichen Zeichen x_i auch gegeben durch

$$\boxed{I(x_i) = -\log_2 P(x_i)}$$

Betrachten wir dazu den Kurvenverlauf der Logarithmusfunktion $y = \log_2 x$ im folgenden Bild 34.

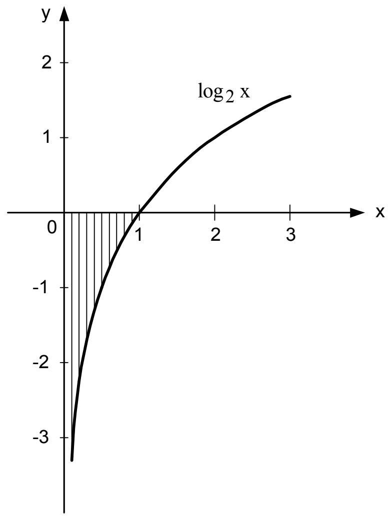

Bild 34: Logarithmusfunktion $y = \log_2 x$

Unser besonderes Interesse gilt dem Definitionsbereich $0 \le x \le 1$ der Funktion $y = \log_2 x$. Dieser ist identisch mit den Wahrscheinlichkeiten $P(x_i)$, die bekanntermaßen im Bereich $0 \le P(x_i) \le 1$ liegen.
Nimmt x die Wahrscheinlichkeiten $P(x_i)$ im Bereich $0 \le P(x_i) < 1$ an, dann sind alle resultierenden Funktionswerte $y = \log_2 P(x_i)$ negativ.
• Für $x = P(x_i) \to 0$ geht $\log_2 P(x_i) \to -\infty$.
Folglich sind im betreffenden Bereich die Funktionswerte von $-\log_2 P(x_i)$ ausschließlich positiv. Es gilt: Für $x = P(x_i) \to 0$ geht $-\log_2 P(x_i) \to +\infty$.

Also ist der Informationsgehalt $I(x_i) = -\log_2 P(x_i)$ niemals negativ, sondern stets **positiv**.

Hierfür schreiben wir: Für $P(x_i) \to 0$ geht $I(x_i) = -\log_2 P(x_i) \to +\infty$

und gewinnen hieraus folgende Erkenntnis:
- *Je kleiner die Wahrscheinlichkeit* $P(x_i)$ *eines Zeichens* x_i *ist, umso größer ist sein Informationsgehalt* $I(x_i)$.

Für $P(x_i) = 1$ ist $I(x_i) = -\log_2 P(x_i) = (-)0$

(siehe Kurvenverlauf). Dies führt zu einer weiteren Erkenntnis:
• *Tritt ein Zeichen* x_i *mit der Wahrscheinlichkeit* $P(x_i) = 1$ *in Erscheinung (sicheres Ereignis), dann ist sein Informationsgehalt* $I(x_i)$ *gleich Null.*

Die Maßeinheit für den Informationsgehalt ist das **bit** (Abkürzung von „binary digit“) mit dem Wertevorrat {0, 1}.

Wir wollen nun die Hartleysche Formel anwenden und den Informationsgehalt der oben dargestellten Informationsquellen Quelle 1 und Quelle 2 berechnen.

Quelle 1: Für ihren Zeichenvorrat $X_1 = \{x_{1_1} : 0, x_{1_2} : 1\}$ mit $|X_1| = n_1 = 2$ erhalten wir den Informationsgehalt $I(x_{1_i}) = -\log_2 P(x_{1_i}) = \log_2 n_1 = \log_2 2$ = **1 bit**.
Das Ergebnis entspricht der im Zeichenvorrat angegebenen Codierung.

Quelle 2: Für ihren Zeichenvorrat $X_2 = \{x_{2_1} : 00, x_{2_2} : 01, x_{2_3} : 10, x_{2_4} : 11\}$ mit $|X_2| = n_2 = 4$ erhalten wir den Informationsgehalt $I(x_{2_i}) = -\log_2 P(x_{2_i}) = \log_2 n_2 = \log_2 4$ = **2 bit**.
Auch dieses Ergebnis entspricht der im Zeichenvorrat angegebenen Codierung.

Hinweis: Sollte die Anzahl n der Zeichen x_i in einem Zeichenvorrat X **keine** Potenz zur Basis 2 sein, dann ist eine einfache Berechnung des Logarithmus (wie in den beiden Beispielen) nicht mehr möglich. In diesem Fall kann dann nach der Umrechnungsregel

$$\boxed{\log_2 n = \frac{1}{\ln 2} \cdot \ln n = 1{,}4427 \cdot \ln n}$$

für jedes beliebige n der Logarithmus zur Basis 2 mit dem Taschenrechner berechnet werden.

Beispiel: Es sei n = 7. Dafür ergibt sich ein Informationsgehalt von

$$\log_2 7 = 1{,}4427 \cdot \ln 7 = 1{,}4427 \cdot 1{,}9458 = 2{,}8074 \text{ bit}$$

2.2 Die Entropie (die Formel von Shannon)

Angenommen, es existiert eine Anzahl m paarweise unvereinbarer Teilquellen $X_1, X_2, ..., X_m$; k = 1, 2, ..., m mit den Zeichenvorräten $X_1 = \{x_{1_1}, x_{1_2}, \ldots\ldots, x_{1_{n_1}}\}$, $X_2 = \{x_{2_1}, x_{2_2}, \ldots\ldots, x_{2_{n_2}}\}$, ..., $X_m = \{x_{m_1}, x_{m_2}, \ldots\ldots, x_{m_{n_m}}\}$ den Wahrscheinlichkeiten $\frac{1}{|X_k|}$ und Informationsgehalten $I(x_{k_i}) = \log_2 |X_k|$. Zudem sei der Zeichenumfang der Teilquellen verschieden: $|X_1| \neq |X_2| \neq \cdots\cdots \neq |X_m|$

$$X_1 = \begin{pmatrix} x_{1_1} & x_{1_2} & & x_{1_{n_1}} \\ \frac{1}{|X_1|} & \frac{1}{|X_1|} & \cdots\cdots & \frac{1}{|X_1|} \\ \log_2 |X_1| & \log_2 |X_1| & & \log_2 |X_1| \end{pmatrix} \qquad X_2 = \begin{pmatrix} x_{2_1} & x_{2_2} & & x_{2_{n_2}} \\ \frac{1}{|X_2|} & \frac{1}{|X_2|} & \cdots\cdots & \frac{1}{|X_2|} \\ \log_2 |X_2| & \log_2 |X_2| & & \log_2 |X_2| \end{pmatrix}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad X_m = \begin{pmatrix} x_{m_1} & x_{m_2} & & x_{m_{n_m}} \\ \frac{1}{|X_m|} & \frac{1}{|X_m|} & \cdots\cdots & \frac{1}{|X_m|} \\ \log_2 |X_m| & \log_2 |X_m| & & \log_2 |X_m| \end{pmatrix}$$

Aus der Menge der Teilquellen wählen wir nun eine beliebige Untermenge aus (jedoch mindestens zwei Teilquellen), nehmen von den ausgewählten Untermengen jeweils eine begrenzte Anzahl von Codierungen willkürlich oder gezielt heraus und „verleiben“ sie der Nachrichtenquelle X ein. Danach stellt X eine inhomogene Menge von Zeichen dar, die beispielhaft wie folgt aussehen könnte:

$$X = \begin{pmatrix} x1_1 & x1_2 & x2_1 & x2_2 & xm_1 & xm_2 \\ \frac{1}{|X_1|} & \frac{1}{|X_1|} & \frac{1}{|X_2|} & \frac{1}{|X_2|} & \frac{1}{|X_m|} & \frac{1}{|X_m|} \\ \log_2|X_1| & \log_2|X_1| & \log_2|X_2| & \log_2|X_2| & \log_2|X_m| & \log_2|X_m| \end{pmatrix}$$

Wir erkennen an den Herkunftsbezeichnungen, dass jeweils die ersten beiden Zeichen von $X_1, X_2, ..., X_m$ ausgesucht wurden. Da sie jetzt zum Zeichenvorrat der Nachrichtenquelle $X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ gehören, benennen wir sie um und so bekommt X folgende Gestalt:

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ \frac{1}{|X_1|} & \frac{1}{|X_1|} & \frac{1}{|X_2|} & \frac{1}{|X_2|} & \frac{1}{|X_m|} & \frac{1}{|X_m|} \\ \log_2|X_1| & \log_2|X_1| & \log_2|X_2| & \log_2|X_2| & \log_2|X_m| & \log_2|X_m| \end{pmatrix}$$

Wie für eine homogene Menge von Zeichen, so gilt auch für eine inhomogene Menge $\sum_i P(X = x_i) = 1$.

Wir interessieren uns jetzt dafür, welchen durchschnittlichen Informationsgehalt die Nachrichtenquelle $X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ hat. In Anlehnung an den Erwartungswert $E(X) = \sum_{i=1}^{n} x_i P(X = x_i)$ einer diskreten Zufallsvariablen X, ist es sinnvoll den Informationsgehalt $I(X = x_i)$ eines Zeichens x_i mit seiner Wahrscheinlichkeit $P(X = x_i)$ zu „gewichten“. So gesehen, betrachten wir dann eigentlich den Erwartungswert des Informationsgehalts einer Nachrichtenquelle X.

Und so ist der **mittlere Informationsgehalt** einer diskreten und stationären Nachrichtenquelle $X = \{x_1, x_2, ..., x_n\}$ mit den Wahrscheinlichkeiten $P(X = x_1), P(X = x_2), ..., P(X = x_n)$ gegeben durch

$$\boxed{H(X) = \sum_{i=1}^{n} P(X = x_i) I(X = x_i) = \sum_{i=1}^{n} P(x_i) I(x_i)}$$

Mit dem Informationsgehalt $I(x_i) = -\log_2 P(x_i) = \log_2 n$ erhalten wir hieraus

$$\boxed{H(X) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i) = \sum_{i=1}^{n} P(x_i) \log_2 \frac{1}{P(x_i)} = \sum_{i=1}^{n} P(x_i) \log_2 n}$$

Die Formel wurde zuerst von Shannon aufgestellt und heißt deswegen **Shannonsche Formel**.

Unabhängig von ihm, hat sie auch Norbert Wiener (amerikanischer Mathematiker, Prof. am MIT, 1894-1964) erkannt. Mit H(X) wird der mittlere Informationsgehalt bzw. die **Entropie** einer Nachrichtenquelle X bezeichnet.
Der Begriff Entropie stammt aus der Thermodynamik, weil dort die mathematische Formulierung der Entropie eine ähnliche Gestalt hat wie die Shannonsche Formel. Doch wie schon oben erwähnt, sie ist eigentlich der Erwartungswert des Informationsgehalts einer Nachrichtenquelle X.

Beispiel 1:
In der folgenden Tabelle ist der inhomogene Zeichenvorrat einer diskreten und stationären Nachrichtenquelle X zusammen mit seinen Wahrscheinlichkeiten $P(x_i)$ und den Informationsgehalten $I(x_i)$ dargestellt:

$X = x_i$	x_1 (00)	x_2 (01)	x_3 (11)	x_4 (100)	x_5 (110)
$P(x_i)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
$I(x_i)$	2 bit	2 bit	2 bit	3 bit	3bit

• Die drei **linken** Zeichen $x_1 \equiv x_{1_1}$ (00), $x_2 \equiv x_{1_2}$ (01) und $x_3 \equiv x_{1_4}$ (11) entstammen der Teilquelle

$$X_1 = \begin{pmatrix} x_{1_1} & x_{1_2} & x_{1_3} & x_{1_4} \\ (00) & (01) & (10) & (11) \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

mit $|X_1| = n_1 = 4$, der Zeichenwahrscheinlichkeit $P(x_i) = \frac{1}{n_1} = \frac{1}{4}$ und dem Informationsgehalt $I(x_i) = \log_2 n_1 = \log_2 4 = 2 \text{ bit}$.

• Die beiden **rechten** Zeichen $x_4 \equiv x_{2_5}$ (100) und $x_5 \equiv x_{2_7}$ (110) entstammen der Teilquelle

$$X_2 = \begin{pmatrix} x_{2_1} & x_{2_2} & x_{2_3} & x_{2_4} & x_{2_5} & x_{2_6} & x_{2_7} & x_{2_8} \\ (000) & (001) & (010) & (011) & (100) & (101) & (110) & (111) \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{pmatrix}$$

mit $|X_2| = n_2 = 8$, der Zeichenwahrscheinlichkeit $P(x_i) = \frac{1}{n_2} = \frac{1}{8}$ und dem Informationsgehalt $I(x_i) = \log_2 n_2 = \log_2 8 = 3 \text{ bit}$.

Für die Nachrichtenquelle X mit |X| = n = 5 erhalten wir nach der Shannonschen Formel

$$H(X) = \sum_{i=1}^{5} P(x_i) \log_2 \frac{1}{P(x_i)}$$

$$H(X)=\frac{1}{4}\log_2 4+\frac{1}{4}\log_2 4+\frac{1}{4}\log_2 4+\frac{1}{8}\log_2 8+\frac{1}{8}\log_2 8=\frac{2}{4}+\frac{2}{4}+\frac{2}{4}+\frac{3}{8}+\frac{3}{8}=\frac{6}{4}+\frac{6}{8}=2{,}25\text{ bit}$$

Das Ergebnis besagt: Die Zeichen x_i der Nachrichtenquelle X haben einen mittleren Informationsgehalt von H(X) = 2,25 bit . Dies bedeutet, durch die in der Tabelle angegebenen Codierungen beansprucht die Übertragung der Zeichen im Mittel 2,25 bit.

Beispiel 2:
In der folgenden Tabelle ist der homogene Zeichenvorrat einer diskreten und stationären Nachrichtenquelle X dargestellt. D.h. alle Zeichen x_i sind gleichwahrscheinlich und haben demzufolge einen identischen Informationsgehalt $I(x_i)$:

$X = x_i$	x_1 (00)	x_2 (01)	x_3 (10)	x_4 (11)
$P(x_i)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$I(x_i)$	2 bit	2 bit	2 bit	2bit

Es gilt $|X| = n = 4$, $P(x_i)=\frac{1}{n}=\frac{1}{4}$ und $I(x_i)=\log_2 n=\log_2 4=2\text{ bit}$.

Nach der Shannonschen Formel $H(X)=\sum_{i=1}^{n}P(x_i)\log_2\frac{1}{P(x_i)}=\sum_{i=1}^{n}P(x_i)\log_2 n$ bekommen wir

$$H(X)=\frac{1}{n}\log_2 n+\frac{1}{n}\log_2 n+\frac{1}{n}\log_2 n+\frac{1}{n}\log_2 n=\frac{4}{n}\log_2 n=\frac{n}{n}\log_2 n=\log_2 n\,.$$

Das Ergebnis $\boxed{H(X)=\log_2 n}$ lässt erkennen, dass sich wegen

$P(x_1)=P(x_2)=\ldots\ldots P(x_n)=\frac{1}{n}$ die Shannonsche Formel auf die Harleysche Formel reduziert.
Mit anderen Worten, für gleichwahrscheinliche Zeichen x_i stimmt der definierte mittlere Informationsgehalt H(X) einer Nachrichtenquelle X mit dem definierten Informationsgehalt $I(x_i)$ überein. Es ist der **maximale** Informationsgehalt, den eine Nachrichtenquelle X haben kann. Für unser Beispiel erhalten wir $I(x_i)=H(X)=\log_2 4=2\text{ bit}$.

Die Beziehung $H(X)=\log_2 n$ wird auch als **Entscheidungsgehalt** H_0 bezeichnet. Es gilt

$$\boxed{H_0=\log_2 n}$$

Woher rührt dieser Begriff? Zur Erklärung zeichnen wir von einer beliebigen homogenen Nachrichtenquelle X z.B. $X=\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ (00) & (01) & (10) & (11)\end{pmatrix}$ einen sogenannten Codebaum mit folgendem Aussehen:

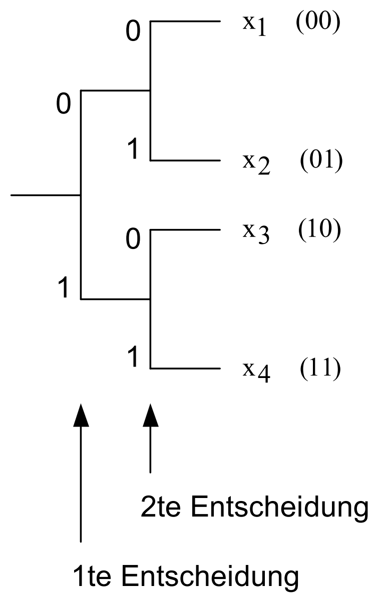

Bild 35: Codebaum = Entscheidungsbaum

Um herauszufinden, ob die Quelle aus dem Zeichenvorat zufällig z.B. das Zeichen $\begin{pmatrix} x_3 \\ (10) \end{pmatrix}$ ausgewählt und gesendet hat, entscheiden wir uns für eine Fragestrategie und achten auf die Antworten:

1te Entscheidung: Ist das erste Zeichen (links beginnend) eine 0? Die Antwort lautet **nein**, folglich muss das 1te Zeichen eine **1** sein.
2te Entscheidung: Ist das zweite Zeichen eine 0? Die Antwort lautet **ja**.

Die Entscheidung für diese 2 Fragen hat das gesendete Zeichen ausfindig gemacht.
Die folgende Frageentscheidung sollte zum gleichen Ergebnis führen:

1te Entscheidung: Ist das erste Zeichen eine 0? Die Antwort lautet **nein**, folglich muss das 1te Zeichen eine **1** sein.
2te Entscheidung: Ist das zweite Zeichen eine 1? Die Antwort lautet **nein**, folglich muss das 2te Zeichen eine **0** sein.

Eine dritte Alternative lautet:

1te Entscheidung: Ist das erste Zeichen eine 1? Die Antwort lautet **ja**.
2te Entscheidung: Ist das zweite Zeichen eine 0? Die Antwort lautet **ja**.

Und eine letzte Alternative lautet:

1te Entscheidung: Ist das erste Zeichen eine 1? Die Antwort lautet **ja**.
2te Entscheidung: Ist das zweite Zeichen eine 1? Die Antwort lautet **nein**, folglich muss das 2te Zeichen eine **0** sein.

Mit jeder der gewählten Fragestellungen gewinnen wir Klarheit und finden das gesuchte Zeichen. Alle führen mit 2 Fragentscheidungen zum gleichen Ergebnis. Sie können natürlich auch auf die verbleibenden Zeichen angewendet werden.

So zeigt sich, dass die **Anzahl** der Entscheidungen (in unserem Beispiel **2**) mit dem mittleren Informationsgehalt $H(X) = \log_2 n = \log_2 4 = 2\ \mathrm{bit}$ einer homogenen Informationsquelle X identisch ist. Demzufolge ist es gerechtfertigt, die Anzahl der Entscheidungen (wir nennen sie Entscheidungsgehalt) durch die mathematischen Formulierung $H_0 = \log_2 n$ auszudrücken.

2.3 Die Entropie unabhängiger Verbundzeichen

Besteht eine Nachrichtenquelle aus mehreren Teilquellen, dann spricht man von einer sogenannten **Verbundquelle**. Zur Erklärung gehen wir davon aus, dass die Verbundquelle aus zwei Teilquellen X und Y besteht.

- Es sei X eine diskrete Nachrichtenquelle mit dem homogenen oder inhomogenen Zeichenvorrat $X = \{x_1, x_2, \ldots\ldots, x_n\}$ und dem Wahrscheinlichkeitsfeld

$$X = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ P(x_1) & P(x_2) & & P(x_n) \end{pmatrix} \text{ und}$$

- es sei Y eine diskrete Nachrichtenquelle mit dem homogenen oder inhomogenen Zeichenvorrat $Y = \{y_1, y_2, \ldots\ldots, y_m\}$ und dem Wahrscheinlichkeitsfeld

$$Y = \begin{pmatrix} y_1 & y_2 & \cdots & y_m \\ P(y_1) & P(y_2) & & P(y_m) \end{pmatrix}.$$

Die mittleren Informationsgehalte (Entropien) der beiden Nachrichtenquellen X und Y lauten

$$H(X) = -\sum_{i=1}^{n} P(X = x_i)\log_2 P(X = x_i) \quad \text{und} \quad H(Y) = -\sum_{j=1}^{m} P(Y = y_j)\log_2 P(Y = y_j).$$

Es sei angenommen, dass X und Y voneinander **unabhängig** sind. Wählt die Nachrichtenquelle X ein Zeichen x_i und zugleich die Nachrichtenquelle Y ein Zeichen y_j aus, dann treten die beteiligten Zeichen als Paar mit der Wahrscheinlichkeit $P(X = x_i, Y = y_j)$ in Erscheinung. Wie für zwei unabhängige Zufallsvariablen, so gilt auch für zwei unabhängige Nachrichtenquellen X und Y der *Multiplikationssatz für unabhängige Ereignisse*:

$$\boxed{P(X = x_i, Y = y_j) = P(X = x_i)\cdot P(Y = y_j) = P(x_i)\cdot P(y_j)} \quad \text{(vgl. Seite 142)}$$

Weil $n\cdot m$-Kombinationen der Teilquellenzeichen (x_i, y_j) möglich sind, resultiert aus den Wahrscheinlichkeitsfeldern der beiden Teilquellen X und Y das Wahrscheinlichkeitsfeld der Verbundquelle X,Y in folgender Gestalt:

$$X, Y = \begin{pmatrix} x_1, y_1 & x_1, y_2 & \cdots & x_n, y_m \\ P(x_1)\cdot P(y_1) & P(x_1)\cdot P(y_2) & & P(x_n)\cdot P(y_m) \end{pmatrix}$$

Auf welche Weise sich die beiden Nachrichtenquellen X und Y zu einer Verbundquelle X,Y formen, illustriert das folgende Bild:

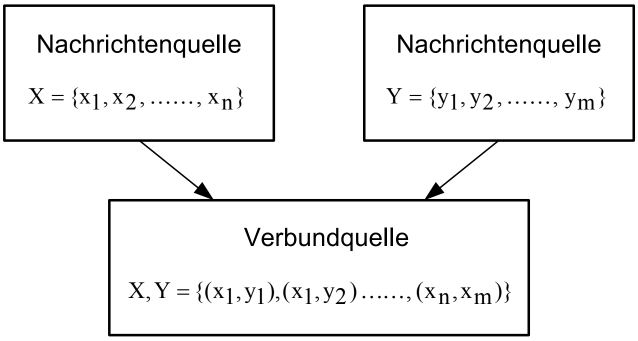

Bild 36: Aus zwei unabhängigen Teilquellen resultierende Verbundquelle

Beispiel:
Wir betrachten die beiden homogenen Nachrichtenquellen $X = \{x_1, x_2\}$ mit den Codierungen $x_1 : (0), x_2 : (1)$ und $Y = \{y_1, y_2\}$ mit den Codierungen $y_1 : (0), y_2 : (1)$. X und Y werden durch die folgenden Wahrscheinlichkeitsfelder beschrieben:

$$X = \begin{pmatrix} x_1 & x_2 \\ (0) & (1) \\ P(x_1) = \frac{1}{2} & P(x_2) = \frac{1}{2} \end{pmatrix} \qquad Y = \begin{pmatrix} y_1 & y_2 \\ (0) & (1) \\ P(y_1) = \frac{1}{2} & P(y_2) = \frac{1}{2} \end{pmatrix}$$

Unser Interesse gilt zunächst den mittleren Informationsgehalten H(X) und HY) der beiden Teilquellen.

• Für die Nachrichtenquelle X lautet er $H(X) = -\sum_{i=1}^{2} P(x_i)\log_2 P(x_i) = \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 1 \ = \mathbf{1}$ bit

und

• Für die Nachrichtenquelle Y lautet er $H(Y) = -\sum_{j=1}^{2} P(y_j)\log_2 P(y_j) = \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 1 \ = \mathbf{1}$ bit.

Als nächstes gilt unser Interesse dem mittleren Informationsgehalt H(X,Y) der Verbundquelle Da X und Y die gemeinsame Verteilung $(x_i, y_j, P(x_i)\cdot P(y_j));\ i, j = 1{,}2$ haben, können wir H(X,Y) aus dem folgenden zweidimensionalen Wahrscheinlichkeitsraum der Verbundquelle gewinnen.

X \ Y	$Y = y_1$	$Y = y_2$
$X = x_1$	$P(x_1)\cdot P(y_1)$	$P(x_1)\cdot P(y_2)$
$X = x_2$	$P(x_2)\cdot P(y_1)$	$P(x_2)\cdot P(y_2)$

=

X \ Y	$Y = y_1$	$Y = y_2$	
$X = x_1$	$\frac{1}{4}$	$\frac{1}{4}$	$P(x_1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$
$X = x_2$	$\frac{1}{4}$	$\frac{1}{4}$	$P(x_2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$
	$P(y_1)$ $= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$	$P(y_2)$ $= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$	$\sum_{j=1}^{2} P(y_j) = \sum_{i=1}^{2} P(x_i) = 1$

Es sei angemerkt, dass die Wahrscheinlichkeitsmatrix einer Verbundquelle die gleichen Eigenschaften besitzt wie die Wahrscheinlichkeitsmatrix von Zufallsvariablen. So ist für ein festgehaltenes i (Zeile i) die **Zeilensumme** gegeben durch

$$\boxed{P(x_i) = \sum_{j=1}^{m} P(X = x_i, Y = y_j) = \sum_{j=1}^{m} P(x_i) \cdot P(y_j)}$$ (vgl. Seite 139)

Demnach erhalten wir für das Zeichen x_1 in der Zeile i = 1 die Auswahlwahrscheinlichkeit $P(x_1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ und für das Zeichen x_2 in der Zeile i = 2 die Auswahlwahrscheinlichkeit $P(x_2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.

Beide Wahrscheinlichkeiten sind in der obigen Matrix eingetragen und stimmen mit den Wahrscheinlichkeiten $P(x_i);\ i = 1{,}2$ der Nachrichtenquelle X überein.

Des weiteren ist für ein festgehaltenes j (Spalte j) die **Spaltensumme** gegeben durch

$$\boxed{P(y_j) = \sum_{i=1}^{n} P(X = x_i, Y = y_j) = \sum_{i=1}^{n} P(x_i) \cdot P(y_j)}$$ (vgl. Seite 140)

So erhalten wir für das Zeichen y_1 in der Spalte j = 1 die Auswahlwahrscheinlichkeit $P(y_1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ und für das Zeichen y_2 in der Spalte j = 2 die Auswahlwahrscheinlichkeit $P(y_2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.

Auch diese beiden Wahrscheinlichkeiten sind in der Matrix der Verbundquelle X,Y eingetragen und stimmen mit den Wahrscheinlichkeiten $P(y_j);\ j = 1{,}2$ der Nachrichtenquelle Y überein. Wie für die zweidimensionale Zufallsvariable Z = X,Y, so ist auch für die Verbundquelle X,Y die Summe der Zeilensummen gegeben durch $\sum_{i=1}^{n=2} P(x_i) = 1$ und die Summe der Spaltensummen durch $\sum_{j=1}^{m=2} P(y_j) = 1$.

Des weiteren ist erkennbar, dass sich die Entropien H(X) und H(Y) der beiden Teilquellen aus der Matrix der Verbundquelle X,Y gewinnen lassen. Die Matrix ist hier nochmals dargestellt:

X \ Y	$Y = y_1$	$Y = y_2$	
$X = x_1$	$\frac{1}{4}$	$\frac{1}{4}$	$P(x_1) = \frac{1}{2}$
$X = x_2$	$\frac{1}{4}$	$\frac{1}{4}$	$P(x_2) = \frac{1}{2}$
	$P(y_1) = \frac{1}{2}$	$P(y_2) = \frac{1}{2}$	

$$H(X) = -\sum_{i=1}^{2} P(x_i)\log_2 P(x_i) = \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 1 = 1 \text{ bit}$$

$$H(Y) = -\sum_{j=1}^{2} P(y_j)\log_2 P(y_j) = \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 1 = 1 \text{ bit}$$

Wie zu sehen ist, liefert analog zu den zweidimensionalen Zufallsvariablen Z=X,Y die Randverteilung von X auf der rechten Seite der Matrix die Entropie H(X) und die Randverteilung von Y auf der unteren Seite liefert die Entropie H(Y).
Wenden wir uns nun der Verbund-Entropie H(X,Y) zu. Zu ihrer Ermittlung gehen wir folgendermaßen vor:
In Anlehnung an den Informationsgehalt $I(x_i) = -\log_2 P(x_i)$ für ein Zeichen x_i erhalten wir für ein Verbundzeichen (x_i, y_j) den Informationsgehalt $I(x_i, y_j) = -\log_2 P(x_i, y_j)$. Wir „gewichten" die Informationsgehalte $I(x_i, y_j)$ mit ihren Verbundwahrscheinlichkeiten $P(x_i, y_j)$ und erhalten die Elemente $I(x_i, y_j) = -P(x_i, y_j)\log_2 P(x_i, y_j)$.

Die Summation aller Elemente führt schließlich zum **mittleren Informationsgehalt H(X,Y)** der Verbundquelle X,Y.

$$\boxed{H(X, Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i, y_j)\log_2 P(x_i, y_j)}$$

oder

$$\boxed{H(X, Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i)\cdot P(y_j)\log_2 P(x_i)\cdot P(y_j)}$$

Durch Anwendung des Logarithmengesetzes $\log_2 P(x_i)\cdot P(y_j) = \log_2 P(x_i) + \log_2 P(y_j)$ resultiert hieraus

$$H(X, Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i)\cdot P(y_j)[\log_2 P(x_i) + \log_2 P(y_j)]$$

$$= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i)\cdot P(y_j)\log_2 P(x_i) - \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i)\cdot P(y_j)\log_2 P(y_j)$$

$$= -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i)\sum_{j=1}^{m} P(y_j) \quad - \quad \sum_{i=1}^{n} P(x_i)\sum_{j=1}^{m} P(y_j)\log_2 P(y_j)\,.$$

Und mit der Summe der Spaltensummen $\sum_{j=1}^{m} P(y_j) = 1$ und der Summe der Zeilensummen $\sum_{i=1}^{n} P(x_i) = 1$ erhalten wir schließlich

$$\boxed{H(X,Y) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i) \quad - \quad \sum_{j=1}^{m} P(y_j)\log_2 P(y_j) \quad = \quad H(X) + H(Y)}$$

<u>In Worten</u>: Sind X und Y zwei unabhängige Nachrichtenquellen mit existierenden Entropien H(X) und H(Y), dann lautet der mittlere Informationsgehalt der Verbundquelle X,Y

$$\boxed{H(X,Y) = H(X) + H(Y)}$$

In unserem Beispiel hat die Verbundquelle X,Y wegen H(X) = 1 bit und H(Y) = 1 bit die Verbund-Entropie H(X,Y) = 1 bit + 1 bit = 2 bit.

Die Gleichung H(X,Y) = H(X) + H(Y) lässt sich nach dem Prinzip der vollständigen Induktion auf beliebig viele Nachrichtenquellen übertragen. Demzufolge erhalten wir für die Teilquellen $X_1, X_2, \ldots\ldots, X_k$ die Verbund-Entropie

$$\boxed{H(X_1, X_2, \ldots, X_k) = H(X_1) + H(X_2) + \cdots\cdots + H(X_k)}$$

<u>In Worten</u>: Sind $X_1, X_2, \ldots\ldots X_k$ beliebig viele unabhängige Nachrichtenquellen mit existierenden Entropien $H(X_1)$, $H(X_2)$, ..., $H(X_k)$, dann ergibt sich die Verbund-Entropie $H(X_1, X_2, \ldots, X_k)$ als Summe der einzelnen Entropien $H(X_i)$ mit i = 1,2, ..., k.

2.4 Die Entropie abhängiger Verbundzeichen

Es sei X eine Nachrichten**quelle**, die Zeichen aus ihrem Zeichenvorrat $X = \{x_1 : 0, x_2 : 1\}$ über einen Binärkanal an eine Nachrichten**senke** Y schickt. Ein Binärkanal ist dadurch gekennzeichnet, dass er nur zwei verschiedene Zeichen (0 oder 1) übertragen kann.
Durch Störungen auf dem Übertragungskanal werden Teile der von X ausgesendeten Zeichen x_1 und x_2 beschädigt und kommen falsch bei Y an. Siehe Bild 37.

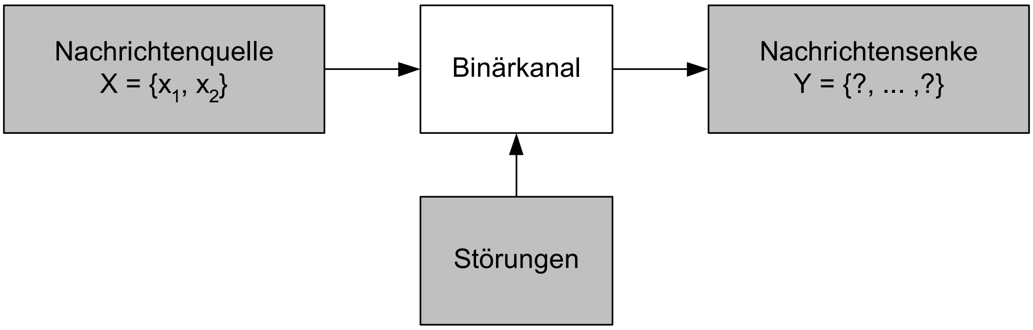

Bild 37: Gestörter Binärkanal

Ein Binärkanal, der die beiden Zeichen mit gleicher Wahrscheinlichkeit verfälscht, heißt *symmetrisch* gestörter Binärkanal. Verfälscht er die Zeichen mit verschiedenen Wahrscheinlichkeiten, wird er *nicht symmetrisch* gestörter Binärkanal genannt.

Es sei bekannt, dass 90% der gesendeten **Nullen** beim Empfänger richtig ankommen. Dafür schreiben wir:

- Auf $x_1 = 0$ folgt $y_1 = 0$
  mit der bedingten Wahrscheinlichkeit $P(Y = y_1 / X = x_1) = P(Y = 0 / X = 0) = 0{,}9$.

Wenn 90% der gesendeten Nullen richtig ankommen, dann werden 10% der gesendeten Nullen falsch empfangen. Dafür schreiben wir:

- Auf $x_1 = 0$ folgt $y_2 = 1$
  mit der bedingten Wahrscheinlichkeit $P(Y = y_2 / X = x_1) = P(Y = 1 / X = 0) = 0{,}1$.

Es sei des weiteren bekannt, dass 85% der gesendeten **Einsen** beim Empfänger korrekt ankommen. Dafür schreiben wir:

- Auf $x_2 = 1$ folgt $y_2 = 1$
  mit der bedingten Wahrscheinlichkeit $P(Y = y_2 / X = x_2) = P(Y = 1 / X = 1) = 0{,}85$.

Wenn 85% der gesendeten Einsen korrekt ankommen, dann kommen 15% der gesendeten Einsen falsch an. Dafür schreiben wir:

- Auf $x_2 = 1$ folgt $y_1 = 0$
  mit der bedingten Wahrscheinlichkeit $P(Y = y_1 / X = x_2) = P(Y = 0 / X = 1) = 0{,}15$.

Zur besseren Übersicht fassen wir die gewonnenen bedingten Wahrscheinlichkeiten $P(y_j/x_i)$ in der Übergangsmatrix Ü = $[P(y_j/x_i)]$ zusammen. Weil diese Matrix das Störverhalten bzw. das Rauschverhalten des Übertragungskanals beschreibt, wird sie oft auch als **Rauschmatrix** bezeichnet.

$$Ü = [P(y_j/x_i)] = \begin{pmatrix} P(Y = y_1/X = x_1) & P(Y = y_2/X = x_1) \\ P(Y = y_1/X = x_2) & P(Y = y_2/X = x_2) \end{pmatrix} = \begin{pmatrix} P(Y = 0/X = 0) & P(Y = 1/X = 0) \\ P(Y = 0/X = 1) & P(Y = 1/X = 1) \end{pmatrix}$$

oder kurz

$$Ü = [P(y_j/x_i)] = \begin{pmatrix} P(y_1/x_1) & P(y_2/x_1) \\ P(y_1/x_2) & P(y_2/x_2) \end{pmatrix} = \begin{pmatrix} 0{,}9 & 0{,}1 \\ 0{,}15 & 0{,}85 \end{pmatrix}$$

Eine Veranschaulichung der Verhältnisse im binären Übertragungskanal zeigt der folgende Übergangsgraph:

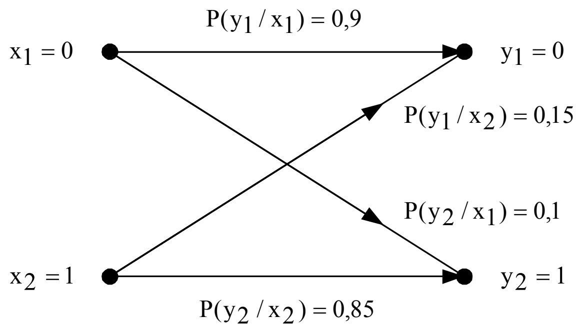

Bild 38: Nicht symmetrisch gestörter Binärkanal

Wie lassen sich die Werte der bedingten Wahrscheinlichkeiten $P(y_j/x_i)$ ermitteln?

Näherungsweise ist dies folgendermaßen möglich:
In einem Experiment lässt man die Nachrichtenquelle X eine sehr große Anzahl **w** das Zeichen $x_1 = 0$ senden. Man beobachtet die Nachrichtensenke Y und stellt fest, dass **h_1-mal** das Zeichen $y_1 = 0$ und **h_2-mal** das Zeichen $y_2 = 1$ empfangen wurde.
Die Gesamtzahl aller Versuche ist gegeben durch $w = h_1 + h_2$. So folgt aus h_1 und w die relative Häufigkeit $r_1 = \frac{h_1}{w}$ für den Empfang der Zeichen $y_1 = 0$ unter der Voraussetzung, dass $x_1 = 0$ gesendet wurde. Dafür schreiben wir

$$\boxed{P(y_1/x_1) = P(Y = 0/X = 0) \approx \frac{h_1}{w}}$$

Aus h_2 und w folgt die relative Häufigkeit $r_2 = \frac{h_2}{w}$ für den Empfang des Zeichens $y_2 = 1$ unter der Voraussetzung, dass $x_1 = 0$ gesendet wurde. Dafür schreiben wir

$$\boxed{P(y_2/x_1) = P(Y = 1/X = 0) \approx \frac{h_2}{w}}$$

In analoger Weise lassen sich auch die restlichen Wahrscheinlichkeiten $P(y_1/x_2)$ und $P(y_2/x_2)$ näherungsweise ermitteln.
Sobald die Nachrichtenquelle X ein Zeichen x_i sendet, empfängt die Nachrichtensenke aufgrund von Störungen das Zeichen y_j/x_i. Die beteiligten Zeichen treten als **Paar** mit der Wahrscheinlichkeit $P(X=x_i, Y=y_j/x_i)$ auf. Da X und Y voneinander **abhängig** sind, gilt für die Elemente der Verbundquelle X,Y der *Multiplikationssatz für abhängige Ereignisse*:

$$\boxed{P(X=x_i, Y=y_j/x_i) = P(X=x_i)\cdot P(Y=y_j/x_i) = P(x_i)\cdot P(y_j/x_i)}$$

Aus den Wahrscheinlichkeitsfeldern der Nachrichtenquelle X und der Nachrichtensenke Y resultiert das Wahrscheinlichkeitsfeld der Verbundquelle X,Y in folgender Gestalt:

$$X,Y = \begin{pmatrix} x_1,y_1/x_1 & x_1,y_2/x_1 & x_2,y_1/x_2 & x_2,y_2/x_2 \\ P(x_1)P(y_1/x_1) & P(x_1)P(y_2/x_1) & P(x_2)P(y_1/x_2) & P(x_2)P(y_2/x_2) \end{pmatrix}$$

Eine Veranschaulichung hierfür zeigt das folgende Bild 39:

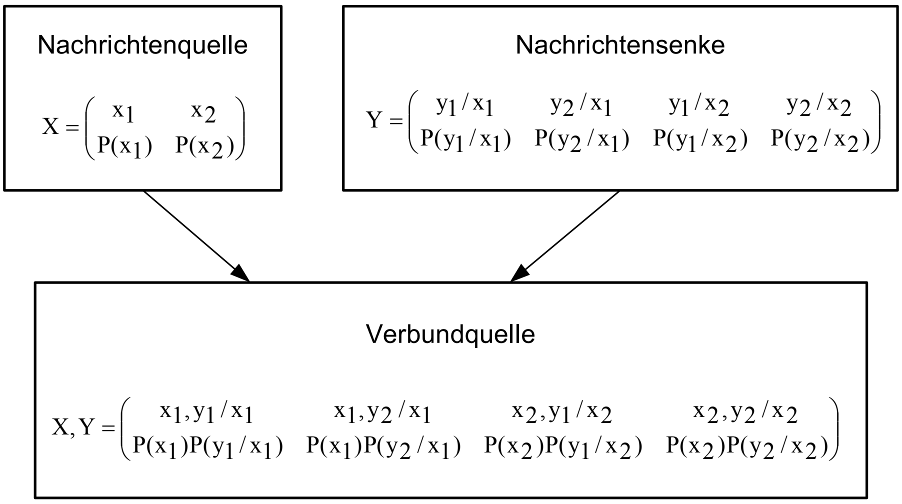

Bild 39: Aus zwei abhängigen Teilquellen resultierende Verbundquelle

Für einen Beobachter, der von außen sowohl die Nachrichtenquelle X als auch die Nachrichtensenke Y im Auge hat, stellt sich Y wie eine Nachrichten**quelle** dar, die eine Folge von Zeichen $Y=\{y_i\},\ i=1,2,......$ produziert.
Wir stellen nun das Wahrscheinlichkeitsfeld der Verbundquelle X,Y als zweidimensionale Wahrscheinlichkeitsmatrix dar:

Y \ X	Y = y_1 (0)	Y = y_2 (1)	
X = x_1 (0)	$P(x_1)\cdot P(y_1/x_1)$	$P(x_1)\cdot P(y_2/x_1)$	$P(x_1)$
X = x_2 (1)	$P(x_2)\cdot P(y_1/x_2)$	$P(x_2)\cdot P(y_2/x_2)$	$P(x_2)$
	$P(y_1)$	$P(y_2)$	

X,Y= ... =

Y \ X	Y = y_1 (0)	Y = y_2 (1)	
X = x_1 (0)	$\frac{1}{2}\cdot 0{,}9 = 0{,}45$	$\frac{1}{2}\cdot 0{,}1 = 0{,}05$	
X = x_2 (1)	$\frac{1}{2}\cdot 0{,}15 = 0{,}075$	$\frac{1}{2}\cdot 0{,}85 = 0{,}425$	

Wir wollen nachweisen, dass auch für bedingte Verbundzeichen für ein festes i die Addition der Zeilenelemente (Zeilensumme) die Wahrscheinlichkeit $P(x_i)$ liefert und für ein festes j die Addition der Spaltenelemente (Spaltensumme) die Wahrscheinlichkeit $P(y_j)$. Ist der Nachweis geführt, können wir wie üblich die Entropien H(X) und H(Y) der Verbundquelle X,Y berechnen. Dafür gehen wir folgendermaßen vor:

Aus $P(A) = P(AB_1) + P(AB_2)$ und $P(AB_j) = P(A)P(B_j/A)$, j = 1,2, ...,m erhalten wir die Gleichung $P(A) = P(A)P(B_1/A) + P(A)P(B_2/A)$.

Damit bekommen wir vom *Satz über die totale Wahrscheinlichkeit*
$P(B) = \sum_{i=1}^{n} P(A_i)P(B/A_i)$ (vgl. Seite 133) die Form $P(A) = \sum_{j=1}^{m} P(A)P(B_j/A)$.
Wir ersetzen darin das Ereignis A durch das Ereignis „*das Zeichen* x_i *wird gesendet*“ und das Ereignis B_j durch das Ereignis „*das Zeichen* y_j *wird empfangen*“. Für ein festgehaltenes i lautet die **Zeilensumme** sodann

$$\boxed{P(x_i) = \sum_{j=1}^{m} P(x_i)P(y_j/x_i)}$$

- Für i=1 und j=1,2
  ergibt sich $P(x_1) = P(x_1)P(y_1/x_1) + P(x_1)P(y_2/x_1) = \frac{1}{2}\cdot 0{,}9 + \frac{1}{2}\cdot 0{,}1 = 0{,}45 + 0{,}05 = \frac{1}{2}$ und
- für i=2 und j=1,2
  ergibt sich $P(x_2) = P(x_2)P(y_1/x_2) + P(x_2)P(y_2/x_2) = \frac{1}{2}\cdot 0{,}15 + \frac{1}{2}\cdot 0{,}85 = 0{,}075 + 0{,}425 = \frac{1}{2}$

Der Vergleich mit der Wahrscheinlichkeitsmatrix X,Y zeigt, dass die Elemente der Matrixzeilen i = 1,2 als Summanden in den korrespondierenden Summengleichungen erscheinen. Und es ist zu sehen, dass die Zeilensummen $P(x_1)$ und $P(x_2)$ mit den Wahrscheinlichkeiten $P(x_i) = \frac{1}{2}$, i = 1,2 der homogenen Nachrichtenquelle

$$X = \begin{pmatrix} x_1 & x_2 \\ (0) & (1) \\ P(x_1) = \frac{1}{2} & P(x_2) = \frac{1}{2} \end{pmatrix}$$ übereinstimmen.

In analoger Weise bestimmen wir nun die Spaltensummen. Wir nehmen wieder den Satz über die totale Wahrscheinlichkeit, belassen ihn in seiner Originalform

$P(B) = \sum_{i=1}^{n} P(A_i)P(B/A_i)$ und ersetzen darin das Ereignis B durch das Ereignis „*das Zeichen* y_j *wird empfangen*“ und das Ereignis A_i durch das Ereignis „*das Zeichen* x_i *wird gesendet*“. Für ein festgehaltenes j lautet demzufolge die **Spaltensumme**

$$\boxed{P(y_j) = \sum_{i=1}^{n} P(x_i)P(y_j/x_i)}$$

- Für j=1 und i=1,2
  ergibt sich $P(y_1) = P(x_1)P(y_1/x_1) + P(x_2)P(y_1/x_2) = \frac{1}{2}\cdot 0{,}9 + \frac{1}{2}\cdot 0{,}15 = 0{,}45 + 0{,}075 = 0{,}525$,
- für j=2 und i=1,2
  ergibt sich $P(y_2) = P(x_1)P(y_1/x_1) + P(x_2)P(y_2/x_2) = \frac{1}{2}\cdot 0{,}1 + \frac{1}{2}\cdot 0{,}85 = 0{,}05 + 0{,}425 = 0{,}475$

Der Vergleich mit der Wahrscheinlichkeitsmatrix X,Y zeigt, dass die Elemente der Matrixspalten j = 1,2 als Summanden in den korrespondierenden Summengleichungen erscheinen. Die Ergebnisse $P(x_1), P(x_2), P(y_1)$ und $P(y_2)$ sind der Matrix X,Y eingetragen. Sie ist hier nochmals dargestellt:

X,Y =

X \ Y	$Y = y_1$ (0)	$Y = y_2$ (1)	
$X = x_1$ (0)	$\frac{1}{2}\cdot 0{,}9 = 0{,}45$	$\frac{1}{2}\cdot 0{,}1 = 0{,}05$	$P(x_1) = 0{,}5$
$X = x_2$ (1)	$\frac{1}{2}\cdot 0{,}15 = 0{,}075$	$\frac{1}{2}\cdot 0{,}85 = 0{,}425$	$P(x_2) = 0{,}5$
	$P(y_1) = 0{,}525$	$P(y_2) = 0{,}475$	$\sum_{j=1}^{2} P(y_j) = \sum_{i=1}^{2} P(x_i) = 1$

Wie für Verbundquellen mit unabhängigen Verbundzeichen, so gilt auch für Verbundquellen

mit abhängigen Verbundzeichen $\sum_{j=1}^{m} P(y_j) = \sum_{i=1}^{n} P(x_i) = 1 \Rightarrow \boxed{\sum_{j=1}^{m}\sum_{i=1}^{n} P(y_j)P(x_i/y_j) = 1}$

Für die Matrix X,Y erhalten wir $\sum_{j=1}^{m=2} P(y_j) = P(y_1) + P(y_2) = 0{,}525 + 0{,}475 = 1$ und

$$\sum_{i=1}^{n=2} P(x_i) = P(x_1) + P(x_2) = 0{,}5 + 0{,}5 = 1$$

Mit den jetzt bekannten Wahrscheinlichkeiten $P(x_1), P(x_2), P(y_1)$ und $P(y_2)$ lassen sich die Entropien H(X) und H(Y) der Nachrichtenquelle X und Y wie folgt berechnen:

$$H(X) = -\sum_{i=1}^{n=2} P(x_i)\log_2 P(x_i) = 0{,}5\cdot 1 + 0{,}5\cdot 1 = 1 \text{ bit}$$

$$H(Y) = -\sum_{j=1}^{m=2} P(y_j)\log_2 P(y_j) = -0525\log_2 0{,}525 - 0{,}475\log_2 0{,}475 = 0{,}488 + 0{,}5102 = 0{,}9982 \text{ bit}$$

Es ist klar erkennbar, dass die Entropie H(Y) der Nachrichtensenke Y = 0,9982 bit **kleiner** ist, als die Entropie der Nachrichtenquelle H(X) = 1 bit. Dies zeigt, dass Information verloren gegangen ist.
Dies ist insofern plausibel, weil durch Störungen auf dem Übertragungskanal nicht jedes von der Quelle gesendete Zeichen am Kanalende richtig interpretiert wird. Es entsteht eine **Rückschlussunsicherheit**, die sogenannte **Äquivokation**. Sie bewirkt, dass aus verschiedenen Zeichen einer Nachrichtenquelle X mit dem Zeichenvorrat $X = \{x_1, x_2, ..., x_n\}$ in der Nachrichtensenke Y das gleiche Zeichen y_j entsteht.
Siehe Bild 40.

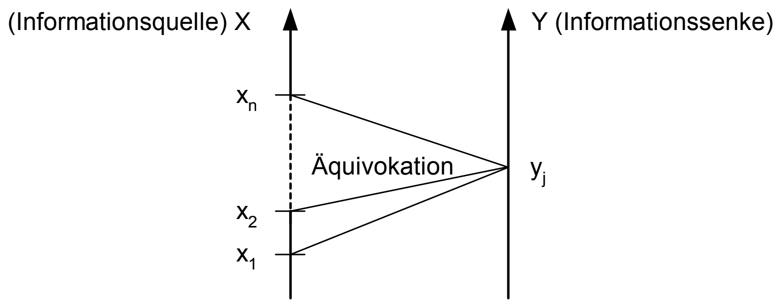

Bild 40: Veranschaulichung der Äquivokation, hervorgerufen durch Störungen

So ist die Äquivokation der Teil der Quellenentropie H(X), der bei der Übertragung durch Störungen verloren geht. Man bezeichnet diesen **Informationsverlust** als bedingte Entropie **H(X/Y)**. Die Berechnung dafür ist noch zu zeigen.
Auf dem Weg dorthin, muss unser Interesse zunächst dem mittleren Informationsgehalt H(X,Y) der Verbundquelle X,Y gehören.

Für die Entropie der Verbundquelle X,Y gilt

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i)\cdot P(y_j)\log_2 P(x_i)\cdot P(y_j) \qquad \text{(vgl. Seite 181)}$$

Da die Unabhängigkeit der Verbundzeichen x_i, y_j **nicht** gegeben ist, muss der Multiplikationssatz für **abhängige** Ereignisse „eingearbeitet“ werden:

$$P(x_i \cdot y_j) = P(x_i)\cdot P(y_j) = P(x_i)\cdot P(y_j/x_i)$$

Eingesetzt in die Gleichung für H(X,Y) erhalten wir

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i)\cdot P(y_j/x_i)\log_2 P(x_i)\cdot P(y_j/x_i)$$

$$= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i)\cdot P(y_j/x_i)[\log_2 P(x_i) + \log_2 P(y_j/x_i)]$$

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i)\cdot P(y_j/x_i)\log_2 P(x_i) - \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i)\cdot P(y_j/x_i)\log_2 P(y_j/x_i)$$

$$= -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i)\sum_{j=1}^{m} P(y_j/x_i) - \sum_{i=1}^{n} P(x_i)\sum_{j=1}^{m} P(y_j/x_i)\log_2 P(y_j/x_i)$$

Betrachten wir als erstes die linke Summe der Gleichung. Für ein festgehaltenes **i** ist $\sum_{j=1}^{m} P(y_j/x_i) = 1$ ein sicheres Ereignis. Denn jedes mal, wenn die Nachrichtenquelle X das Zeichen x_i überträgt, entsteht (überwiegend hervorgerufen durch Fehlinformationen) in der Nachrichtensenke Y irgendein Zeichen y_j. Beide Zeichen treten somit stets zusammen auf. Daher gilt für die **linke** Summe

$$\boxed{-\sum_{i=1}^{n} P(x_i)\log_2 P(x_i)\sum_{j=1}^{m} P(y_j/x_i) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i) = H(X)}$$

Zur Auswertung der rechten Summe machen wir folgende Überlegungen:
Bekanntlich ist X,Y eine Verbundquelle, bei der die Nachrichtenquelle X die Zeichen $X = \{x_i\},\ i = 1,2,\ldots,$ liefert, während daraufhin in der Nachrichtensenke Y die Zeichen $Y = \{y_j\},\ j = 1,2,\ldots,$ entstehen. Die bedingte Entropie der Nachrichtensenke Y unter der Voraussetzung, dass ein Zeichen x_i gesendet wird, ist gegeben durch

$$\boxed{H(Y/x_i) = -\sum_{j=1}^{m} P(y_j/x_i)\log_2 P(y_j/x_i)}$$

Betrachten wir dazu die Übergangsmatrix (Rauschmatrix) Ü = $[P(y_j/x_i)]$ unseres Beispiels:

$$Ü = [P(y_j/x_i)] = \begin{pmatrix} P(y_1/x_1) & P(y_2/x_1) \\ P(y_1/x_2) & P(y_2/x_2) \end{pmatrix} = \begin{pmatrix} 0{,}9 & 0{,}1 \\ 0{,}15 & 0{,}85 \end{pmatrix}$$

Wir sehen, für

- $Y = \{y_1, y_2\}$ lautet unter der Voraussetzung des Zeichens x_1 (Zeile i = 1) die bedingte Entropie

$$H(Y/x_1) = -\sum_{j=1}^{2} P(y_j/x_1)\log_2 P(y_j/x_1)$$

$$= -P(y_1/x_1)\log_2 P(y_1/x_1) - P(y_2/x_1)\log_2 P(y_2/x_1)$$

$$H(Y/x_1) = -0{,}9\log_2 0{,}9 \ - \ 0{,}1\log_2 0{,}1 = 0{,}1368 + 0{,}3322 = 0{,}469 \text{ bit}$$

und für

- $Y = \{y_1, y_2\}$ lautet unter der Voraussetzung des Zeichens x_2 (Zeile i = 2) die bedingte Entropie

$$H(Y/x_2) = -\sum_{j=1}^{2} P(y_j/x_2)\log_2 P(y_j/x_2)$$

$$= -P(y_1/x_2)\log_2 P(y_1/x_2) - P(y_2/x_2)\log_2 P(y_2/x_2)$$

$$H(Y/x_2) = -0{,}15\log_2 0{,}15 \;-\; 0{,}85\log_2 0{,}85 = 0{,}4105 + 0{,}1993 = 0{,}6098 \text{ bit}$$

Der Auftritt dieser Entropien hängt davon ab, mit welcher Wahrscheinlichkeit $P(x_i)$ das ursächliche Zeichen x_i ausgewählt wird. Deshalb binden wir die so erklärten Entropien $H(Y/x_i)$ an die Wahrscheinlichkeiten $P(x_i)$, mit der sie in Erscheinung treten. Durch diese „Gewichtung“ erhalten wir die Folge $\sum_{i=1}^{n} P(x_i)H(Y/x_i)$. Sie lautet für unser Beispiel

$$\sum_{i=1}^{n} P(x_i)H(Y/x_i) = P(x_1)H(Y/x_1) + P(x_2)H(Y/x_2).$$

Die Summation aller Glieder der Folge liefert den **Mittelwert** der bedingten Entropie $H(Y/x_i)$. Dafür schreiben wir

$$\boxed{H(Y/X) = \sum_{i=1}^{n} P(x_i)H(Y/x_i)}$$

Mit $H(Y/x_i) = -\sum_{j=1}^{m} P(y_j/x_i)\log_2 P(y_j/x_i)$ erhalten wir hieraus

$$\boxed{H(Y/X) = -\sum_{i=1}^{n} P(x_i)\sum_{j=1}^{m} P(y_j/x_i)\log_2 P(y_j/x_i)}$$

Das Ergebnis entspricht der **rechten** Summe aus der Gleichung H(X,Y). Somit ist die *bedingte Entropie der Verbundquelle* X,Y gegeben durch

$$\boxed{H(X,Y) = H(X) + H(Y/X)}$$

Wie ist H(Y/X) zu interpretieren? In Folge von Störungen auf dem Übertragungskanal unterscheidet sich ein Teil der empfangenen Zeichen y_j von den gesendeten Zeichen x_i. Es entsteht eine **Vorhersageunsicherheit**, die sogenannte **Irrelevanz**. Sie erzeugt je nach Störung eines Zeichens x_i verschiedene Zeichen y_j. Siehe Bild 41.

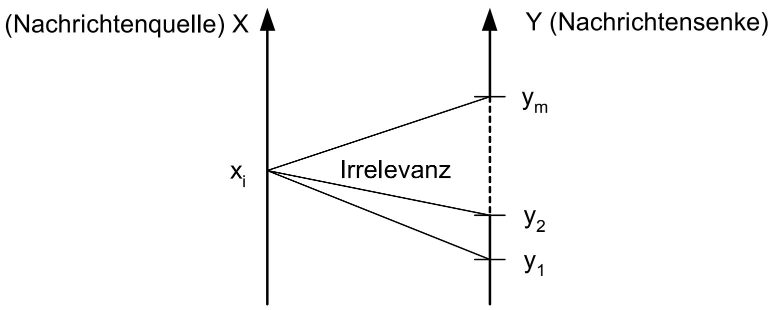

Bild 41: Veranschaulichung der Irrelevanz, hervorgerufen durch Störungen

So ist die Irrelevanz H(Y/X) eine **Fehlinformation**, die in der Nachrichtensenke Y auftritt, die aber nicht von der Nachrichtenquelle X stammt.

Wir benutzen die gewonnenen Gleichungen für unser Beispiel und berechnen die Irrelevanz H(Y/X) und die bedingte Entropie H(X,Y) der Verbundquelle X,Y.

- Die Irrelevanz **H(Y/X)** ist gegeben durch

$$H(Y/X) = \sum_{i=1}^{2} P(x_i) H(Y/x_i) = P(x_1) H(Y/x_1) + P(x_2) H(Y/x_2)\,.$$

Mit $P(x_1) = P(x_2) = \frac{1}{2}$ und $H(Y/x_1) = 0{,}469$ bit, $H(Y/x_2) = 0{,}6098$ bit erhalten wir

$$H(Y/X) = \frac{1}{2} \cdot 0{,}469 \text{ bit} + \frac{1}{2} \cdot 0{,}6098 \text{ bit} = 0{,}2345 + 0{,}3049 = 0{,}5394 \text{ bit}$$

- Die bedingte Entropie **H(X,Y)** ergibt sich mit der bereits berechneten Entropie H(X) = 1 bit zu

$$H(X,Y) = H(X) + H(Y/X) = 1 \text{ bit} + 0{,}5394 \text{ bit} = 1{,}5394 \text{ bit.}$$

Zur Ermittlung der Äquivokation H(X/Y) führen wir eine weitgehend gleiche Rechnung aus. Wir setzen jetzt den *Multiplikationssatz für abhängige Ereignisse* in der Form

$$P(x_i \cdot y_j) = P(x_i) \cdot P(y_j) = P(y_j) \cdot P(x_i / y_j)$$

in die Gleichung $H(X,Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i) \cdot P(y_j) \log_2 P(x_i) \cdot P(y_j)$ (vgl. Seite 181)

ein und erhalten

$$H(X,Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P(y_j) \cdot P(x_i / y_j) \log_2 P(y_j) \cdot P(x_i / y_j)$$

$$= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(y_j)\cdot P(x_i/y_j)[\log_2 P(y_j) + \log_2 P(x_i/y_j)]$$

$$= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(y_j)\cdot P(x_i/y_j)\log_2 P(y_j) - \sum_{i=1}^{n}\sum_{j=1}^{m} P(y_j)\cdot P(x_i/y_j)\log_2 P(x_i/y_j)]$$

$$= -\sum_{j=1}^{m} P(y_j)\log_2 P(y_j)\sum_{i=1}^{n} P(x_i/y_j) - \sum_{j=1}^{m} P(y_j)\sum_{i=1}^{n} P(x_i/y_j)\log_2 P(x_i/y_j)$$

Betrachten wir auch hier als erstes die linke Summe der Gleichung. Für ein festgehaltenes **j** ist $\sum_{i=1}^{n} P(x_i/y_j) = 1$ ein sicheres Ereignis. Denn durch irgendein Zeichen x_i der Quelle X entsteht in der Nachrichtensenke Y (überwiegend hervorgerufen durch Informationsverlust) dasselbe Zeichen y_j (siehe Bild 40, Seite 188). Die beteiligten Zeichen treten somit stets zusammen auf. Deshalb gilt für die **linke** Summe

$$\boxed{= -\sum_{j=1}^{m} P(y_j)\log_2 P(y_j)\sum_{i=1}^{n} P(x_i/y_j) = -\sum_{j=1}^{m} P(y_j)\log_2 P(y_j) = H(Y)}$$

Zur Auswertung der rechten Summe gehen wir folgendermaßen vor:
Bekanntlich ist X,Y eine Verbundquelle, bei der die Nachrichtensenke Y die Zeichen $Y = \{y_j\},\ j = 1,2,\dots,$ empfängt, während dafür die Nachrichtenquelle X die Zeichen $X = \{x_i\},\ i = 1,2,\dots,$ liefert. Die bedingte Entropie der Nachrichtenquelle X unter der Voraussetzung, dass das Zeichen y_j empfangen wird, ist gegeben durch

$$\boxed{H(X/y_j) = -\sum_{i=1}^{n} P(x_i/y_j)\log_2 P(x_i/y_j)}$$

Wir binden die so erklärten Entropien $H(X/y_j)$ an die Wahrscheinlichkeiten $P(y_j)$, mit der sie in Erscheinung treten. So erhalten wir durch diese „Gewichtung“ die Folge $\sum_{j=1}^{m} P(y_j)H(X/y_j)$. Sie lautet für unser Beispiel

$$\sum_{j=1}^{m} P(y_j)H(X/y_j) = P(y_1)H(X/y_1) + P(y_2)H(X/y_2)\,.$$

Die Summation aller Mitglieder der Folge liefert den **Mittelwert** der bedingten Entropie $H(X/y_j)$.

Dafür schreiben wir

$$H(X/Y) = \sum_{j=1}^{m} P(y_j) H(X/y_j)$$

Mit $H(X/y_j) = -\sum_{i=1}^{n} P(x_i/y_j) \log_2 P(x_i/y_j)$ erhalten wir hieraus

$$H(X/Y) = -\sum_{j=1}^{m} P(y_j) \sum_{i=1}^{n} P(x_i/y_j) \log_2 P(x_i/y_j)$$

Das Ergebnis entspricht der **rechten** Summe aus der Gleichung H(X,Y). Somit ist die *bedingte Entropie der Verbundquelle* X,Y auch gegeben durch

$$H(X,Y) = H(Y) + H(X/Y)$$

Durch Gleichsetzung mit H(X,Y) = H(X) + H(Y/X) (vgl. Seite 190) erhalten wir

$$H(X) + H(Y/X) = H(Y) + H(X/Y).$$

Hieraus folgt durch Umstellung

$$T = H(X) - H(X/Y) = H(Y) - H(Y/X)$$

Die linke Seite der Gleichung besagt, dass nur die Differenz

$$T = H(X) - H(X/Y)$$

der Quellenentropie H(X) als sogenannte (mittlere) **Transinformation T** unbeschädigt zum Kanalende gelangt. Die Äquivokation H(X/Y) geht durch Störungen verloren und vermindert so den mittleren Informationsgehalt H(X) der Nachrichtenquelle X.
Aus der rechten Seite der Gleichung T = H(Y) – H(Y/X) erhalten wir durch Umstellung

$$H(Y) = T + H(Y/X)$$

Diese Beziehung besagt, dass für den Empfänger das Kanalende wie eine Nachrichtenquelle mit der Entropie H(Y) wirkt. Sie setzt sich zusammen aus der Transinformation T und der Irrelevanz H(Y/X).

Mit den gewonnenen Erkenntnissen lässt sich ein gestörter Übertragungskanal wie folgt modellieren.

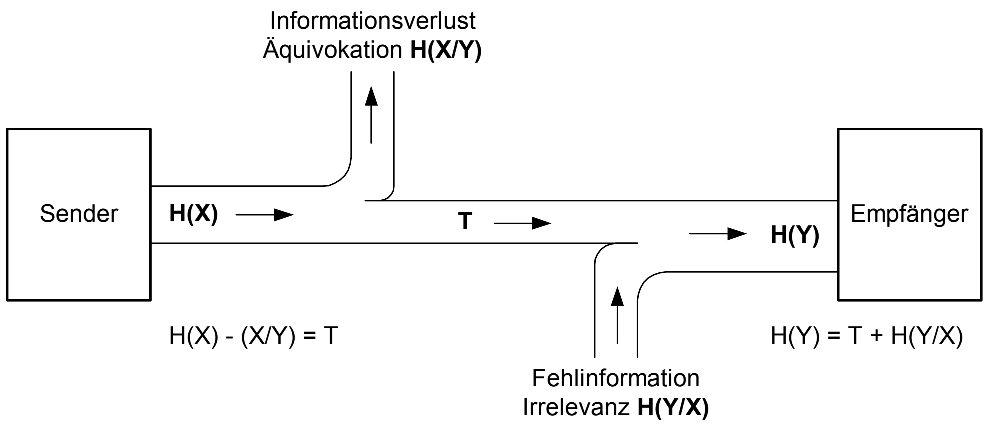

Bild 42: Modell eines gestörten Nachrichtenkanals

Bis hierher haben wir folgende Entropien berechnet:

die Entropie der Nachrichtenquelle	H(X) = 1 bit	vgl. Seite 187
die Entropie der Nachrichtensenke	H(Y) = 0,9982 bit	vgl. Seite 188
die Irrelevanz (Fehlinformation)	H(Y/X) = 0,5394 bit	vgl. Seite 191
die bedingte Entropie der Verbundquelle	H(X,Y) = 1,5394 bit	vgl. Seite 191

Da H(Y) = 0,9982 bit und H(Y/X) = 0,5394 bit bekannt sind, lässt sich durch Auflösung der Beziehung H(Y) = T + H(Y/X) nach **T = H(Y) – H(Y/X)** die mittlere Transinformation T berechnen.

Wir erhalten T = 0,9982 bit – 0,5394 bit = 0,4588 bit.

Abschließend interessieren wir uns für die Äquivokation H(X/Y) unseres Beispiels und benutzen für ihre Berechnung die Formel T = H(X) – H(X/Y). Durch Auflösung nach H(X/Y) = H(X) – T erhalten wir den Informationsverlust

H(X/Y) = 1 bit – 0,4588 bit = 0,5412 bit.

Für eine alternative Berechnung der Äquivokation benutzen wir die Formel

$$H(X/Y) = \sum_{j=1}^{m} P(y_j) H(X/y_j) = -\sum_{j=1}^{m} P(y_j) \sum_{i=1}^{n} P(x_i/y_j) \log_2 P(x_i/y_j)$$

Wir stellen fest, dass die $P(y_j)$ bereits berechnet sind ($P(y_1) = 0{,}525$, $P(y_2) = 0{,}475$ (siehe Seite 187), während die bedingten Wahrscheinlichkeiten $P(x_i/y_j)$ noch nicht bekannt sind.

Für ihre Ermittlung vegleichen wir als erstes die uns bereits bekannte Wahrscheinlichkeitsmatrix X,Y mit der Wahrscheinlichkeitsmatrix Y,X.

X,Y =

X \ Y	$Y=y_1$ $Y=0$	$Y=y_2$ $Y=1$	
$X=x_1$ $X=0$	$P(x_1)P(y_1/x_1)$ $0{,}5\cdot 0{,}9=0{,}45$	$P(x_1)P(y_2/x_1)$ $0{,}5\cdot 0{,}1=0{,}05$	$P(x_1)=0{,}5$
$X=x_2$ $X=1$	$P(x_2)P(y_1/x_2)$ $0{,}5\cdot 0{,}15=0{,}075$	$P(x_2)P(y_2/x_2)$ $0{,}5\cdot 0{,}85=0{,}425$	$P(x_2)=0{,}5$
	$P(y_1)=0{,}525$	$P(y_2)=0{,}475$	1

Y,X

Y \ X	$X=x_1$ $X=0$	$X=x_2$ $X=1$	
$Y=y_1$ $Y=0$	$P(y_1)P(x_1/y_1)$	$P(y_1)P(x_2/y_1)$	$P(y_1)$
$Y=y_2$ $Y=1$	$P(y_2)P(x_1/y_2)$	$P(y_2)P(x_2/y_2)$	$P(y_2)$
	$P(x_1)$	$P(x_2)$	

Wir erkennen, dass sich die $P(x_i / y_j)$ der Y,X-Matrix aus den $P(y_j/x_i)$ der X,Y-Matrix mit dem *Multiplikationssatz für abhängige Ereignisse*

$$P(AB) = P(B)P(A/B) = P(A)P(B/A)$$

berechnen lassen.

(1) Wir setzen $A = x_1$ und $B = y_1$ $(X = 0, Y = 0)$ und bekommen

$P(x_1y_1) = P(y_1)P(x_1/y_1) = P(x_1)P(y_1/x_1)$. Hieraus folgt

$$P(x_1/y_1) = \frac{P(x_1)P(y_1/x_1)}{P(y_1)} = \frac{0{,}5\cdot 0{,}9}{0{,}525} = 0{,}8571$$

(2) Für $A = x_1$ und $B = y_2$ $(X = 0, Y = 1)$ ergibt sich

$$P(x_1y_2) = P(y_2)P(x_1/y_2) = P(x_1)P(y_2/x_1)$$

$$P(x_1/y_2) = \frac{P(x_1)P(y_2/x_1)}{P(y_2)} = \frac{0{,}5\cdot 0{,}1}{0{,}475} = 0{,}1053$$

(3) Für $A = x_2$ und $B = y_1$ $(X = 1, Y = 0)$ ergibt sich

$$P(x_2y_1) = P(y_1)P(x_2/y_1) = P(x_2)P(y_1/x_2).$$

$$P(x_2/y_1) = \frac{P(x_2)P(y_1/x_2)}{P(y_1)} = \frac{0{,}5\cdot 0{,}15}{0{,}525} = 0{,}1429$$

(4) Für $A = x_2$ und $B = y_2$ $(X = 1, Y = 1)$ ergibt sich

$$P(x_2y_2) = P(y_2)P(x_2/y_2) = P(x_2)P(y_2/x_2)$$

$$P(x_2/y_2) = \frac{P(x_2)P(y_2/x_2)}{P(y_2)} = \frac{0{,}5\cdot 0{,}85}{0{,}475} = 0{,}8947$$

Mit diesen Ergebnissen lauten die Werte der Elemente $P(y_j)P(x_i/y_j)$ von Y,X wie folgt:

- Zeile j =1: $P(y_1)P(x_1/y_1) = 0{,}525 \cdot 0{,}8571 = 0{,}45$
  $P(y_1)P(x_2/y_1) = 0{,}525 \cdot 0{,}1429 = 0{,}075$

- Zeile j =2: $P(y_2)P(x_1/y_2) = 0{,}475 \cdot 0{,}1053 = 0{,}05$
  $P(y_2)P(x_2/y_2) = 0{,}475 \cdot 0{,}8947 = 0{,}425$

Zur besseren Übersicht sind sie in der folgenden Wahrscheinlichkeitsmatrix Y,X eingetragen. Dort erkennen wir, dass die **Zeilensummen** $P(y_1) = 0{,}525$ und $P(y_2) = 0{,}475$ erwartungsgemäß mit den Spaltensummen der Wahrscheinlichkeitsmatrix X,Y übereinstimmen. Und auch die **Spaltensummen** $P(x_1) = P(x_2) = 0{,}5$ in Y,X sind mit den Zeilensummen in X,Y identisch.

Des weiteren gilt auch hier $\sum_{j=1}^{2} P(y_j) = \sum_{i=1}^{2} P(x_i) = 1 \Rightarrow \boxed{\sum_{j=1}^{m}\sum_{i=1}^{n} P(y_j)P(x_i/y_j) = 1}$;m, n = 2

Y,X =

Y \ X	$X = x_1$ $X = 0$	$X = x_2$ $X = 1$	
$Y = y_1$ $Y = 0$	$P(y_1)P(x_1/y_1)$ $0{,}525 \cdot 0{,}8571 = 0{,}45$	$P(y_1)P(x_2/y_1)$ $0{,}525 \cdot 0{,}1429 = 0{,}075$	$P(y_1) = 0{,}45 + 0{,}075 = 0{,}525$
$Y = y_2$ $Y = 1$	$P(y_2)P(x_1/y_2)$ $0{,}475 \cdot 0{,}1053 = 0{,}05$	$P(y_2)P(x_2/y_2)$ $0{,}475 \cdot 0{,}8947 = 0{,}425$	$P(y_2) = 0{,}05 + 0{,}425 = 0{,}475$
	$P(x_1) = 0{,}45 + 0{,}05$ $= 0{,}5$	$P(x_2) = 0{,}075 + 0{,}425$ $= 0{,}5$	1

Wir sehen, für

- $X = \{x_1, x_2\}$ lautet unter der Voraussetzung des Zeichens y_1 (Zeile j = 1) die bedingte Entropie

$$H(X/y_1) = -\sum_{i=1}^{2} P(x_i/y_1)\log_2 P(x_i/y_1)$$

$$= -P(x_1/y_1)\log_2 P(x_1/y_1) - P(x_2/y_1)\log_2 P(x_2/y_1)$$

$$H(X/y_1) = -0{,}8571\log_2 0{,}8571 - 0{,}1429\log_2 0{,}1429 = 0{,}1907 + 0{,}401 = 0{,}5917 \text{ bit}$$

und für

- $X = \{x_1, x_2\}$ lautet unter der Voraussetzung des Zeichens y_2 (Zeile j = 2) die bedingte Entropie

$$H(X/y_2) = -\sum_{i=1}^{2} P(x_i/y_2)\log_2 P(x_i/y_2)$$

$$= -P(x_1/y_2)\log_2 P(x_1/y_2) - P(x_2/y_2)\log_2 P(x_2/y_2)$$

$$H(X/y_2) = -0{,}1053\log_2 0{,}1053 - 0{,}8947\log_2 0{,}8947 = 0{,}3419 + 0{,}1436 = 0{,}4855 \text{ bit}$$

• Die Äquivokation **H(X/Y)** ist gegeben durch

$$H(X/Y) = \sum_{j=1}^{m} P(y_j)H(X/y_j) = P(y_1)H(X/y_1) + P(y_2)H(X/y_2)$$

Mit $P(y_1) = 0{,}525$, $P(y_2) = 0{,}475$ und $H(X/y_1) = 0{,}5917$ bit, $H(X/y_2) = 0{,}4855$ bit erhalten wir

$$H(X/Y) = 0{,}525 \cdot 0{,}5917 \text{ bit} + 0{,}475 \cdot 0{,}4855 \text{ bit} = 0{,}3106 + 0{,}2306 = 0{,}5412 \text{ bit}$$

Das Ergebnis entspricht der obigen Berechnung mit der Formel H(X/Y) = H(X) – T. Dort ergab sich H(X/Y) = 1 bit – 0,4588 bit = 0,5412 bit (vgl. Seite 194).

Zur Veranschaulichung unseres Beispiels eines *nicht symmetrisch* gestörten Binärkanals sind die Ergebnisse aller Berechnungen im folgenden Bild eingetragen.

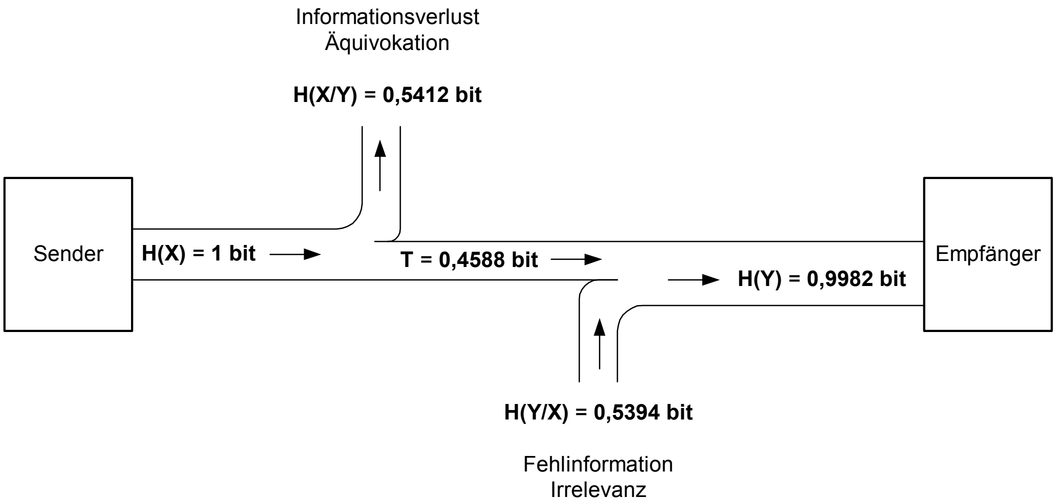

Bild 43: Beispiel eines nicht symmetrisch gestörten Binärkanals

Bleibt ein Übertragungskanal von Störungen verschont, dann gehen weder Informationen verloren, noch kommen Fehlinformationen an der Nachrichtensenke an. Somit gilt für die bedingten Entropen

Äquivokation H(X/Y) = 0 und Irrelevanz H(Y/X) = 0

Aus T = H(X) – H(X/Y) = H(Y) – H(Y/X) entsteht T = H(X) = H(Y) bzw.

$$\boxed{T = H(X)}$$

Dies ist gleichzeitig der maximale Wert (T_{max}), den die mittlere Transinformation T beim **ungestörten** Übertragungskanal erreichen kann. Alle Informationen der Nachrichtenquelle X kommen in der Nachrichtensenke Y an.

Enthält die Nachrichtenquelle X eine homogene Menge von Zeichen, dann ist T_{max} gegeben durch

$$\boxed{T_{max} = H(X) = \log_2 n}$$

Siehe dazu Bild 44.

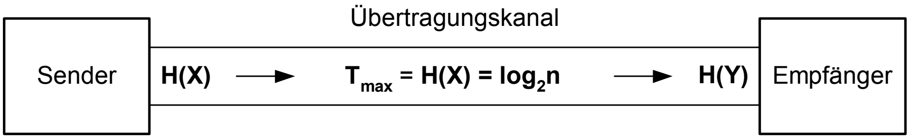

Bild 44: Ungestörter Übertragungskanal

Gleichwohl folgt für die bedingte Entropie der Verbundquelle X ,Y wegen H(Y/X) = 0 und H(X/Y) = 0 aus

$$H(X,Y) = H(X) + H(Y/X) \quad \text{bzw.} \quad H(X,Y) = H(Y) + H(X/Y)$$

die Gleichung $\qquad H(X,Y) = H(X) = H(Y)$ bzw.

$$\boxed{H(X,Y) = H(X)}$$

Die bedingte Entropie der Verbundquelle X ,Y **reduziert** sich auf die Entropie der Nachrichtenquelle X. Enthält die Nachrichtenquelle X eine homogene Menge von Zeichen, dann ist H(X,Y) gegeben durch

$$\boxed{H(X,Y) = H(X) = \log_2 n}$$

Hier schließt sich die folgende interessante Frage an. Gibt es eine Quelle, mit der auch für einen **gestörten** Übertragungskanal eine maximale Transinformation erreichbar ist?
Wir gehen der Frage nach und untersuchen dazu drei verschiedene inhomogene Nachrichtenquellen mit folgenden Wahrscheinlichkeitsfeldern:

Nachrichtenquelle 1

$$X = \begin{pmatrix} x_1 & x_2 & x_3 \\ P(x_1) = \frac{1}{2} & P(x_2) = \frac{1}{4} & P(x_3) = \frac{1}{4} \end{pmatrix}$$

Nachrichtenquelle 2

$$X = \begin{pmatrix} x_1 & x_2 & x_3 \\ P(x_1) = \frac{1}{4} & P(x_2) = \frac{1}{4} & P(x_3) = \frac{1}{2} \end{pmatrix}$$

Nachrichtenquelle 3

$$X = \begin{pmatrix} x_1 & x_2 & x_3 \\ P(x_1) = \frac{1}{3} & P(x_2) = \frac{1}{3} & P(x_3) = \frac{1}{3} \end{pmatrix}$$

Der Übertragungskanal, der die Zeichen x_1, x_2, x_3 an die Nachrichtensenke Y transportiert, habe die Rauschmatrix

$$Ü = [P(y_j / x_i)] = \begin{pmatrix} P(y_1/x_1) & P(y_2/x_1) & P(y_3/x_1) \\ P(y_1/x_2) & P(y_2/x_2) & P(y_3/x_2) \\ P(y_1/x_3) & P(y_2/x_3) & P(y_3/x_3) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0{,}1 & 0{,}8 & 0{,}1 \\ 0{,}2 & 0{,}6 & 0{,}2 \end{pmatrix}$$

In einem Gedankenexperiment schließen wir nun nacheinander die drei verschiedenen Nachrichtenquellen an den gegebenen Kanal an. Wir berechnen für jede Quelle die Entropien H(X) und H(Y), die bedingten Entropien H(Y/X) und mit der Beziehung T = H(Y) – H(Y/X) die interessierenden Transinformationen.
Die Ergebnisse werden zeigen, welche Quelle die größte Transinformation erreicht.
Anfänglich lässt sich vermuten, dass dies der Nachrichtenquelle mit der größten Quellenentropie H(X) gelingt.
Wir beginnen mit der Berechnung der bedingten Entropien HY/X). Für alle drei Quellen gilt

- $H(Y/x_1) = -1 \cdot \log_2 1 - 0 \cdot \log_2 0 - 0 \cdot \log_2 0 \;=\; 0\,\text{bit}$

- $H(Y/x_2) = -0{,}1 \cdot \log_2 0{,}1 - 0{,}8 \cdot \log_2 0{,}8 - 0{,}1 \cdot \log_2 0{,}1$
  = 0,3322 bit + 0,2575 bit + 0,3322 bit = 0,9219 bit

- $H(Y/x_3) = -0{,}2 \cdot \log_2 0{,}2 - 0{,}6 \cdot \log_2 0{,}6 - 0{,}2 \cdot \log_2 0{,}2$
  = 0,4644 bit + 0,4422 bit + 0,4644 bit = 1,371 bit

Hieraus erhalten wir mit der Beziehung

$$\boxed{H(Y/X) = P(x_1)H(Y/x_1) + P(x_2)H(Y/x_2) + P(x_3)H(Y/x_3)}$$

• für die Nachrichtenquelle 1: H(Y/X) = 1/2·0 + 1/4·0,9219 + 1/4·1,371
H(Y/X) = 0 bit + 0,2305 bit + 0,3428 bit = <u>0,5733 bit</u>

• für die Nachrichtenquelle 2: H(Y/X) = 1/4·0 + 1/4·0,9219 + 1/2·1,371
H(Y/X) = 0 bit + 0,2305 bit + 0,6855 bit = <u>0,916 bit</u>

und

• für die Nachrichtenquelle 3: H(Y/X) = 1/3·0 + 1/3·0,9219 + 1/3·1,371
H(Y/X) = 0 bit + 0,3073 bit + 0,457 bit = <u>0,7643 bit</u>

Nun stellen wir auf übliche Weise für jede Quelle das Wahrscheinlichkeitsfeld der resultierenden Verbundquelle X,Y als zweidimensionale Wahrscheinlichkeitsmatrix X,Y dar, ermitteln aus den individuellen Zeilen- und Spaltensummen die Wahrscheinlichkeiten $P(x_i)$ und $P(y_j)$ und berechnen daraus für jede Quelle die Entropien H(X) und H(Y).

Schließlich gewinnen wir mit T = H(Y) – H(Y/X) die individuellen Transinformationen.

Die Wahrscheinlichkeitsmatrix X,Y jeder Nachrichtenquelle hat die folgende allgemeine Form:

X,Y =

Y / X	$Y = y_1$	$Y = y_2$	$Y = y_3$	
$X = x_1$	$P(x_1)P(y_1/x_1)$	$P(x_1)P(y_2/x_1)$	$P(x_1)P(y_3/x_1)$	$P(x_1)$
$X = x_2$	$P(x_2)P(y_1/x_2)$	$P(x_2)P(y_2/x_2)$	$P(x_2)P(y_3/x_2)$	$P(x_2)$
$X = x_3$	$P(x_3)P(y_1/x_3)$	$P(x_3)P(y_3/x_3)$	$P(x_3)P(y_3/x_3)$	$P(x_3)$
	$P(y_1)$	$P(y_2)$	$P(y_3)$	

Und so setzen wir die uns bekannten Auswahlwahrscheinlichkeiten $P(x_i)$ der Zeichen und die Übergangswahrscheinlichkeiten $P(y_j/x_i)$ des Kanals als Verbundwahrscheinlichkeiten $P(x_i)P(y_j/x_i)$ in die Elemente der Matrix ein und erhalten

• für die **Nachrichtenquelle 1**:

X,Y =

Y / X	$Y = y_1$	$Y = y_2$	$Y = y_3$	
$X = x_1$	$\frac{1}{2} \cdot 1$	$\frac{1}{2} \cdot 0$	$\frac{1}{2} \cdot 0$	$P(x_1) = 0{,}5$
$X = x_2$	$\frac{1}{4} \cdot 0{,}1$	$\frac{1}{4} \cdot 0{,}8$	$\frac{1}{4} \cdot 0{,}1$	$P(x_2) = 0{,}25$
$X = x_3$	$\frac{1}{4} \cdot 0{,}2$	$\frac{1}{4} \cdot 0{,}6$	$\frac{1}{4} \cdot 0{,}2$	$P(x_3) = 0{,}25$
	$P(y_1) = 0{,}575$	$P(y_2) = 0{,}35$	$P(y_3) = 0{,}075$	

Daraus ergibt sich:

- $H(X) = -0{,}5 \cdot \log_2 0{,}5 - 0{,}25 \log_2 0{,}25 - 0{,}25 \log_2 0{,}25$
  H(X) = 0,5 bit + 0,5 bit + 0,5 bit = <u>1,5 bit</u>

- $H(Y) = -0{,}575 \cdot \log_2 0{,}575 - 0{,}35 \log_2 0{,}35 - 0{,}075 \log_2 0{,}075$
  H(Y) = 0,4591 + 0,5301 + 0,2803 = <u>1,2695 bit</u>

- T = H(Y) – H(Y/X) = 1,2695 bit – 0,5733 bit = <u>0,6962 bit</u>

• Für die **Nachrichtenquelle 2**:

X,Y =

Y / X	$Y = y_1$	$Y = y_2$	$Y = y_3$	
$X = x_1$	$\frac{1}{4} \cdot 1$	$\frac{1}{4} \cdot 0$	$\frac{1}{4} \cdot 0$	$P(x_1) = 0{,}25$
$X = x_2$	$\frac{1}{4} \cdot 0{,}1$	$\frac{1}{4} \cdot 0{,}8$	$\frac{1}{4} \cdot 0{,}1$	$P(x_2) = 0{,}25$
$X = x_3$	$\frac{1}{2} \cdot 0{,}2$	$\frac{1}{2} \cdot 0{,}6$	$\frac{1}{2} \cdot 0{,}2$	$P(x_3) = 0{,}5$
	$P(y_1) = 0{,}375$	$P(y_2) = 0{,}5$	$P(y_3) = 0{,}125$	

Daraus ergibt sich:

- $H(X) = -0{,}25 \cdot \log_2 0{,}25 - 0{,}25 \log_2 0{,}25 - 0{,}5 \log_2 0{,}5$
  $H(X) = 0{,}5 \text{ bit} + 0{,}5 \text{ bit} + 0{,}5 \text{ bit} = \underline{1{,}5 \text{ bit}}$

- $H(Y) = -0{,}375 \cdot \log_2 0{,}375 - 0{,}5 \log_2 0{,}5 - 0{,}125 \log_2 0{,}125$
  $H(Y) = 0{,}5306 \text{ bit} + 0{,}5 \text{ bit} + 0{,}375 \text{ bit} = \underline{1{,}4056 \text{ bit}}$

- $T = H(Y) - H(Y/X) = 1{,}4056 \text{ bit} - 0{,}916 \text{ bit} = \underline{0{,}4896 \text{ bit}}$

• Für die **Nachrichtenquelle 3**:

X,Y =

X \ Y	$Y = y_1$	$Y = y_2$	$Y = y_3$	
$X = x_1$	$\frac{1}{3} \cdot 1$	$\frac{1}{3} \cdot 0$	$\frac{1}{3} \cdot 0$	$P(x_1) = \frac{1}{3}$
$X = x_2$	$\frac{1}{3} \cdot 0{,}1$	$\frac{1}{3} \cdot 0{,}8$	$\frac{1}{3} \cdot 0{,}1$	$P(x_2) = \frac{1}{3}$
$X = x_3$	$\frac{1}{3} \cdot 0{,}2$	$\frac{1}{3} \cdot 0{,}6$	$\frac{1}{3} \cdot 0{,}2$	$P(x_3) = \frac{1}{3}$
	$P(y_1) = 0{,}433$	$P(y_2) = 0{,}4667$	$P(y_3) = 0{,}1$	

Daraus ergibt sich:

- $H(X) = \log_2 3 = \underline{1{,}5845 \text{ bit}}$

- $H(Y) = -0{,}433 \cdot \log_2 0{,}433 - 0{,}4667 \log_2 0{,}4667 - 0{,}1 \log_2 0{,}1$
  $H(Y) = 0{,}5229 \text{ bit} + 0{,}5131 \text{ bit} + 0{,}3322 \text{ bit} = \underline{1{,}3682 \text{ bit}}$

- $T = H(Y) - H(Y/X) = 1{,}3682 \text{ bit} - 0{,}7643 \text{ bit} = \underline{0{,}6039 \text{ bit}}$

Die Ergebnisse fassen wir in der folgenden Tabelle übersichtlich zusammen:

Nachrichtenquelle 1	Nachrichtenquelle 2	Nachrichtenquelle 3
H(X) = 1,5 bit	H(X) = 1,5 bit	H(X) = 1,5845 bit
H(Y) = 1,2695 bit	H(Y) = 1,4056 bit	H(Y) = 1,2695 bit
H(Y/X) = 0,5733 bit	H(Y/X) = 0,916 bit	H(Y/X) = 0,7643 bit
T = 0,6962 bit	T = 0,4896 bit	T = 0,6039 bit

Wir erkennen sofort, die Nachrichtenquellen 1 und 2 haben die gleiche Quellenentropie H(X) = 1,5 bit, während die Nachrichtenquelle 3 mit H(X) = 1,5845 bit die größte Entropie aufweist. Trotzdem erreicht die Nachrichtenquelle 3 mit T = 0,6039 bit **nicht** die Transinformation der Nachrichtenquelle 1 mit T = 0,6962 bit.
Wir erkennen auch, dass trotz gleicher Entropien der Quellen 1 und 2 unterschiedliche Transinformationen auftreten. Dies führt zu der berechtigten Frage: Warum tritt beim Betrieb mit Quelle 1 eine höhere Transinformation auf, als beim Betrieb mit Quelle 2 oder Quelle 3? Alle drei Quellen übertragen das Zeichen x_1 mit der bedingten Wahrscheinlichkeit (Sörwahrscheinlichkeit) $P(y_1/x_1) = 1$ zur Senke Y. Mit anderen Worten, 100% der gesendeten Zeichen x_1 kommen als y_1 in der Senke an.

Dabei überträgt die Quelle 1 das Zeichen x_1 mit der Auswahlwahrscheinlichkeit $P(x_1) = \frac{1}{2}$,

die Quelle 2 mit $P(x_1) = \frac{1}{4}$ und die Quelle 3 mit $P(x_1) = \frac{1}{3}$. Wir sehen, die Quelle 1 sendet das Zeichen **häufiger** als die Quelle 2 oder die Quelle 3. Die Untersuchung aller Übertagungen lässt folgendes klar erkennen:

Wenig gestörte Zeichen, die häufig übertragen werden, vergrößern die Transinformation.

Um für einen gestörten Übertragungskanal die maximal mögliche Transinformation T_{max} zu erreichen, muss die Quelle an den Kanal angepasst werden. Dies bedeutet, die Quelle muss eine Zeichencodierung bereit halten, die ein solches T_{max} ermöglicht. Shannon hat nachgewiesen, dass es einen optimalen Code geben muss, der eine beliebig fehlerfreie Übertragung von Zeichen erlaubt. Eine konkrete Codierungsvorschrift liefert die Informationstheorie nicht.
Unter diesem Gesichtspunkt ist die Nachrichtenquelle 1 zumindest besser an den gegebenen Kanal angepasst, als Quelle 2 oder 3.
An dieser Stelle sei darauf hingewiesen, dass es für einen gegebenen Übertragungskanal keinen eindeutigen Zusammenhang zwischen der Quellenentropie H(X) und der erreichbaren Transinformation T gibt.
Bei Quelle 1 und Quelle 2 sehen wir, dass es trotz gleicher Quellenentropien H(X) = 1,5 bit für die Transinformation einen oberer Wert $T_{High} = 0{,}6962$ bit und einen unterer Wert $T_{Low} = 0{,}4896$ bit gibt.
Eine optimale Anpassung einer Quelle an den gegebenen Kanal ist nur mit aufwendigen numerischen Mitteln möglich.
(Siehe Spataru, Al.: Theorie der Informationsübertragung, Vieweg-Verlag, Braunschweig 1972).

Den prinzipiellen Zusammenhang zwischen Quellenentropie H(X) und Transinformation T zeigt für einen gegebenen Kanal (und abhängigen Teilquellen) das folgende Diagramm.
(Siehe Mildenberger, O.: Informationstheorie und Codierung, Vieweg-Verlag, Braunschweig 1992)

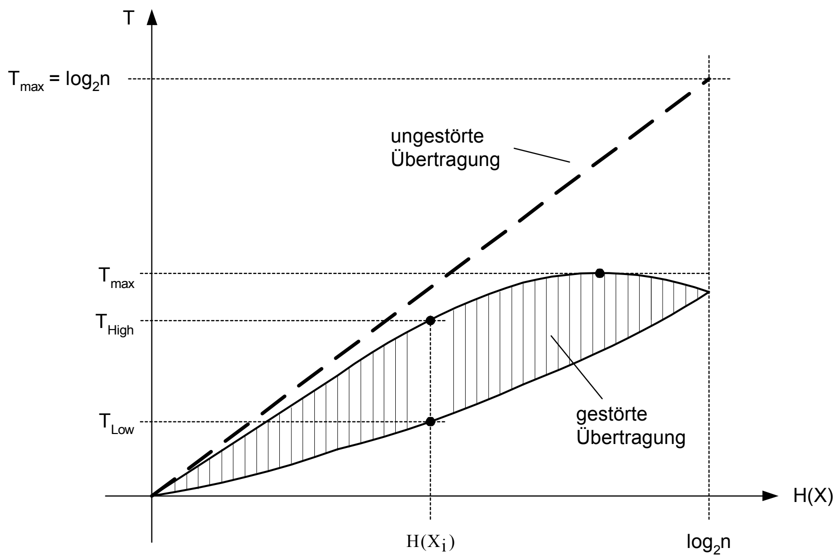

Bild 45: Abhängigkeit der Transinformation von der Quellenentropie bei gegebenem Übertragungskanal

Wir erkennen daraus folgendes:

1) **Ungestörter Übertragungskanal**:

Ist der Zeichenvorrat einer Nachrichtenquelle X homogen, dann gilt $P(x_i) = \frac{1}{n}$ und $H(X) = H(X)_{max}$. Demzufolge erreicht die Transinformation T = H(X) – HX/Y) wegen H(X/Y) = 0 den Maximalwert T_{max}. Es gilt $T_{max} = H(X)_{max} = \log_2 n$ (siehe Seite 198).
Ist der Zeichenvorrat einer Nachrichtenquelle X inhomogen, dann ist deren Quellenentropie H(X) kleiner als die einer homogenen Quelle $H(X)_{max}$. Daher ist auch die erreichbare Transinformation T kleiner als T_{max}. Diesen Zusammenhang verdeutlicht die gestrichelte Linie im Diagramm.

2) **Gestörter Übertragungskanal**:

Die Nachrichtenquellen können sowohl homogen als auch inhomogen sein. Die möglichen Transinformationen T = H(X) – H(X/Y) liegen alle im schraffierten Bereich des Diagramms. Haben verschiedene Quellen die gleichen Quellenentropien H(X), dann gibt es für die Transinformationen (abhängig vom Grad der Anpassung) einen oberen Wert T_{High} und einen unteren Wert T_{Low}.
Ist eine Quelle X optimal an den gegebenen Übertragungskanal angepasst, dann erreicht sie für diesen Kanal die maximal mögliche Transinformation $T_{max} = [H(X) - H(X/Y)]_{max}$. Dieses T_{max} ist jedoch kleiner als das T_{max} des ungestörten Kanals (siehe Diagramm).

2.5 Informationsfluss und Kanalkapazität

Wir nehmen an, dass eine Nachrichtenquelle X aus ihrem homogenen Zeichenvorrat $X = \{x_i\}, i = 1,2, ...,n$ eine Teilmenge von k Zeichen zeitlich nacheinander an den Übertragungskanal abgibt. Das Wahrscheinlichkeitsfeld X und die Entropie H(X) lauten

$$X = \begin{pmatrix} x_1 & x_2 & \cdots\cdots & x_n \\ P(x_1) & P(x_2) & & P(x_n) \end{pmatrix}, \qquad H(X) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i) = \log_2 n$$

Ist z.B. n = 4, dann hat das Wahrscheinlichkeitsfeld die Gestalt

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ (00) & (01) & (10) & (11) \\ P(x_1) = \frac{1}{4} & P(x_2) = \frac{1}{4} & P(x_3) = \frac{1}{4} & P(x_4) = \frac{1}{4} \end{pmatrix}$$ und die Entropie beträgt

$H(X) = \log_2 4 = 2 \text{ bit}$.

Wählen wir k=3, dann schickt die Quelle ein beliebiges Zeichen-Tripel aus ihrem Vorrat in den Übertragungskanal. Die ausgewählten Zeichen stellen sich dort als eine Folge von k Zeichen dar. Siehe Bild 46.

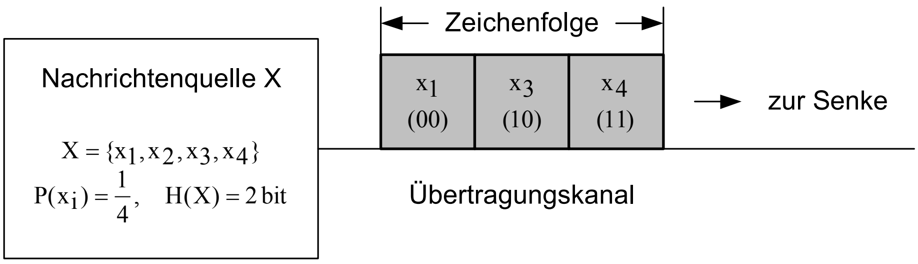

Bild 46: Beliebiges Zeichen-Tripel als Zeichenfolge im Übertragungskanal

Wie jedes Zeichen der Nachrichtenquelle X, so hat auch jedes Zeichen in der Zeichenfolge die Entropie $H(X) = \log_2 4 = 2 \text{ bit}$. Wegen der Homogenität von X gilt: H(X) = $I(x_i) = \log_2 n$.
Wir interessieren uns nun für die Frage: Wie groß ist der mittlere Informationsgehalt der kompletten **Zeichenfolge**? Um dies herauszufinden, gehen wir folgendermaßen vor:
Alle Zeichen der Zeichenfolge sind Zeichen ohne Gedächtnis, d.h. sie sind voneinander unabhängig. So können die von der Nachrichtenquelle X abgegebenen k=3 Zeichen genauso gut von k=3 unabhängigen Teilquellen X_1,X_2 und X_3 geliefert werden. Zusammen bilden diese Teilquellen die Verbundquelle X_1, X_2, X_3 mit der Verbundentropie $H(X_1,X_2,X_3)$. Die Wahrscheinlichkeitsfelder der Teilquellen sind identisch und haben das gleiche Format wie das Wahrscheinlichkeitsfeld der Nachrichtenquelle X. Sie lauten

<u>Teilquelle 1</u>: $$X_1 = \begin{pmatrix} x_{1_1} & x_{1_2} & x_{1_3} & x_{1_4} \\ (00) & (01) & (10) & (11) \\ P(x_{1_1}) = \frac{1}{4} & P(x_{1_2}) = \frac{1}{4} & P(x_{1_3}) = \frac{1}{4} & P(x_{1_4}) = \frac{1}{4} \end{pmatrix}$$

mit $H(X_1) = \log_2 4 = 2 \text{ bit}$

$$\underline{\text{Teilquelle 2}}: \quad X_2 = \begin{pmatrix} x2_1 & x2_2 & x2_3 & x2_4 \\ (00) & (01) & (10) & (11) \\ P(x2_1) = \frac{1}{4} & P(x2_2) = \frac{1}{4} & P(x2_3) = \frac{1}{4} & P(x2_4) = \frac{1}{4} \end{pmatrix}$$

$$\text{mit } H(X_2) = \log_2 4 = 2 \text{ bit}$$

$$\underline{\text{Teilquelle 3}}: \quad X_3 = \begin{pmatrix} x3_1 & x3_2 & x3_3 & x3_4 \\ (00) & (01) & (10) & (11) \\ P(x3_1) = \frac{1}{4} & P(x3_2) = \frac{1}{4} & P(x3_3) = \frac{1}{4} & P(x3_4) = \frac{1}{4} \end{pmatrix}$$

$$\text{mit } H(X_3) = \log_2 4 = 2 \text{ bit}$$

Auf welche Weise die Teilquellen X_1,X_2 und X_3 die gleiche Zeichenfolge wie die einzelne Nachrichtenquelle X abgeben, verdeutlicht das folgende Bild 47.

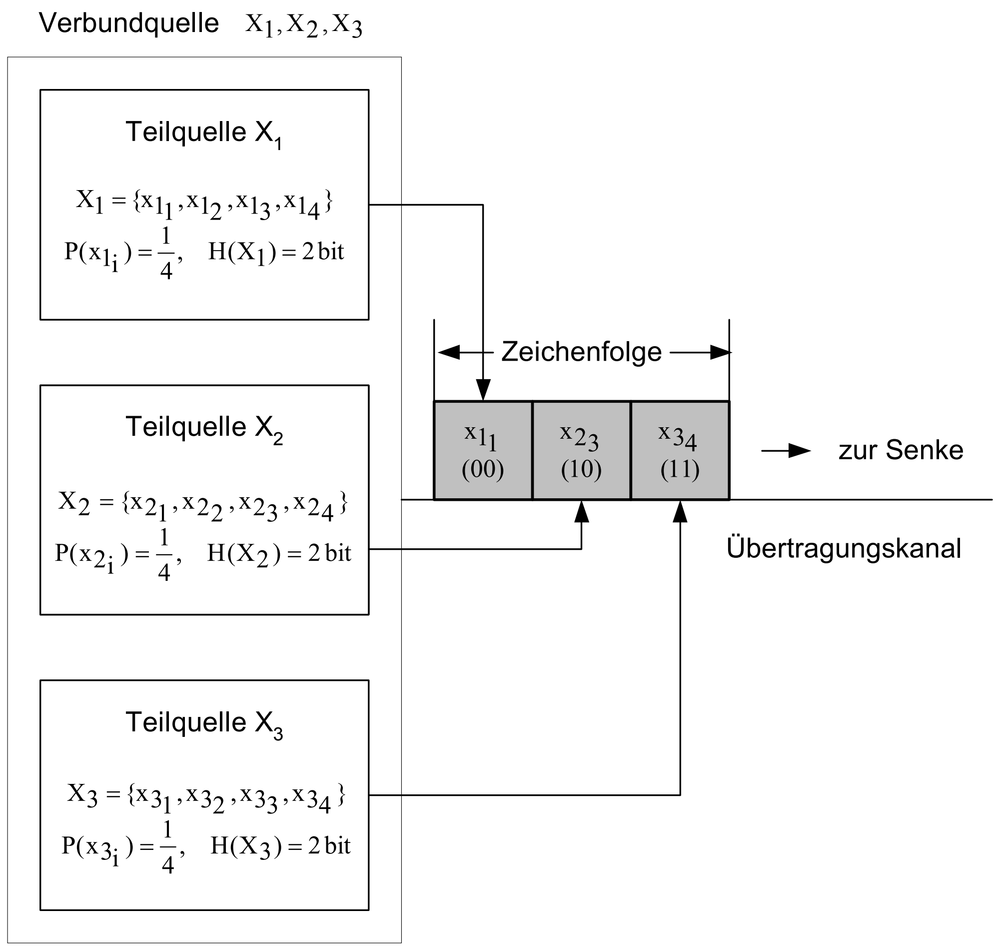

Bild 47: Die Teilquellen X_1,X_2 und X_3 liefern die gleiche Zeichenfolge wie die Nachrichtenquelle X

Wir erkennen, dass das i-te Zeichen der Zeichenfolge von der Teilquelle X_i stammt.

So liefert

- Die Teilquelle X_1 das 1-te Zeichen $X_1 = x_{1_1}$ (00). Dieses entspricht dem 1-ten Zeichen $X = x_1$ (00) der Nachrichtenquelle X,
- die Teilquelle X_2 das 2-te Zeichen $X_2 = x_{2_3}$ (10). Dieses entspricht dem 2-ten Zeichen $X = x_3$ (10) der Nachrichtenquelle X und
- Die Teilquelle X_3 das 3-te Zeichen $X_3 = x_{3_4}$ (11). Dieses entspricht dem 3-ten Zeichen $X = x_4$ (11) der Nachrichtenquelle X,

Die Teilquellen X_1,X_2 und X_3 generieren mit $x_{1_1}, x_{2_3}, x_{3_4} = 00,10,11$ die gleiche Zeichenfolge wie die Nachrichtenquelle X mit $x_1, x_2, x_3 = 00,10,11$. Dies bedeutet, die von X abgegebenen Zeichen der Zeichenfolge können durch die von den Teilquellen X_1,X_2 und X_3 stammenden Zeichen ersetzt werden.
Insgesamt können die drei Teilquellen 4·4·4 = 64 verschiedene Zeichenfolgen zu je 3 Zeichen bilden. Allgemein gilt: Besitzt eine Teilquelle einen Zeichenvorrat von n Zeichen und besteht jede Zeichenfolge aus k Zeichen, dann sind insgesamt n^k verschiedene Zeichenfolgen möglich. Es ist die gleiche Anzahl, die auch eine einzelne Nachrichtenquelle $X = \{x_i\}, i = 1,2, ...,n$ für ein gewähltes k bilden kann.
Jedes Zeichen der Zeichenfolge $x_{1_1}, x_{2_3}, x_{3_4}$ besitzt die Entropie der Teilquelle, von der sie herstammt. So hat das Zeichen x_{1_1} die Entropie $H(X_1) = \log_2 4 = 2$ bit, das Zeichen x_{2_3} die Entropie $H(X_2) = \log_2 4 = 2$ bit und das Zeichen x_{3_4} die Entropie $H(X_3) = \log_2 4 = 2$ bit.

So können wir den mittleren Informationsgehalt einer Zeichenfolge, die aus k Zeichen besteht, als Entropie der Verbundquelle $X_1, X_2, ..., X_k$ *auffassen.*

Bekanntlich ist für unabhängige Teilquellen $X_1, X_2, ..., X_k$ die Entropie der Verbundquelle gegeben durch

$$H(X_1, X_2, ..., X_k) = H(X_1) + H(X_2) +, ..., + H(X_k) \quad \text{(vgl. Seite 182)}$$

In unseren Überlegungen haben wir den Sonderfall gleicher Entropien aller Teilquellen vorausgesetzt. Dafür gilt

$$H(X_1) = H(X_2) =, ... = H(X_k) = H(X)$$

Hieraus ergibt sich für die Entropie der zusammengesetzten Quelle und damit für den mittleren Informationsgehalt einer aus k Zeichen bestehenden Zeichenfolge

$$\boxed{H(X_1) + H(X_2) +, ..., + H(X_k) = k \cdot H(X)}$$

oder kurz

$$\boxed{H(X_1, X_2, ..., X_k) = k \cdot H(X)}$$

Da die beteiligten Teilquellen unabhängig voneinander sind, erreicht die Verbundquelle $X_1, X_2, ..., X_k$ und damit die Zeichenfolge aus k Zeichen die **größte** Entropie k·H(X).

Für unser Beispiel mit der Zeichenfolge $x_{11}, x_{23}, x_{34} = 00,10,11$ aus den Teilquellen X_1, X_2, X_3 ergibt sich demzufolge der mittlere Informationsgehalt

$$H(X_1, X_2, X_3) = k \cdot H(X) = k \cdot 2 \text{ bit} = 6 \text{ bit} .$$

Sind die Teilquellen $X_1, X_2, \ldots\ldots, X_k$ voneinander **abhängig**, dann erhalten wir für den mittleren Informationsgehalt von k Zeichen einer Zeichenfolge eine Verallgemeinerung. Grundlage dafür ist die Beziehung

$$H(X,Y) = H(X) + H(Y/X) \qquad \text{(vgl. Seite 190).}$$

Sie gilt für eine Verbundquelle, deren zwei Teilquellen X,Y voneinander abhängig sind.

Wenden wir sie auf die Teilquellen X_1, X_2 an, dann bekommt sie die Form

- $\underline{H(X_1, X_2) = H(X_1) + H(X_2 / X_1)}$. Gesellt sich die Teilquelle X_3 hinzu, setzen wir

  $Z_1 = X_1, X_2$ und erhalten für $Z_1, X_3 = X_1, X_2, X_3$ die Entropie

  $H(Z_1, X_3) = H(Z_1) + H(X_3 / Z_1) = H(X_1, X_2) + H(X_3 / X_1, X_2)$. Setzen wir die rechte Seite der Gleichung $H(X_1, X_2) = H(X_1) + H(X_2 / X_1)$ darin ein, erhalten wir

- $\underline{H(X_1, X_2, X_3) = H(X_1) + H(X_2 / X_1) + H(X_3 / X_1, X_2)}$. Kommt die Teilquelle X_4 hinzu, setzen wir $Z_2 = X_1, X_2, X_3$ und erhalten für $Z_2, X_4 = X_1, X_2, X_3, X_4$ die Entropie

  $H(Z_2, X_4) = H(Z_2) + H(X_4 / Z_2) = H(X_1, X_2, X_3) + H(X_4 / X_1, X_2, X_3)$. Der Einsatz der rechten Seite der Gleichung $H(X_1, X_2, X_3) = H(X_1) + H(X_2 / X_1) + H(X_3 / X_1, X_2)$ führt zu

- $\underline{H(X_1, X_2, X_3, X_4) = H(X_1) + H(X_2 / X_1) + H(X_3 / X_1, X_2) + H(X_4 / X_1, X_2, X_3)}$

•
•
•

usw.

Für k abhängige Teilquellen $X_1, X_2, \ldots\ldots, X_k$ bekommen wir folgerichtig die Entropie

$$H(X_1, X_2, \ldots\ldots, X_k) = H(X_1) + H(X_2 / X_1) + \ldots\ldots\ldots + H(X_k / X_1, X_2, \ldots\ldots, X_{k-1})$$

Die verkürzte Schreibweise dafür lautet

$$\boxed{H(X_1, X_2, \ldots\ldots, X_k) = H(X_1) + \sum_{i=2}^{k} H(X_i / X_1, X_2, \ldots\ldots, X_{i-1})}$$

Darin bedeutet z.B. $H(X_3 / X_1, X_2)$ die bedingte Entropie des 3-ten Zeichens einer Zeichenfolge, unter der Voraussetzung, dass die vorangegangenen 2 Zeichen (i-1 = 3-1 = 2) aus den Teilquellen X_1 und X_2 bekannt sind.

Betrachten wir noch einmal die Gleichung für abhängige Teilquellen:

$$H(X,Y) = H(X) + H(Y/X)$$

Ist der Übertragungskanal störungsfrei, dann ist H(Y/X) = 0. Die Entropie der Verbundquelle X,Y reduziert sich auf H(X). Es gilt H(X,Y) = H(X).
Ist Y von X unabhängig, dann gilt H(Y/X) = H(Y). Die Entropie der Verbundquelle lautet in diesem Fall H(X,Y) = H(X) + H(Y) und entspricht der Verbundentropie unabhängiger Teilquellen. Für abhängige Teilquellen gilt folgerichtig H(X,Y) < H(Y).
So lauten die Erkenntnisse zusammenfassend

$$\boxed{0 \leq H(Y/X) < H(Y)}$$

Die Anwendung dieser Ungleichung auf die Teilquellen $X_1, X_2, \ldots\ldots, X_k$ führt zu folgenden Ergebnissen:

$$\begin{aligned}
&H(X_2/X_1) < H(X_2)\\
&H(X_3/X_1,X_2) < H(X_3)\\
&H(X_4/X_1,X_2,X_3) < H(X_4)\\
&\quad\bullet\\
&\quad\bullet\\
&\quad\bullet\\
&H(X_k/X_1,X_2,\ldots\ldots,X_{k-1}) < H(X_k)
\end{aligned}$$

Hieraus folgt

$$\boxed{H(X_1,X_2,\ldots\ldots,X_k) = H(X_1) + \sum_{i=2}^{k} H(X_i/X_1,X_2,\ldots\ldots,X_{i-1}) < k \cdot H(X)}$$

oder kurz

$$\boxed{H(X_1,X_2,\ldots\ldots,X_k) < k \cdot H(X)}$$

<u>In Worten</u>: Eine aus k abhängigen Teilquellen zusammengesetzte Nachrichtenuelle hat eine **kleinere** Entropie als eine Nachrichtenquelle, die aus k unabhängigen Teilquellen besteht.

Somit gilt allgemein für eine Verbundquelle $X_1, X_2, \ldots\ldots, X_k$ die Gleichung

$$\boxed{H(X_1,X_2,\ldots\ldots,X_k) \leq k \cdot H(X)}$$

Das Gleichheitszeichen steht für eine Verbundquelle mit unabhängigen Teilquellen. Das Zeichen < steht für eine Verbundquelle mit abhängigen Teilquellen.

Teilen wir beide Seiten der Beziehung durch k, dann ergibt sich

$$\boxed{\frac{H(X_1,X_2,\ldots\ldots,X_k)}{k} \leq H(X)}$$

Die linke Seite der Beziehung heißt **Markow´sche Entropie** H_M (zu Ehren von Markow, Andrei Andrejewitsch, russ. Mathematiker 1856 – 1922). Es gilt

$$\boxed{H_M = \frac{H(X_1, X_2, \ldots\ldots, X_k)}{k}}$$

Die Gleichung bedeutet folgendes: Gibt eine stationäre Nachrichtenquelle eine Folge von k Zeichen an den Übertragungskanal ab, dann ist der mittlere Informationsgehalt pro Zeichen gegeben durch die Markow´sche Entropie.

Aus $\frac{H(X_1, X_2, \ldots\ldots, X_k)}{k} \leq H(X)$ und $H_M = \frac{H(X_1, X_2, \ldots\ldots, X_k)}{k}$ folgt

$$\boxed{H_M \leq H(X)}$$

Diese Beziehung hat folgende Bedeutung: Ist eine Verbundquelle aus unabhängigen Teilquellen zusammengesetzt, dann entspricht die Markow´sche Entropie der Quellenentropie H(X): $H_M = H(X)$.
Besteht die Verbundquelle aus abhängigen Teilquellen, dann ist die Markow´sche Entropie kleiner als die Quellenentropie H(X): $H_M < H(X)$.

Jedes Zeichen x_i, das sich im Übertragungskanal befindet, beansprucht für sich die Sendezeit Δt. Tritt ein Zeichen x_i in der Zeichenfolge mit der Wahrscheinlichkeit $P(X = x_i)$ auf, dann lautet die mittlere Sendezeit pro Zeichen

$$\boxed{\Delta t = \sum_{i=1}^{k} \Delta t_i P(X = x_i)}$$

Bilden wir den Quotienten aus der Quellenentropie H(X) und der mittleren Sendezeit Δt eines Zeichens x_i, dann führt dies zum sogenannten **Informationsfluss I** der Nachrichtenquelle X.

$$\boxed{I = \frac{H(X)}{\sum_{i=1}^{k} \Delta t_i P(X = x_i)} = \frac{H(X)}{\Delta t} \quad \frac{\text{bit}}{\text{s}}}$$

<u>In Worten</u>: Schickt eine Nachrichtenquelle X ihre Zeichen x_i zeitlich nacheinander in den Übertragungskanal, dann heißt die auf die Sendezeit Δt bezogene Quellenentropie H(X) Informationsfluss.

Setzen wir in die Gleichung die **Schrittgeschwindigkeit** $v = \frac{1}{\Delta t}$ ein, dann erhalten wir

$$\boxed{I = v \cdot H(X) \quad \frac{\text{bit}}{\text{s}}}$$

Darin bezeichnet v die Anzahl der Zeichen, die pro Sekunde von der Quelle zur Senke übertragen wird. Sie trägt deshalb die Einheit s^{-1} oder auch die Pseudoeinheit **baud**.
Zur Bestimmung der Kanalkapazität gehen wir folgendermaßen vor: Bekanntlich ist die Transinformation T gegeben durch

$$T = H(X) - H(X/Y); \quad \text{für gestörte Übertragungen}$$
$$T = H(X); \quad \text{für ungestörte Übertragungen}$$

Das Diagramm im Bild 45 (Seite 202) zeigt anschaulich den Zusammenhang zwischen der Quellenentropie H(X) und der erreichbaren Transinformation T einer Nachrichtenquelle X für ungestörte und gestörte Übertragungen.

So können wir den Informationsfluss $I = v \cdot H(X) \quad \frac{\text{bit}}{\text{s}}$ auch durch $\boxed{I = v \cdot T \quad \frac{\text{bit}}{\text{s}}}$ ausdrücken.

- Erfolgt die Übertragung über einen ungestörten Kanal, dann gilt $T_{max} = H(X)_{max} = \log_2 n$.
- Erfolgt die Übertragung über einen gestörten Kanal und ist die Nachrichtenquelle an den gegebenen (gestörten) Übertragungskanal angepasst, dann erreicht die Nachrichtenquelle X die maximale Transinformation $T_{max} = [H(X) - H(X/Y)]_{max}$.

Die Transinformation T_{max} bewirkt, dass auch der Informationsfluss I sein Maximum I_{max} erreicht. Dieser Maximalwert ist als **Kanalkapazität** C definiert und hat folgende Form:

$$\boxed{C = I_{max} = v \cdot T_{max} \quad \frac{\text{bit}}{\text{s}}}$$

Die Kanalkapazität ist eine elementare Eigenschaft des Übertragungskanals. Sie gibt den maximalen Informationsfluss an, den eine Nachrichtenquelle X **auch** bei Störungen des Übertragungskanals erreichen kann.

Wir betreiben den Übertragungskanal weiterhin als binären Kanal und schließen eine Nachrichtenquelle X mit dem Wertevorrat $X = \{x_1, x_2\}$ an. Gibt X seine Zeichen an den Kanal ab, dann wird dort das Zeichen $x_1 = 0$ als **negativer** Impuls und das Zeichen $x_2 = 1$ als **positiver** Impuls dargestellt. Jeder Impuls beansprucht im Kanal die Sendezeit Δt.
Gibt X eine Folge von n Zeichen zeitlich nacheinander an den Übertragungskanal ab, dann erscheinen sie an seinem Eingang als Impulsfolge $x(n \cdot \Delta t)$. Bei einer fehlerfreien Übertragung erhält man dann am Ausgang des Kanals die Impulsfolge $y(n \cdot \Delta t)$. So wird für eine Übertragung die Zeichenfolge der Quelle X in eine Impulsfolge umgewandelt.

Im folgenden Bild 48 ist ein Beispiel für die Umwandlung der Zeichenfolge $X = \{x_2, x_1, x_2, x_2, x_1, x_2\} = \{1,0,1,1,0,1\}$ in eine Impulsfolge skizziert.

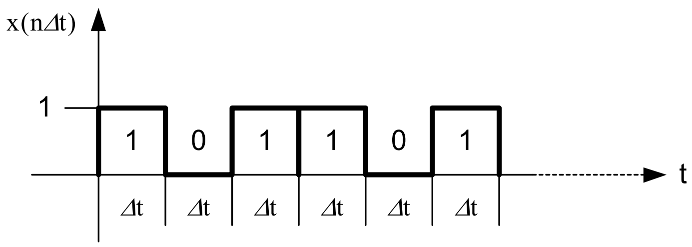

Bild 48: Impulsfolge $x(n \cdot \Delta t)$ am Kanaleingang

Wir interessieren uns für die Frage: Von welcher Dauer muss die Sendezeit Δt eines Zeichens mindestens sein, damit die Impulsfolge (Signale) am Eingang des Kanals ungedämpft am Kanalausgang ankommt? Die Kenntnis darüber hat direkten Einfluss auf die weiterführende Bestimmung der Kanalkapazität C.

Zur Beantwortung gehen wir folgendermaßen vor. In den meisten Fällen zeigt ein Übertragungskanal die gleichen Eigenschaften wie ein Tiefpass. So lässt sich ein Übertragungskanal als Tiefpass modellieren. Unter einem Tiefpass sind alle Schaltungen zu verstehen, die nur Signale x(t) mit Frequenzen unterhalb einer Grenzfrequenz f_g ungedämpft durchlassen. Bei Frequenzen oberhalb der Grenzfrequenz f_g bewirken die Schaltungen eine Abschwächung von x(t) und eine Phasennacheilung. Die einfachste Form des Tiefpasses ist eine Schaltung mit RC-Gliedern. Siehe Bild 49.

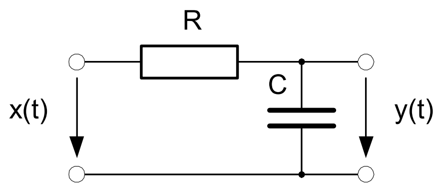

Bild 49: Einfachster Tiefpass

Das Verhalten des Tiefpasses im Zeitbereich zeigen die folgenden Abbildungen.

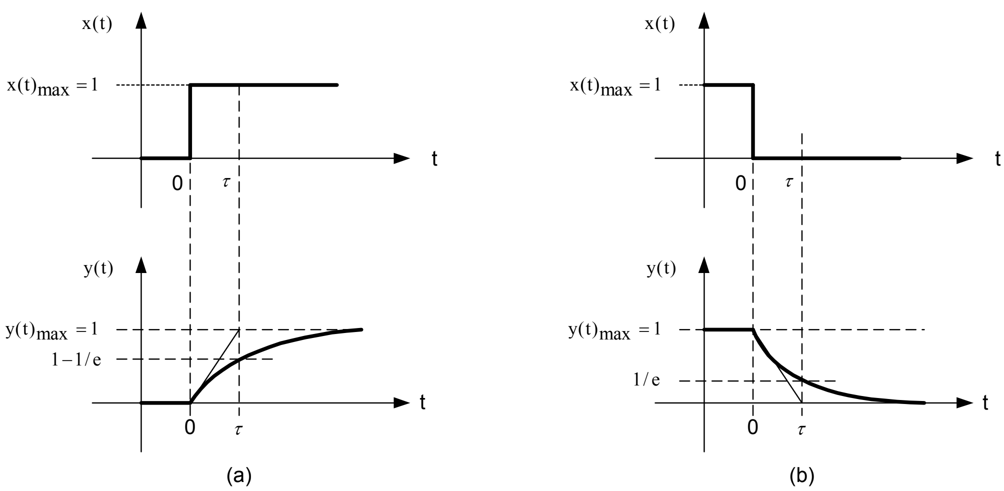

Bild 50: Sprungantworten eines Tiefpasses

Zu Bild 50a: Ändert sich das Eingangssignal x(t) zum Zeitpunkt t = 0 sprungartig von $0 \rightarrow x(t)_{max}$, dann erreicht das Ausgangssignal y(t) den stationären Wert $y(t)_{max} = 1$ nur asymptotisch.
Zu Bild 50b: Ändert sich das Eingangssignal x(t) zum Zeitpunkt t = 0 sprungartig von $x(t)_{max} \rightarrow 0$, dann erreicht das Ausgangssignal y(t) den stationären Wert y(t) = 0 ebenfalls nur asymptotisch.

Als Maß für die sogenannte **Einstellzeit** ist die Zeitkonstante τ definiert. Sie gibt an, wie lange es dauert, bis die Abweichung des Ausgangssignals y(t) vom stationären Wert $y(t)_{max} = 1$ bzw. y(t) = 0 nur noch den e-ten Teil (1/e) der Sprunghöhe $y(t)_{max} = 1$ beträgt (Abweichung: $\frac{1}{e} \cdot 1 = \frac{1}{2{,}718} = 0{,}3679$).

• Der Verlauf von y(t) nach Bild 50a lautet: $y(t) = y_{max}(1 - e^{-\frac{t}{RC}}) = 1 - e^{-\frac{t}{RC}}$ und

• der Verlauf von y(t) nach Bild 50b lautet: $y(t) = y_{max} e^{-\frac{t}{RC}} = e^{-\frac{t}{RC}}$

Zur Berechnung der Zeitkonstanten τ ergibt sich daraus folgender Ansatz:

Für die Abweichung bei einem negativen Impuls gilt $\frac{1}{e} = e^{-\frac{\tau}{RC}}$ und bei einem positiven Impuls gilt $1 - \frac{1}{e} = 1 - e^{-\frac{\tau}{RC}} \Rightarrow -\frac{1}{e} = -e^{-\frac{\tau}{RC}} \Rightarrow \frac{1}{e} = e^{-\frac{\tau}{RC}}$.

Erwartungsgemäß liefern beide Impulsformen die gleichen Abweichungen. Durch Logarithmieren der Gleichung erhalten wir

$\ln\frac{1}{e} = \ln e^{-\frac{\tau}{RC}} \Rightarrow \ln 1 - \ln e = \ln e^{-\frac{\tau}{RC}}$. Hieraus folgt $0 - 1 = -\frac{\tau}{RC}$ und schließlich die

Zeitkonstante $\boxed{\tau = RC}$

Wir wollen nun das Verhalten des Tiefpasses im Frequenzbereich betrachten und ermitteln zunächst den Frequenzgang der RC-Schaltung. Dafür berechnen wir die komplexe Spannungsverstärkung $\underline{A}(j\omega)$ und benutzen dazu die Spannungsteilerformel in komplexer Schreibweise. Ist x(t) die Eingangsspannung $U_1(t)$ und y(t) die Ausgangsspannung $U_2(t)$ des Tiefpasses, dann ist die komplexe Spannungsverstärkung gegeben durch

$$\underline{A}(j\omega) = \frac{\underline{U}_2}{\underline{U}_1} = \frac{\frac{1}{j\omega C}}{R + \frac{1}{j\omega C}} = \frac{\frac{1}{j\omega C}}{\frac{1 + j\omega RC}{j\omega C}} = \frac{1}{1 + j\omega RC}$$

Darin bezeichnen $\underline{U}_1$ und $\underline{U}_2$ die komplexen Spannungsamplituden.

Bilden wir den Betrag $|\underline{A}(j\omega)|$, dann erhalten wir den Amplitudengang

$$|\underline{A}(j\omega)| = \frac{1}{\sqrt{1+\omega^2 R^2 C^2}}$$

Wir beziehen uns nun auf die folgende Mitteilung der Nachrichtentechnik:

- Bei einer Frequenz $f = f_g$, ist $|\underline{A}(j\omega)| = \frac{1}{\sqrt{2}}$ (entspricht Signal-Rauschverhältnis -3db).

Demzufolge setzen wir statt der Kreisfrequenz $\omega = 2\pi f$ die Kreisfrequenz $\omega_g = 2\pi f_g$ ein und erhalten als Amplitudengang

$$\boxed{|\underline{A}(j\omega)| = \frac{1}{\sqrt{2}} = \frac{1}{\sqrt{1+\omega_g{}^2 R^2 C^2}}}$$

Hieraus ergibt sich

$$\sqrt{1+\omega_g{}^2 R^2 C^2} = \sqrt{2} \quad \Rightarrow \quad 1+\omega_g{}^2 R^2 C^2 = 2 \quad \Rightarrow \quad \omega_g{}^2 R^2 C^2 = 1 \quad \Rightarrow \quad \omega_g RC = 1$$

und mit $\omega_g = 2\pi f_g$ folgt $2\pi f_g RC = 1$. Wir lösen nach f_g auf und erhalten für die Grenzfrequenz des Tiefpasses

$$\boxed{f_g = \frac{1}{2\pi \cdot RC} = \frac{1}{2\pi \cdot \tau}}$$

Ist eine Impulsfolge $x(n \cdot \Delta t)$ auf die Grenzfrequenz f_g tiefpassbegrenzt, dann kann nach dem Abtasttheorem die Impulsfolge in **Abständen** von

$$\boxed{\Delta t = \frac{1}{2f_g}}$$

fehlerfrei übertragen werden. Hier bezeichnet Δt die Periodendauer des Abtastsignals s(t). Ist sie genauso groß wie die Sendezeit eines Zeichens (Impulsdauer), dann passt in das Δt eines jeden Impulses der Folge $x(n \cdot \Delta t)$ genau 1 Periode der Abtastfrequenz.

Von Interesse ist deshalb die Impulsdauer. Setzen wir die Grenzfrequenz $f_g = \frac{1}{2\pi \cdot \tau}$ in das Abtasttheorem ein, dann erhalten wir

$$\boxed{\Delta t = \frac{1}{2\frac{1}{2\pi \cdot \tau}} = \pi \cdot \tau \quad \approx 3 \cdot \tau}$$

Im folgenden Bild 51 ist erkennbar, dass alle Impulse der Folge $x(n \cdot \Delta t)$ nach Ablauf von τ eingeschwungen sind. So kann das Abtastsignal s(t) jeden Impuls in seiner zeitlichen Mitte ($3/2 \cdot \tau = 1{,}5\tau$) fehlerfrei abtasten.

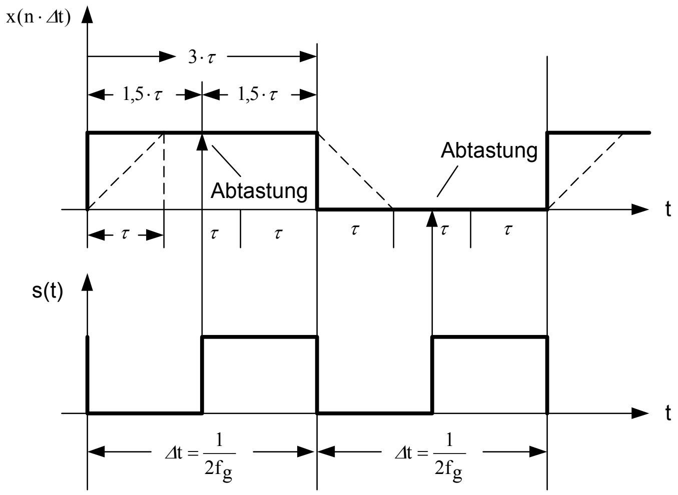

Bild 51: Fehlerfreie Abtastung

- Ist die Sendezeit der Zeichen (Impulsdauer) **kleiner** als die Abtastperiodendauer $\Delta t = \frac{1}{2f_g}$,

  dann liegen die Zeitabschnitte der Impulse und die Perioden der Abtastfrequenz nicht mehr im gleichen Raster. Eine korrekte Interpretation der Impulse ist demzufolge nicht mehr möglich. Dies bedeutet, in einer Folge von Zeichen, darf die Sendezeit eines Zeichens niemals die Zeit $\Delta t = \frac{1}{2f_g}$ unterschreiten.

- Ist die Sendezeit der Zeichen k-mal **größer** als die Abtastperiodendauer $\Delta t = \frac{1}{2f_g}$, dann

  haben im Zeitabschnitt eines Impulses k Abtastperioden Platz. Dies bedeutet, jeder Impuls der Folge $x(n \cdot \Delta t)$ wird k-mal abgetastet. Doch wie sich gleich zeigen wird, reduziert sich unter dieser Voraussetzung die Kanalkapazität.

Setzen wir die optimale Sendezeit der Zeichen $\Delta t = \frac{1}{2f_g}$ in die Beziehung $C = v \cdot T_{max} \frac{bit}{s}$

ein, dann erhalten wir mit $v = \frac{1}{\Delta t} \quad \Rightarrow \quad v = \frac{1}{\frac{1}{2f_g}} = 2f_g$ die optimale Kanalkapazität

$$\boxed{C = 2f_g \cdot T_{max} \frac{bit}{s}}$$

Für ein Zeichen mit der Sendedauer $k \cdot \Delta t$ ergibt sich $v = \frac{1}{k \cdot \Delta t} \quad \Rightarrow \quad v = \frac{1}{\frac{k}{2f_g}} = \frac{2f_g}{k}$.

D.h. die Kanalkapazität verringert sich auf $C = \frac{2f_g \cdot T_{max}}{k} \frac{bit}{s}$.

2.6 Kontinuierliche Nachrichtenquellen und zufällige Signale

Wir stellen uns eine Schachtel vor, in der sich eine Anzahl n rauschender Widerstände gleicher Bauart befindet. Jeder einzelne Widerstand erzeugt eine individuelle Zeitfunktion $x_i(t)$, i = 1, 2, ... , n, die durch einen zufälligem Verlauf gekennzeichnet ist. Bei entsprechender Verstärkung der Signale werden die Widerstände zu Rauschgeneratoren. Im folgenden Bild 52 sind einige mögliche Signalverläufe skizziert.

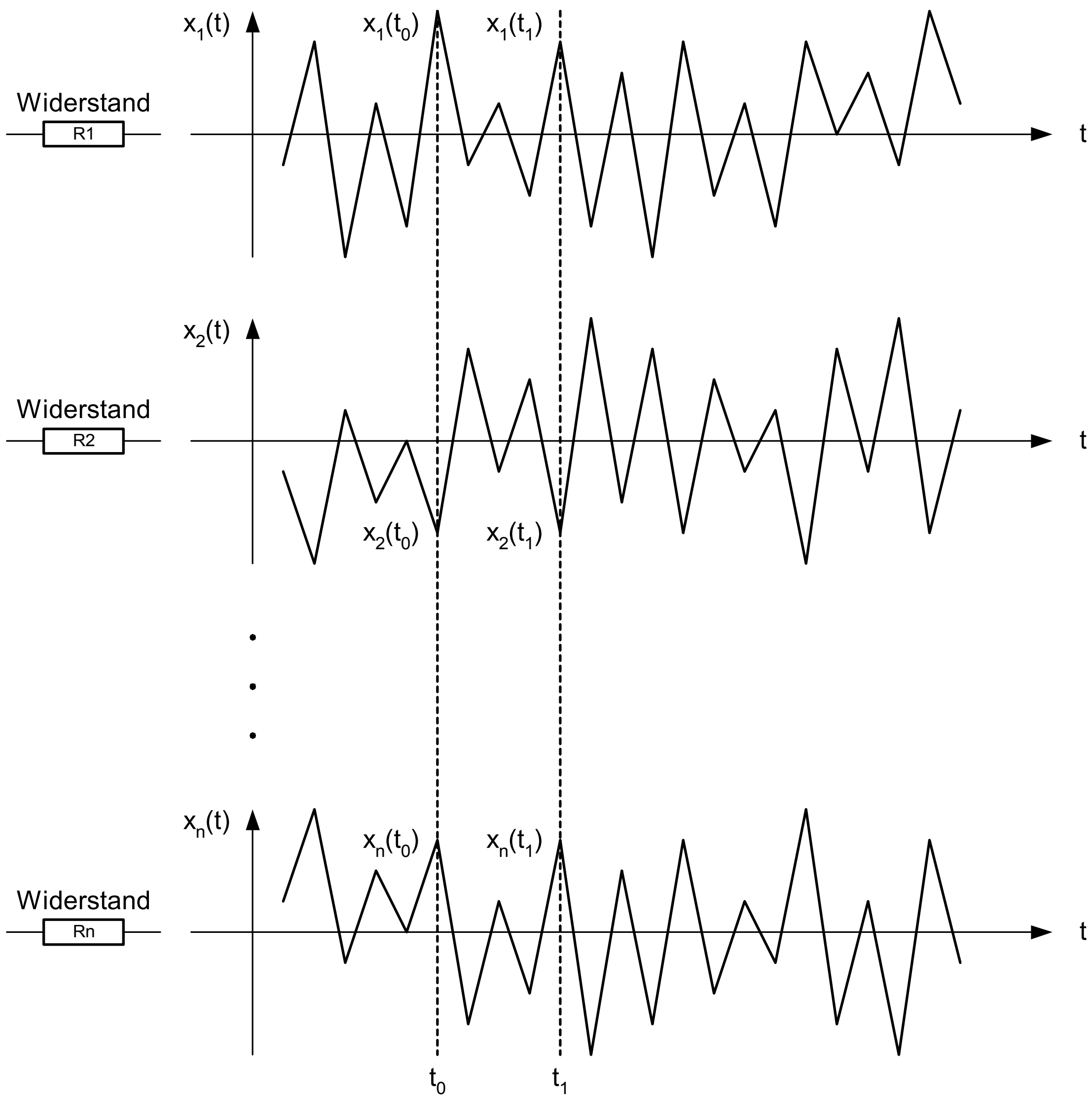

Bild 52: Mögliche Signalverläufe von n rauschenden Widerständen gleicher Bauart

Jede Zeitfunktion $x_i(t)$ stellt eine unabhängige **Musterfunktion** dar. Zu einem bestimmten Zeitpunkt t wählt die Nachrichtenquelle X eine von ihnen mit der Wahrscheinlichkeit p = P(A) aus und sendet sie. A bezeichnet dieses Ereignis.
In ihrer Gesamtheit realisieren die Zeitfunktionen einen sogenannten **Zufallsprozess** bzw. einen stochastischen Prozess X(t). Formal ist X(t) eine Zufallsvariable, die zusätzlich vom Zeitparameter t abhängig ist. Für ein festgehaltenes t entspricht X(t) einer diskreten Zufallsvariable X. So kann zum Zeitpunkt t_0 die Zufallsvariable $X(t_0)$ die Signalwerte $x_1(t_0)$ oder $x_2(t_0)$ oder .... $x_n(t_0)$ annehmen (siehe Bild 52).

$X(t_0)$ hat demzufolge den Signalwertevorrat $W[X(t_0)] = \{x_1(t_0), x_2(t_0), \ldots\ldots x_n(t_0)\}$.
Der Signalwert zum Zeitpunkt t_0 hängt davon ab, welche der Musterfunktionen ausgewählt wurde. Auch die zufälligen Signale $x_i(t_0)$, i = 1, 2, ..., n sind wie andere Zufallswerte nach einer bestimmten Funktion verteilt. Zum Zeitpunkt t_0 hat die Zufallsvariable $X(t_0)$

• die Verteilungsfunktion $F(x, t_0) = P(X(t_0) \le x) = \sum_{x_k \le x} P(X(t_0) = x_k)$ (vgl. Seite 16) und

• den Erwartungswert $\mu(t_0) = E[X(t_0)]$.

Analog hat zum Zeitpunkt t_1 die Zufallsvariable $X(t_1)$

• den Signalwertevorrat $W[X(t_1)] = \{x_1(t_1), x_2(t_1), \ldots\ldots x_n(t_1)\}$ (siehe Bild 52),
• die Verteilungsfunktion $F(x, t_1) = P(X(t_1) \le x) = \sum_{x_k \le x} P(X(t_1) = x_k)$ und

• den Erwartungswert $\mu(t_1) = E[X(t_1)]$

usw. für die Zeitpunkte t_k, k = 2,3, ...,m.

Um einen Zufallsprozess (Gesamtheit aller Zeitfunktionen) X(t) zu charakterisieren, begnügt man sich meist mit dem **arithmetischen** (statistischen) Mittelwert von X(t).
Er liegt in der Nähe des Erwartungswerts E[X(t)] und bestimmt das Anfangsmoment 2ter Ordnung (2tes Moment) $E[X^2(t)]$ und die Varianz $D^2[X(t)] = \sigma^2$ des Essembles X(t).
So lautet zum Zeitpunkt t_0

• der Erwartungswert der Zufallsvariablen $X(t_0)$ mit dem Signalwertevorrat $W[X(t_0)] = \{x_i(t_0)\}$, i = 1, 2, ..., n näherungsweise

$$\boxed{E[X(t_0)] \approx \frac{1}{n} \sum_{i=1}^{n} x_i(t_0)}$$

• das 2te Moment $$\boxed{E[X^2(t_0)] \approx \frac{1}{n} \sum_{i=1}^{n} x_i^2(t_0)}$$ und

• die Varianz $$\boxed{D^2[X(t_0)] = \sigma^2(t_0) \approx E[X^2(t_0)] - (E[X(t_0)])^2}$$

Analog gilt zum Zeitpunkt t_1 für die Zufallsvariable $X(t_1)$ mit dem Signalwertevorrat $W[X(t_1)] = \{x_i(t_1)\}$, i = 1, 2, ..., n

• der Erwartungswert $$\boxed{E[X(t_1)] \approx \frac{1}{n} \sum_{i=1}^{n} x_i(t_1)}$$

• das 2te Moment $\boxed{E[X^2(t_1)] \approx \frac{1}{n}\sum_{i=1}^{n} x_i^2(t_1)}$ und

• die Varianz $\boxed{D^2[X(t_1)] = \sigma^2(t_1) \approx E[X^2(t_1)] - (E[X(t_1)])^2}$

usw. für die Zeitpunkte t_k, k = ..,2,3, ...,m.

Zu jedem Zeitpunkt t_k stimmen die Verteilungsfunktionen, die Erwartungswerte und die Varianzen der Zufallsvariablen $X(t_k)$ und X(t) überein. Somit gilt

$$F(x,t) = F(x,t_0) = F(x,t_1) = \ldots\ldots = F(x,t_m)$$
$$E[X(t)] = E[X(t_0)] = E[X(t_1)] = \ldots\ldots = E[X(t_m)] \text{ und}$$
$$D^2[X(t)] = D^2[X(t_0)] = D^2[X(t_1)] = \ldots\ldots = D^2[X(t_m)]$$

Wir sehen, die statistischen Eigenschaften (Verteilung, Erwartungswert, Varianz) der Zufallsvariablen $X(t_k)$ sind vom Zeitpunkt t_k unabhängig, sie gelten für **jedes t**. Demzufolge stellt sich das Essemble der Musterfunktionen als ein System **stationärer** (zeitunabhängiger) Zufallssignale dar. Dafür lauten die Eigenschaften

• Essemblemittelwert $\boxed{E(X) = E[X(t)] = \frac{1}{n}\sum_{i=1}^{n} x_i(t) = \mu}$

• 2te Moment $\boxed{E(X^2) = E[X^2(t)] = \frac{1}{n}\sum_{i=1}^{n} x_i^2(t)}$ und

• Varianz $\boxed{D^2(X) = D^2[X(t)] = E[X^2(t)] - (E[X(t)])^2 = \sigma^2}$

Die Ergebnisse führen zu der Frage: Kann die statistische Auswertung des Essembles (Gesamtheit aller Musterfunktionen) durch die Auswertung einer **einzigen** Musterfunktion ersetzt werden? Nach dem *Ergodensatz* (Kolmogorow) müssen dafür paarweise unabhängige Zufallsvariablen vorliegen, die alle die gleiche Verteilung haben. Diese Voraussetzungen sind für die Zufallsvariablen $X(t_0), X(t_1), \ldots\ldots, X(t_m)$ erfüllt.

Und so ersetzen wir den Mittelwert E(X) = E[X(t)] = μ und den Mittelwert des 2ten Moments $E(X^2) = E[X^2(t)] = \frac{1}{n}\sum_{i=1}^{n} x_i^2(t)$ des Essembles durch entsprechende **Zeitmittelwerte** einer beliebig ausgewählten Musterfunktion $x_i(t)$, i = 1 **oder** 2 **oder** ....n.

Für den Mittelwert E(X) lautet die Approximation:

$$\boxed{E(X) = E[X(t)] = \frac{1}{n}\sum_{i=1}^{n} x_i(t) = \lim_{T\to\infty}\frac{1}{T}\int_0^T x_i(t)dt} \quad (i = 1 \textbf{ oder } 2 \textbf{ oder } ....n)$$

und für den Mittelwert $E(X^2)$ des 2ten Moments lautet sie:

$$\boxed{E(X^2) = E[X^2(t)] = \frac{1}{n}\sum_{i=1}^{n} x_i{}^2(t) = \lim_{T\to\infty}\frac{1}{T}\int_0^T x_i{}^2(t)dt} \quad (i = 1 \textbf{ oder } 2 \textbf{ oder } ....n)$$

Wenn jeder Essemblemittelwert eines stationären Prozesses X(t) durch den Zeitmittelwert einer beliebigen Musterfunktion ersetzt werden kann, dann nennt man einen solchen Prozess ergodisch. Die aus X(t) hervorgegangene Musterfunktion wird deshalb auch als **ergodisches** Zufallssignal bezeichnet.

Wird bei der betreffenden Musterfunktion ein genügend langer Zeitraum betrachtet ($T \to \infty$), dann treten oberhalb und unterhalb der Zeitachse t gleichgroße Flächen auf. In diesem Fall ist der Mittelwert $E(X) = E[X(t)] = \mu = 0$ und es liegt ein **mittelwertfreies** Zufallssignal vor.

Während es beim binären Übertragungskanal nur zwei Amplitudenstufen gibt, werden beim kontinuierlichen Übertragungskanal wesentlich mehr als zwei Stufen unterschieden. Doch auch hier darf die Sendezeit eines Signals (Amplitudenstufe) nicht kleiner als $\Delta t = \frac{1}{2f_g}$ sein.

Betrachten wir nun die mittlere Leistung eines ergodischen Zufallssignals. Es sei x(t) die Zeitfunktion eines zufällig verlaufenden Stroms, der durch einen Widerstand R fließt. Die von ihm erbrachte Augenblicksleistung P lautet $P = x^2(t)\cdot R$. Da es sich bei x(t) um ein ergodisches Zufallssignal handelt, lässt sich die mittlere Leistung als Zeitmittelwert berechnen. Es gilt

$$P = \lim_{T\to\infty}\int_0^T x^2(t)\cdot R dt \;=\; R\cdot\lim_{T\to\infty}\int_0^T x^2(t)dt \;=\; R\cdot E[X^2(t)]$$

Wir erkennen, dass P zum 2ten Moment $E[X^2(t)]$ direkt proportional ist. Deshalb lässt sich die mittlere Leistung auch durch $\boxed{P = E[X^2(t)]}$ ausdrücken.

Ist x(t) eine zufällig verlaufende Spannung, dann gilt $P = x^2(t)/R$. Die mittlere Leistung als Zeitmittelwert ist somit gegeben durch

$$P = \lim_{T\to\infty}\int_0^T x^2(t)\cdot\frac{1}{R}dt \;=\; \frac{1}{R}\cdot\lim_{T\to\infty}\int_0^T x^2(t)dt \;=\; \frac{1}{R}\cdot E[X^2(t)]$$

Wir sehen, die Proportionalität der mittleren Leistung zum 2ten Moment bleibt erhalten.

Andererseits resultiert aus der Varianz $\sigma^2 = E[X^2(t)] - (E[X(t)])^2$ für ein mittelwertfreies Zufallssignal wegen $E[X(t)] = 0 \Rightarrow (E[X(t)])^2 = 0$ die Form $\sigma^2 = E[X^2(t)] = P$. Somit stellt $\sigma^2 = E[X^2(t)]$ die mittlere Leistung eines **mittelwertfreien** Zufallssignals dar.

2.6.1 Die differentielle Entropie und Transinformation

Bekanntlich ist der mittlere Informationsgehalt einer diskreten und stationären Nachrichtenquelle $X = \{x_1, x_2, \ldots\ldots, x_n\}$ mit den Wahrscheinlichkeiten $P(x_1), P(x_2), \ldots\ldots, P(x_n)$ gegeben durch

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i)$$

Wir setzen nun ein **mittelwertfreies** ergodisches Zufallssignal X(t) voraus, dessen Wahrscheinlichkeitsdichte p(x) im Bild 53 dargestellt ist. Da X(t) zeitunabhängig ist, gilt dies gleichermaßen auch für die Dichte p(x).

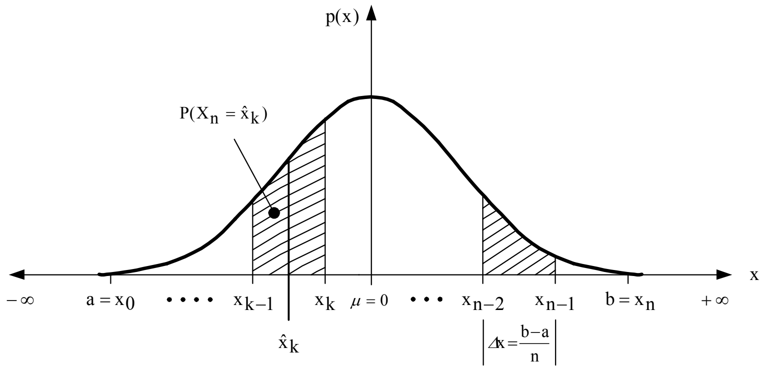

Bild 53: Wahrscheinlichkeitsdichte eines zeitunabhängigen Zufallssignals

Innerhalb eines endlichen Intervalls [a,b] kann das Zufallssignal x jeden Wert annehmen. Die Dichte ist in diesem Bereich deshalb stetig, während sie außerhalb davon verschwindet. Nun diskretisieren wir das kontinuierliche Zufallssignal, indem wir das Intervall [a,b] in n gleich große Abschnitte zerlegen. Die Länge eines jeden Abschnitts ist $\Delta x = \frac{b-a}{n}$. Durch diese Maßnahme ergeben sich auf der x-Achse zwischen a und b die Abschnittspunkte

$$x_0 = a,\ \ x_1 = a + \Delta x,\ \ x_2 = a + 2\Delta x,\ \ \ldots,\ \ x_n = b.$$

Im Abschnitt $[x_{k-1}, x_k]$ wählen wir nun einen Punkt $\hat{x}_k$ aus, der in der Mitte dieses Abschnitts liegt. Für ihn gilt:

$$\boxed{P(X_n = \hat{x}_k) = P(x_{k-1} < X \le x_k) \approx p(\hat{x}_k)\Delta x}$$

Wir sehen, dass sich die Wahrscheinlichkeit $P(X_n = \hat{x}_k)$ eines diskretisierten kontinuierlichen Zufallssignals als Fläche unter der Kurve p(x) zwischen x_{k-1} und x_k darstellt. Wenn wir nun n immer mehr vergrößern, dann wird der Abschnitt $\Delta x = \frac{b-a}{n}$ beliebig klein.

In Anlehnung an die Formel $H(X) = -\sum_{i=1}^{n} P(x_i)\log_2 P(x_i)$ können wir deshalb für die Entropie des Zufallssignals X_n die folgende Näherungsformel angeben:

$$\boxed{H(X_n) = -\sum_{k=1}^{n} P(X_n = \hat{x}_k)\log_2 P(X_n = \hat{x}_k) \approx -\sum_{k=1}^{n} p(\hat{x}_k)\Delta x \log_2(p(\hat{x}_k)\Delta x)}$$

Aus $H(X_n) \approx -\sum_{k=1}^{n} p(\hat{x}_k)\Delta x \log_2(p(\hat{x}_k)\Delta x)$ folgt

$$H(X_n) \approx -\sum_{k=1}^{n} p(\hat{x}_k)\Delta x \cdot [\log_2 p(\hat{x}_k) + \log_2 \Delta x]$$ und hieraus

$$\boxed{H(X_n) \approx -\sum_{k=1}^{n} p(\hat{x}_k)\log_2 p(\hat{x}_k)\Delta x - \sum_{k=1}^{n} p(\hat{x}_k)\Delta x \log_2 \Delta x}$$

Wir sehen, dass $H(X_n)$ aus zwei Summanden besteht. Betrachten wir als erstes den rechten Summanden $-\sum_{k=1}^{n} p(\hat{x}_k)\Delta x \log_2 \Delta x = -\log_2 \Delta x \sum_{k=1}^{n} p(\hat{x}_k)\Delta x$

Bei $n \to \infty$ bzw. $\Delta x \to 0$ existiert für die Summe $\sum_{k=1}^{n} p(\hat{x}_k)\Delta x$ der Grenzwert

$\lim_{n\to\infty} \sum_{k=1}^{n} p(\hat{x}_k)\Delta x = \int_{-\infty}^{\infty} p(x)dx = 1$. Doch dieser Grenzübergang ist **nicht** erlaubt.

Jeder Abschnitt Δx stellt sich als Differenz zwischen zwei benachbarten Signalwerten (Amplitudenstufen) dar, und ist somit ein Maß für die Unterscheidbarkeit der betreffenden Amplitudenstufen. Beispiel: $x_1 = a + \Delta x,\ x_2 = a + 2\Delta x \ \Rightarrow\ x_2 - x_1 = a + 2\Delta x - a - \Delta x = \Delta x$.

Strebt $\Delta x \to 0$, wird diese Unterscheidbarkeit immer schwieriger. Oder anders ausgedrückt, die unterscheidbaren Amplitudenstufen werden immer weniger und verschwinden bei $\Delta x = 0$. Dabei nimmt $-\log_2 \Delta x$ einen unendlich großen Wert an. Da $-\log_2 \Delta x$ in der Gleichung als Summand erscheint, steigert sich die Entropie des kontinuierlicher Signals ins Unendliche. Doch glücklicherweise lassen physikalische Gegebenheiten diesen Grenzübergang nicht zu. Dadurch bleibt n auf einen endlichen Wert begrenzt und Δx wird niemals Null. Dies bedingt stets eine endliche Anzahl **unterscheidbarer** Amplitudenstufen.

Der Grenzübergang $\lim_{n\to\infty} \sum_{k=1}^{n} p(\hat{x}_k)\Delta x = \int_{-\infty}^{\infty} p(x)dx = 1$ hat also nur eine theoretische Bedeutung.

Wegen des zwar großen, aber sinnvoll begrenzten n, kann die Summe $\sum_{k=1}^{n} p(\hat{x}_k)\Delta x$ nur einen Wert in der Nähe von 1 annehmen.

Und so gilt

$$\boxed{-\log_2 \Delta x \sum_{k=1}^{n} p(\hat{x}_k)\Delta x \approx -\log_2 \Delta x}$$

Wenden wir uns nun dem linken Summanden der Gleichung zu. Wir nennen ihn $H_d(X_n)$ und schreiben dafür

$$\boxed{H_d(X_N) \approx -\sum_{k=1}^{n} p(\hat{x}_k)\log_2 p(\hat{x}_k)\Delta x}$$

Für n → ∞ bzw. Δx → 0 strebt $X_n \to X$. Die Annäherung von $H_d(X_n)$ an H(X) wird also mit wachsendem n immer besser. Augenscheinlich existiert der Grenzwert

$$\lim_{n\to\infty} H_d(X_n) = \lim_{n\to\infty} -\sum_{k=1}^{n} p(\hat{x}_k)\log_2 p(\hat{x}_k)\Delta x,$$

gegen den die Folge $H_d(X_n)$ konvergiert. Wir bezeichnen ihn mit

$$\boxed{-\int_a^b p(x)\log_2 p(x)dx}$$

Da die Folge $H_d(X_n)$ gegen diesen Grenzwert konvergiert, macht es Sinn, das angegebene Integral zur sogenannten differentiellen Entropie $H_d(X)$ des kontinuierlichen Zufallssignals zu erklären. Streben a → – ∞ und b → + ∞ und existiert das Integral auf der rechten Seite der Gleichung

$$\lim_{\substack{a\to-\infty \\ b\to+\infty}} -\int_a^b p(x)\log_2 p(x)dx = -\int_{-\infty}^{\infty} p(x)\log_2 p(x)dx,$$

dann ist

$$\boxed{H_d(X) = -\int_{-\infty}^{\infty} p(x)\log_2 p(x)dx}$$

als **differentielle Entropie** definiert. $H_d(X)$ liefert für ein gegebenes kontinuierliches Signal einen eindeutigen Wert. Sie ist vollkommen unabhängig von der Unterteilung des Intervalls [a,b] in Abschnitte $\Delta x = \frac{b-a}{n}$ und ist wie die Entropie einer diskreten Nachrichtenquelle H(X) allein durch die Wahrscheinlichkeitsdichte p(x) festgelegt. Wie sich zeigen wird, ist diese Eigenschaft bedeutungsvoll für die Transinformation T kontinuierlicher Übertragungskanäle. Denn dort wird die Transinformation ausschließlich von differentiellen Entropien bestimmt.

Wenden wir uns nochmals der ursprünglichen Gleichung zu:

$$H(X_n) \approx -\sum_{k=1}^{n} p(\hat{x}_k)\log_2 p(\hat{x}_k)\Delta x - \sum_{k=1}^{n} p(\hat{x}_k)\Delta x \log_2 \Delta x$$

Beim Grenzübergang $n \to \infty$ bzw. $\Delta x \to 0$ strebt $X_n \to X$ und damit die Folge $H(X_n) \to H(X)$. Folgerichtig setzen wir darin die oben gewonnenen Grenzwerte der beiden Summanden ein und erhalten schließlich für die Entropie einer kontinuierlichen Nachrichtenquelle X die Näherung

$$\boxed{H(X) \approx -\int_{-\infty}^{\infty} p(x)\log_2 p(x)dx - \log_2 \Delta x = H_d(X) - \log_2 \Delta x}$$

Analog liefert eine weitgehend gleiche Rechnung die Entropie $H(Y)$ einer kontinuierlichen Nachrichtensenke Y. Für sie gilt

$$\boxed{H(Y) \approx -\int_{-\infty}^{\infty} p(y)\log_2 p(y)dy - \log_2 \Delta y = H_d(Y) - \log_2 \Delta y}$$

Zur Berechnung der differentiellen Äquivokation $H_d(X/Y)$) gehen wir in der bisherigen Weise vor. Bekanntlich ist für einen diskreten Übertragungskanal die Äquivokation gegeben durch

$$H(X/Y) = -\sum_{j=1}^{m} P(y_j)\sum_{i=1}^{n} P(x_i/y_j)\log_2 P(x_i/y_j) \quad \text{(vgl. Seite 193)}$$

Deshalb können wir in Anlehnung daran für einen kontinuierlichen Übertragungskanal die Äquivokation des bedingten Zufallssignals X_n/X_m in folgender Näherung angeben:

$$H(X_n/Y_m) \approx -\sum_{j=1}^{m} p(\hat{y}_j)\Delta y \sum_{i=1}^{n} p(\hat{x}_i/\hat{y}_j)\Delta x\Delta y \log_2 p(\hat{x}_i/\hat{y}_j)\Delta x\Delta y \quad \text{bzw.}$$

$$H(X_n/Y_m) \approx -\sum_{j=1}^{m}\sum_{i=1}^{n} p(\hat{y}_j)\Delta y\, p(\hat{x}_i/\hat{y}_j)\Delta x\Delta y \log_2 p(\hat{x}_i/\hat{y}_j)\Delta x\Delta y$$

Durch Umformung erhalten wir

$$H(X_n/Y_m) \approx -\sum_{j=1}^{m}\sum_{i=1}^{n} p(\hat{y}_j)\Delta y\, p(\hat{x}_i/\hat{y}_j)\Delta x\Delta y[\log_2 p(\hat{x}_i/\hat{y}_j) + \log_2 \Delta x\Delta y]$$

Hieraus folgt

$$\boxed{\begin{aligned} H(X_n/Y_m) \approx &-\sum_{j=1}^{m}\sum_{i=1}^{n} p(\hat{y}_j)\Delta y\, p(\hat{x}_i/\hat{y}_j)\log_2 p(\hat{x}_i/\hat{y}_j)\cdot \Delta x\Delta y \\ &-\sum_{j=1}^{m}\sum_{i=1}^{n} p(\hat{y}_j)\Delta y\, p(\hat{x}_i/\hat{y}_j)\Delta x\Delta y\log_2 \Delta x\Delta y \end{aligned}}$$

Wir sehen, dass $H(X_n/X_m)$ aus zwei Summenden besteht. Betrachten wir als erstes den zweiten Summanden

$$-\sum_{j=1}^{m}\sum_{i=1}^{n} p(\hat{y}_j)\Delta y\, p(\hat{x}_i/\hat{y}_j)\Delta x\Delta y\log_2 \Delta x\Delta y$$

$$=-\sum_{j=1}^{m}\sum_{i=1}^{n} p(\hat{y}_j)\Delta y\, p(\hat{x}_i/\hat{y}_j)\Delta x\Delta y[\log_2 \Delta x+\log_2 \Delta y]$$

$$=-\sum_{j=1}^{m}\sum_{i=1}^{n} p(\hat{y}_j)\Delta y\, p(\hat{x}_i/\hat{y}_j)\Delta x\Delta y\log_2 \Delta x-\sum_{j=1}^{m}\sum_{i=1}^{n} p(\hat{y}_j)\Delta y\, p(\hat{x}_i/\hat{y}_j)\Delta x\Delta y\log_2 \Delta y$$

$$=-\log_2 \Delta x\sum_{j=1}^{m}\sum_{i=1}^{n} p(\hat{y}_j)\Delta y\, p(\hat{x}_i/\hat{y}_j)\Delta x\Delta y-\log_2 \Delta y\sum_{j=1}^{m}\sum_{i=1}^{n} p(\hat{y}_j)\Delta y\, p(\hat{x}_i/\hat{y}_j)\Delta x\Delta y$$

Bei n, m $\to \infty$ bzw. $\Delta x, \Delta y \to 0$ gilt

$$\lim_{\substack{n\to\infty\\ m\to\infty}} \sum_{j=1}^{m}\sum_{i=1}^{n} p(\hat{y}_j)\Delta y\, p(\hat{x}_i/\hat{y}_j)\Delta x\Delta y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(y)\cdot p(x/y)dxdy=1 \quad \text{(vgl. Seite 196)}$$

Eingesetzt in den letzten Ausdruck erhalten wir dafür das Ergebnis $-\log_2 \Delta x-\log_2 \Delta y$

Somit ist der behandelte Summand gegeben durch

$$\boxed{-\sum_{j=1}^{m}\sum_{i=1}^{n} p(\hat{y}_j)\Delta y\, p(\hat{x}_i/\hat{y}_j)\Delta x\Delta y\log_2 \Delta x\Delta y=-\log_2 \Delta x-\log_2 \Delta y}$$

Wenden wir uns nun dem ersten Summanden der Gleichung zu. Wir nennen ihn $H_d(X_n / Y_m)$ und schreiben dafür

$$H_d(X_n / Y_m) \approx -\sum_{j=1}^{m} \sum_{i=1}^{n} p(\hat{y}_j)\Delta y p(\hat{x}_i / \hat{y}_j) \log_2 p(\hat{x}_i / \hat{y}_j) \cdot \Delta x \Delta y$$

Bei n, m → ∞ bzw. Δx, Δy → 0 strebt $X_n / Y_m \rightarrow X / Y$. Die Annäherung von $H_d(X_n / X_m)$ an $H_d(X / Y)$ wird also mit wachsendem n, m immer besser. Demzufolge konvergiert die Folge $H_d(X_n X_m)$ gegen den Grenzwert $-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(y)p(x / y)\log_2 p(x / y)dxdy$. Deshalb macht es Sinn, das angegebene Integral zur sogenannten **differentiellen Äquivokation** $H_d(X / Y)$ des kontinuierlichen Zufallssignals zu erklären. Es gilt

$$H_d(X / Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(y)p(x / y)\log_2 p(x / y)dxdy$$

Auch die differentielle Äquivokation ist vollkommen unabhängig von den Abschnitten Δx, Δy und wird, wie die Äquivokation diskreter Übertragungskanäle, allein durch die Wahrscheinlichkeitsdichte p(x,y) festgelegt.

Wird setzen die oben gewonnenen Grenzwerte der beiden Summanden ein und erhalten für die Äquivokation H(X/Y) eines kontinuierlichen Übertragungskanals die Näherung

$$H(X / Y) \approx -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(y)p(x / y)\log_2 p(x / y)dxdy - \log_2 \Delta x - \log_2 \Delta y \quad \text{bzw.}$$

$$H(X / Y) \approx H_d(X / Y) - \log_2 \Delta x - \log_2 \Delta y$$

Analog liefert eine weitgehend gleiche Rechnung die **Irrelevanz** $H(Y / X)$ eines kontinuierlichen Übertragungskanals. Für sie gilt

$$H(Y / X) \approx -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x)p(y / x)\log_2 p(y / x)dxdy - \log_2 \Delta x - \log_2 \Delta y \quad \text{bzw.}$$

$$H(Y / X) \approx H_d(Y / X) - \log_2 \Delta x - \log_2 \Delta y$$

Ersetzen wir in den gleichwertigen Beziehungen

$$T = H(Y) - H(Y/X) \quad \text{und} \quad T = H(X) - H(X/Y)$$

die Entropien H(X) und H(Y), die Äquivokation H(X/Y) und die Irrelevanz H(Y/X) durch die differentiellen Entsprechungen $H_d(X), H_d(Y)$, $H_d(X/Y)$ und $H_d(Y/X)$, dann erhalten wir für die **Transinformation** eines **kontinuierlichen** Übertragungskanals die folgenden alternativen Gleichungen:

$$\boxed{T = H_d(Y) - H_d(Y/X)} \quad \text{oder} \quad \boxed{T = H_d(X) - H_d(X/Y)}$$

Hieraus erkennen wir folgendes: Da $H_d(X), H_d(Y), H_d(X/Y)$ und $H_d(Y/X)$ ausschließlich von den Wahrscheinlichkeitsdichten p(y), p(y/x), p(x) und p(x/y) bestimmt werden, gilt dies gleichermaßen auch für die Transinformation eines kontinuierlichen Übertragungskanals.

2.6.2 Gleichmäßig verteilte Zufallssignale

Das Zufallssignal X(t) besitze die im Bild 54a skizzierte Wahrscheinlichkeitsdichte p(x).

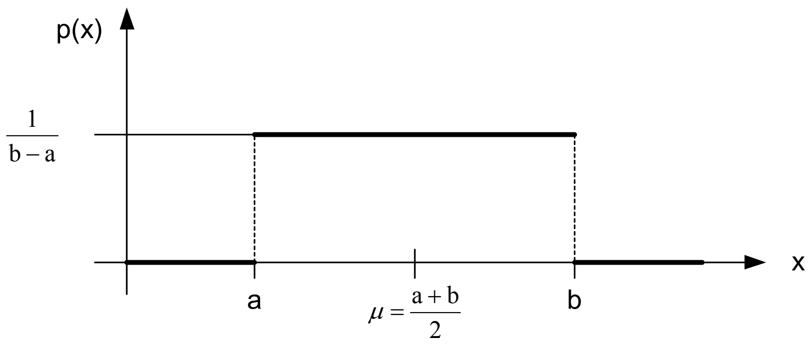

Bild 54a: Dichte eines gleichmäßig verteilten Zufallssignals mit Mittelwert

Wie zu sehen ist, hat X(t) im Intervall [a, b] die gleichmäßige Dichte $p(x) = \frac{1}{b-a}$.

Beweis: Aus $\int_a^b p(x)dx = 1$ folgt $\int_a^b \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b dx = \frac{1}{b-a} x \Big|_a^b = \frac{1}{b-a} \cdot (b-a) = 1$.

Somit gilt: $p(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \le x \le b \\ 0 & \text{für } x < a, x > b \end{cases}$

Da die Dichte zur Achse $x = a + \frac{b-a}{2} = \frac{2a+b-a}{2} = \frac{a+b}{2}$ symmetrisch ist, hat X(t) den

Mittelwert bzw. Erwartungswert $\boxed{\mu = E(X) = \frac{a+b}{2}}$

Ist das Zufallssignal X(t) mittelwertfrei, dann besitzt X(t) die im Bild 54b skizzierte Wahrscheinlichkeitsdichte p(x).

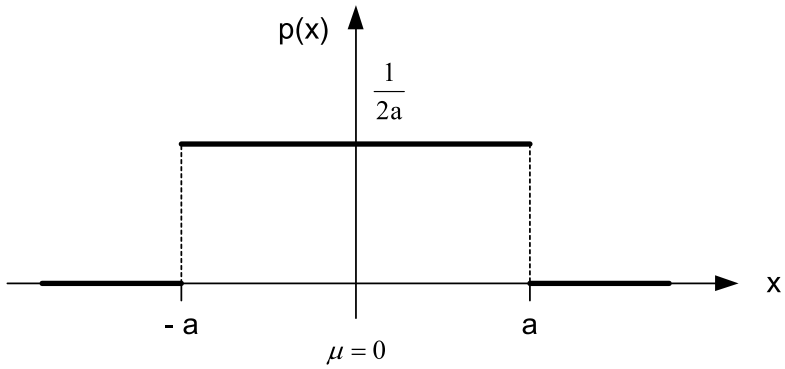

Bild 54b: Dichte eines gleichmäßig verteilten mittelwertfreien Zufallssignals

Da jetzt X(t) im Intervall [- a, a] liegt, hat sie die gleichmäßig Dichte $p(x) = \frac{1}{2a}$.

Beweis: Aus $\int_a^b p(x)dx = 1$ folgt $\int_{-a}^{a} \frac{1}{2a} dx = \frac{1}{2a}\int_{-a}^{a} dx = \frac{x}{2a}\Big|_{-a}^{a} = \frac{1}{2a}\cdot 2a = 1$.

Somit gilt: $p(x) = \begin{cases} \frac{1}{2a} & \text{für } -a \le x \le a \\ 0 & \text{für } x < -a, x > a \end{cases}$

Wir nennen das Intervall [- a, a] den **Aussteuerbereich** des Zufallssignals X(t).

Und so gilt unser Interesse der Leistung und der differentiellen Entropie des mittelwertfreien Zufallssignals X(t) im Bereich [- a, a]. Zur Berechnung der Leistung P benutzen wir die Beziehungen

$$P = E[X^2(t)] = E(X^2) = \int_{-\infty}^{\infty} x^2 p(x)dx \qquad \text{(vgl. Seiten 101 und 218)}$$

Im Aussteuerbereich [- a, a] ist damit die Leistung gegeben durch

$$P = \int_{-a}^{a} x^2 \cdot \frac{1}{2a} dx = \frac{1}{2a}\int_{-a}^{a} x^2 dx = \frac{1}{2a}\cdot \frac{x^3}{3}\Big|_{-a}^{a} = \frac{1}{2a}\left(\frac{a^3}{3} + \frac{a^3}{3}\right) = \frac{1}{2a}\cdot\frac{2a^3}{3} = \frac{a^2}{3}; \qquad \boxed{P = \frac{a^2}{3}}$$

Mit $H_d(X) = -\int_{-\infty}^{\infty} p(x)\log_2 p(x)dx$ und $p(x) = \frac{1}{2a}$ erhalten wir als differentielle Entropie

$$H_d(X) = -\int_{-a}^{a} \frac{1}{2a}\log_2 \frac{1}{2a} dx = -\frac{1}{2a}[\log_2 1 - \log_2 2a]\int_{-a}^{a} dx = \frac{1}{2a}\log_2 2a \int_{-\infty}^{\infty} dx$$

$$= \frac{1}{2a}\log_2 2a \cdot x\Big|_{-a}^{a} = \frac{1}{2a}\log_2 2a \cdot (a+a) = \log_2 2a \qquad \rightarrow \qquad \boxed{H_d(X) = \log_2 2a}$$

In dieser Beziehung ist $H_d(X)$ durch die Aussteuerung a ausgedrückt. Soll sie durch die Leistung P ausgedrückt werden, dann erhält man zunächst aus $P = \frac{a^2}{3}$ die Aussteuerung $a = \sqrt{3P}$. Diese in $H_d(X) = \log_2 2a$ eingesetzt, ergibt

$$H_d(X) = \log_2 2a = \log_2 2\sqrt{3P} = \log_2 \sqrt{4 \cdot 3P} = \log_2 \sqrt{12P}$$

$$\boxed{H_d(X) = \frac{1}{2}\log_2 12P}$$

Wir unterteilen jetzt die Wahrscheinlichkeitsdichte p(x) in n gleichgroße Abschnitte bzw. Amplitudenstufen $\Delta x = \frac{a-(-a)}{n} = \frac{2a}{n}$. Siehe Bild 54c.

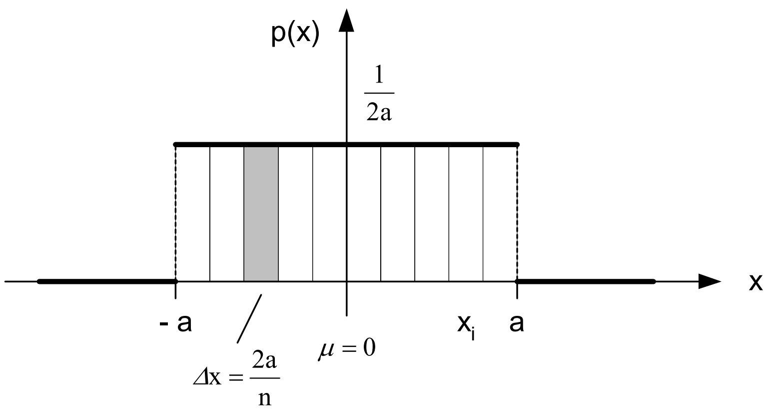

Bild 54c: Unterteilung der Dichte eines gleichmäßig verteilten mittelwertfreien Zufallssignals in gleichgroße Stufen

Durch diese Maßnahme erhalten wir eine konstante Wahrscheinlichkeitsdichte mit n unterscheidbaren Stufen. Jede von ihnen tritt mit der Wahrscheinlichkeit $P(x_i) = \frac{1}{n} = \frac{\Delta x}{2a}$ in Erscheinung. Somit repräsentiert sich die Nachrichtenquelle X als kontinuierliche Quelle mit dem Signalwertevorrat $W(X) = W[X(t)] = \{x_1(t), x_2(t), \ldots\ldots, x_n(t)\}$. Da jedes Signal x_i den Informationsgehalt $I(x_i) = \log_2 n$ besitzt, beträgt die maximale Entropie H(X) der Quelle nach der Shannonschen Formel

$$H(X) = \sum_{i=1}^{n} P(x_i)\log_2 n = \sum_{i=1}^{n} \frac{1}{n}\log_2 n = \frac{n}{n}\log_2 n \qquad \Rightarrow \qquad \boxed{H(X) = \log_2 n}$$

Wir erkennen an diesem Ergebnis die Analogie zur diskreten Nachrichtenquelle X. Bekanntlich hat sie die maximale Entropie $H(X) = \log_2 n$, wenn ihre Zeichen x_i mit der gleichen Wahrscheinlichkeit auftreten. Anmerkung: Ist n=1, ist der Informationsgehalt $I(x_i) = \log_2 1$ gleich Null.

Andererseits ist die Entropie H(X) einer kontinuierlichen Nachrichtenquelle X gegeben durch

$$H(X) \;=\; H_d(X) - \log_2 \Delta x \quad \text{(vgl. Seite 222)}$$

Mit $Hd(X) = \log_2 2a$ und $\Delta x = \dfrac{2a}{n}$ bekommen wir dafür

$$H(X) \;=\; \log_2 2a - \log_2 \frac{2a}{n} = \log_2 2a - [\log_2 2a - \log_2 n] = \log_2 n \quad \Rightarrow \quad \boxed{H(X) = \log_2 n}$$

Wir sehen, diese Berechnung liefert das gleiche das Ergebnis wie die Shannonsche Formel.

Oft wird im Zuammenhang mit einem kontinuierlichen Quellensignal X(t) der Begriff „weißes Rauschen“ genannt. Weißes Rauschen bedeutet, das die zeitlich nacheinander abgegebenen Signale **unabhängig** voneinander sind. Quellensignale mit dieser Eigenschaft, die im zeitlichen Abstand $\Delta t = \dfrac{1}{2f_g}$ abgetastet und identifiziert werden, können deshalb aus der Sicht der Informationstheorie durch zeitdiskrete Signale (Binärcodes) ersetzt werden. Dazu sei folgendes angemerkt: Besteht der Signalwertevorat einer kontinuierlichen Quelle aus n Amplitudenstufen, dann sind für die Umwandlung einer Amplitudenstufe in ein diskretes Signal $\log_2 n$ Entscheidungen notwendig. Diesen Vorgang nennt man Dekodierung.

2.6.3 Normalverteilte Zufallssignale

Das Zufallssignal X(t) besitze die im Bild 55 skizzierte Wahrscheinlichkeitsdichte $p(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Sie geht aus der Dichtefunktion eines $N(\mu,\sigma^2)$ - verteilten Zufallssignals hervor, die uns als $\varphi(x;\mu,\sigma^2) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ bekannt ist (vgl. Seite 109). Setzen wir darin den Erwartungswert $\mu = E(X) = 0$ und die Varianz $\sigma^2 = E(X^2) - (E[X])^2 = E(X^2) - 0 = 1$, dann führt dies zur Dichtefunktion eines N(0,1) – verteilten Zufallssignals $\varphi(x;0,1) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Mit $\varphi(x;0,1) = p(x)$ erhalten wir hieraus die Wahrscheinlichkeitsdichte

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$ eines normalverteilten Zufallssignals.