Quantenkanalkapazität
Seminar: Quanteninformation - Entwicklung, Protokolle, Technologien

Marco Möller

27.06.2007

Inhaltsverzeichnis

1 Einleitung

Das Thema dieser Seminarausarbeitung ist die Kapazität von Kanälen. Dabei werden sowohl klassische als auch quantenmechanische Kanäle betrachtet, die jeweils verrauscht oder ideal sein können.

2 Definitionen

2.1 Entropie

Ich werde im Folgenden die Problemstellung informationstheoretisch betrachten und lösen. Dafür bilden die folgenden Definitionen die Grundlage, deren Bedeutungen später noch behandelt werden. Der Logarithmus log ist in dieser Ausarbeitung als $\log_2$ zu verstehen.

Shannon Entropie $H(X)$ ist für ein Zufallsvariable X, die mit der Wahrscheinlichkeit p_x den Wert x annimmt, definiert als:

$$H(X) \equiv -\sum_x p_x \log p_x \tag{1}$$

Sie lässt sich auch äquivalent als Erwartungswert von $-\log(X)$ schreiben:

$$H(X) = E(-\log(X)) \tag{2}$$

Die Einheit für die Shannon Entropie ist das *bit*.

Von Neumann Entropie $S(\rho)$ ist für ein quantenmechanische Ensemble $\{|x\rangle\}$ mit Auftrittswahrscheinlichkeit $\{p_x\}$ definiert als:

$$S(\rho) \equiv -\mathrm{tr}(\rho \log \rho) \tag{3}$$

Dabei ist $\rho = \sum_x p_x |x\rangle \langle x|$ der quantenmechanische Dichteoperator. Für den Fall, dass die Zustände $\{|x\rangle\}$ orthonormal sind (d.h. $\langle x|y\rangle = \delta_{xy}$), entspricht die von Neumann Entropie der Shannon Entropie $S(\rho) = H(X)$. Die Einheit für die von Neumann Entropie ist das *qubit*.

Über diese elementaren Definitionen hinaus gibt es weitere abgeleitete Formen von Entropieen. Die folgenden Definitionen gibt es analog für die von Neumann Entropie.

Bedingte Entropie $H(X|Y)$ ist das Analogon zur bedingten Wahrscheinlichkeit. Sie misst die (fehlende) Information über X bei zwei korrelierten Zufallsvariablen X, Y im Falle, dass Y komplett bekannt ist. Siehe hierzu auch Abbildung 1 auf der nächsten Seite.

$$H(X|Y) = H(X,Y) - H(Y) \tag{4}$$

Transinformation (mutual information) $H(X:Y)$ ist ein Art Maß für die Korrelation von zwei Zufallsvariablen X, Y. Sie misst die Information, die in beiden Variablen vorhanden ist. Durch Messen von X kann man also bei $H(X:Y) > 0$ auch eine gewisse Aussage über Y machen und umgekehrt. Siehe hierzu auch Abbildung 1 auf der nächsten Seite.

$$\begin{aligned} H(X:Y) &= H(X) + H(Y) - H(X,Y) && (5)\\ &= H(X) - H(X|Y) && (6) \end{aligned}$$

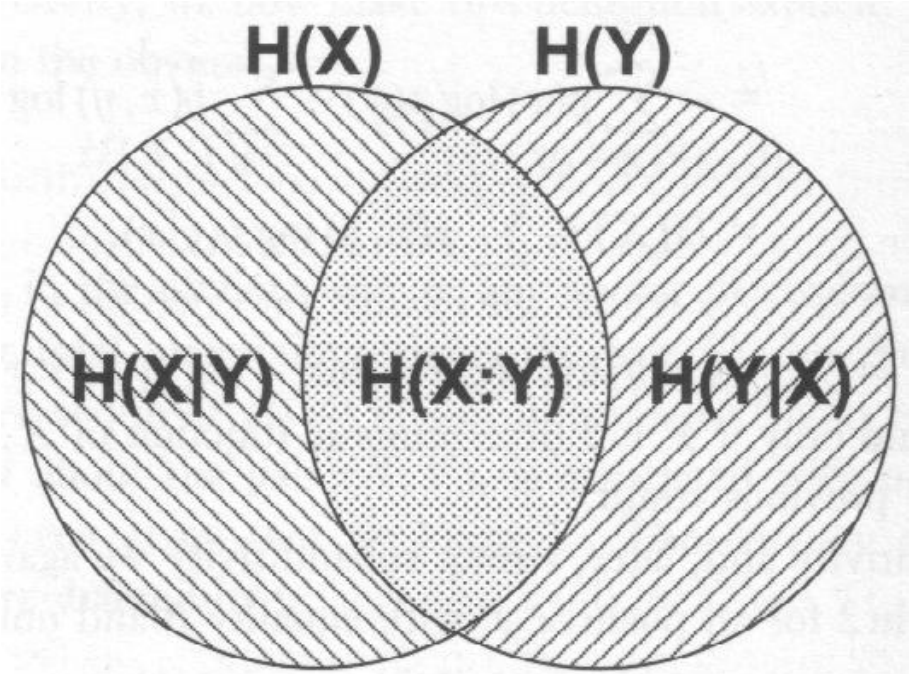

Abbildung 1: Veranschaulichung der bedingten Entropie und Transinformation

2.2 Fidelity

Um nach einer Kompression / Dekompression beurteilen zu können, in wie weit das Dekomprimierte mit dem Original übereinstimmt, benötigen wir hierfür Maße[1]. Eine naheliegende Definition für eine Funktion, die die Ähnlichkeit von σ und ρ beschreibt, wäre die folgende:

$$f(\rho,\sigma) = \begin{cases} 1 & \rho = \sigma \\ 0 & sonst \end{cases} \tag{7}$$

Dies wäre aber nicht zur Bildung von Grenzübergängen geeignet, da diese Funktion unstetig ist.

Ein oft genutztes Maß ist die *Fidelity*. Die statische Version der Fidelity zum Vergleich zweier Quantenzustände ρ, σ ist

$$F(\rho,\sigma) = \mathrm{tr}\sqrt{\rho^{\frac{1}{2}}\sigma\rho^{\frac{1}{2}}} \tag{8}$$

und die dynamische Version zum Vergleich eines Quantenzustands ρ vor und nach der Anwendung der spurerhaltenden Quantenoperation ε ist

$$F(\rho,\epsilon) = \sum_i \left|\mathrm{tr}\left(\rho E_i\right)\right|^2 \tag{9}$$

wobei E_i die Krausoperatoren von ε sind. Hier ist zu beachten, dass

$$F(\rho,\varepsilon(\rho)) \neq F(\rho,\varepsilon) \tag{10}$$

gilt. Es handelt sich also um verschiedene Maße. Sie lassen sich zwar in Beziehung setzen, aber das würde den Rahmen der Ausarbeitung überschreiten. Beiden Fidelitys ist gemein, dass folgendes gilt:

$$0 \leq F(\cdot,\cdot) \leq 1 \tag{11}$$

$$F(\rho,\sigma) = 1 \quad \Leftrightarrow \quad \sigma = \rho \tag{12}$$

$$F(\rho,\epsilon) = 1 \quad \Leftrightarrow \quad \epsilon = \mathrm{id} \tag{13}$$

[1] nicht im mathematischen Sinne

3 Kodierung ohne Rauschen

3.1 Problemstellung

3.1.1 Problemstellung klassisch

Eine Quelle wird im klassischen Fall durch eine Folge von Zufallsvariablen $X_1, X_2, \ldots$ über einem endlichen Alphabet modelliert. Die Werte der Variablen entsprechen den möglichen Ausgaben der Quelle. Eine weitere vereinfachende Annahme ist, dass alle X_i identisch verteilt und unabhängig sind. Diese Annahme wird im Folgenden mit i.i.d. (independent, identically distributed) abgekürzt.

Die zu behandelnde Frage ist nun, wie oft man einen störungsfreien binären Kanal (der nur 2 unterschiedliche Werte annehmen kann, oBdA 0 & 1) im Mittel pro generierten Symbol mindestens benutzen muss, um die Ausgabe der Quelle kodiert zu übertragen. Diese Anzahl wird als *Rate* bezeichnet. Diese Frage kann man auch als die Frage nach der maximalen Kompression des Datenstroms auffassen.

Die Antwort liefert der folgende Satz, den ich nicht seperat beweisen werde, da er ein Spezialfall[2] des entsprechenden quantenmechanischen Satzes ist.

Shannons Satz über rauschfreie Kanalkodierung Sei $\{X_i\}$ eine i.i.d. Informationsquelle mit einer Entropierate[3] $H(X)$. Sei $R > H(X)$. Dann existiert ein zuverlässiges Kompressionsschema der Rate R für diese Quelle. Umgekehrt ist jedes Kompressionsschema für $R < H(X)$ nicht zuverlässig.

3.1.2 Motivation: typische Sequenzen

Um ein Gefühl für die Korrektheit des folgenden Beweises zu bekommen, möchte ich hier die dort verwendeten typischen Sequenzen motivieren.

Betrachten wir einmal eine Ausgabe der Länge n einer binären i.i.d. Quelle. Diese Quelle lässt sich vollständig dadurch charakterisieren, dass die Wahrscheinlichkeit p für die Ausgabe einer 1 angegeben wird. Die Wahrscheinlichkeit für eine Ausgabe $x_1, \ldots, x_n$ beträgt

$$\begin{aligned} p(x_1, \ldots, x_n) &= p(x_1)\, p(x_2) \cdots p(x_n) && (14) \\ &\approx p^{np} (1-p)^{n(1-p)} && (15) \end{aligned}$$

wobei die Abschätzung dadurch zu Stande kommt, dass man für großes n erwarten würde, dass die 1 ca. $n \cdot p$ mal vorkommt. Dies wird im Folgenden durch das Gesetz der Großen Zahlen noch präzisiert.

Durch Logarithmieren beider Seiten erhalten wir

$$\begin{aligned} \log p(x_1, \ldots, x_n) &\approx np \cdot \log p + (1-p)\, n \cdot \log(1-p) && (16) \\ &= -nH(X) && (17) \end{aligned}$$

und durch anschließendes Exponenzieren folgende Beziehung zwischen p und H:

$$p(x_1, \ldots, x_n) \approx 2^{-nH(X)} \tag{18}$$

[2] siehe Definitionen der Entropien

[3] Entropie pro generiertem Symbol

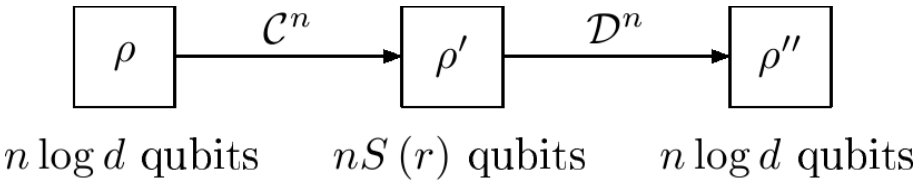

Abbildung 2: (De-)Kompression einer Quanteninformationsquelle [1]

Das heißt, dass es nur etwa $2^{nH(X)} \leq 2^n$ Sequenzen gibt, die häufig vorkommen.[4] Deshalb muss eine effiziente Kodierung nur solche *typischen* Sequenzen berücksichtigen und kann somit die Daten um den Faktor $H(X)$ *komprimieren*! Ein Kompressionsschema, dass diese Beobachtung nutzt, ist die Huffmann-Kodierung.

3.1.3 Problemstellung quantenmechanisch

Eine quantenmechanische i.i.d. Quelle wird durch einen Hilbertraum H und einem Dichteoperator ρ darauf beschrieben. Eine Ausgabe der Quelle der Länge n wird durch einen Dichteoperator $\underbrace{\rho \otimes \rho \otimes \cdots \otimes \rho}_{n\text{-mal}} = \rho^{\otimes n}$ auf $H^{\otimes n}$ beschrieben.

Wie im klassischen Fall stellen wir uns im Folgenden die Frage nach der *maximalen Kompression*. Dazu müssen wir definieren, was wir unter einer quantenmechanischen Kompression verstehen. Ein Kompressionsschema der Rate R besteht aus zwei Familien von Quantenoperationen, der Kompression $\mathcal{C}^n$ und der Dekompression $\mathcal{D}^n$ mit

$$\mathcal{C}^n : \quad H^{\otimes n} \quad \rightarrow 2^{Rn} \tag{19}$$
$$\mathcal{D}^n : \quad 2^{Rn} \quad \rightarrow H^{\otimes n} \tag{20}$$

wie in Abb. 2 visualisiert ist. Zuverlässig ist ein Kompressionsschema, wenn die Fehlerrate von $\mathcal{D}^n \circ \mathcal{C}^n$ auf $\rho^{\otimes n}$ für große n gegen 0 geht. Eine geringe Fehlerrate entspricht einer hohen Fidelity, somit ist dies gleichbedeutend zu

$$F\left(\rho^{\otimes n}, \mathcal{D}^n \circ C^n\right) \rightarrow 1 \tag{21}$$

Die Antwort auf die Frage nach der maximalen Rate R für ein zuverlässiges Kompressionsschema liefert folgender Satz:

Schumacher: quantenmechanische rauschfreie Kanalkodierung Sei $\{H, \rho\}$ eine i.i.d. Quanteninformationsquelle. Für $R > S(\rho)$ existiert ein zuverlässiges Kompressionsschema der Rate R für diese Quelle. Umgekehrt ist jedes Kompressionsschema für $R < S(\rho)$ nicht zuverlässig.

Hier sieht man direkt die Analogie zum klassischen Fall, indem man $S(\rho)$ durch $H(X)$ ersetzt.

3.2 Typische Unterräume

In Abschnitt 3.1.2 auf der vorherigen Seite haben wir typische Sequenzen ja bereits motiviert, nun folgt die formale Definition von ϵ-typisch.

[4] Für eine binäre Wahrscheinlichkeitsverteilung gilt $H(X) \leq 1$.

3.2.1 ϵ-typisch

Sei $\rho = \sum_x p(x) |x\rangle \langle x|$ orthonormale Dekomposition. Eine Sequenz $x_1, \ldots, x_n$ wird ϵ-typisch genannt, falls

$$\left| \frac{1}{n} \log \left(\frac{1}{p(x_1) p(x_2) \cdots p(x_n)} \right) - S(\rho) \right| \leq \epsilon \tag{22}$$

gilt. Entsprechend wird ein zugehöriger Zustand $|x_1\rangle |x_2\rangle \cdots |x_n\rangle$ ϵ-typisch genannt.

Der Unterraum aller ϵ-typischen Zustände wird mit $T(n, \epsilon)$ bezeichnet. Der entsprechende Projektor auf diesen Unterraum ist

$$P(n, \epsilon) = \sum_{x\ \epsilon\text{-typisch}} |x_1\rangle \langle x_1| \otimes \cdots \otimes |x_n\rangle \langle x_n| \tag{23}$$

3.2.2 Satz über typische Unterräume

ϵ-typischen Unterräume haben einige Eigenschaften, die sich zur Kompression nutzen lassen:

1. Sei $\epsilon > 0$. Für jedes $\delta > 0$ und genügend große n gilt:

$$\operatorname{tr}\left(P(n, \epsilon) \rho^{\otimes n} \right) \geq 1 - \delta \tag{24}$$

2. Für jedes $\epsilon > 0$ und $\delta > 0$ und genügend große n erfüllt die Dimension $|T(n, \epsilon)| = \operatorname{tr}(P(n, \epsilon))$ von $T(n, \epsilon)$ die folgende Ungleichung:

$$(1 - \delta) 2^{n(S(\rho) - \epsilon)} \leq |T(n, \epsilon)| \leq 2^{n(S(\rho) + \epsilon)} \tag{25}$$

3. Sei $S(n)$ ein Projektor auf einen beliebigen Unterraum von $H^{\otimes n}$ mit Dimension kleiner 2^{nR}. Sei $R < S(\rho)$. Dann gilt für alle $\delta > 0$ und genügend große n

$$\operatorname{tr}\left(S(n) \rho^{\otimes n} \right) \leq \delta \tag{26}$$

Die obigen Aussagen bedeuten in etwa das Folgende:

1. Im Fall, dass die Spur 1 werden würde, entspräche $P(n, \epsilon)$ der Identität auf $\rho^{\otimes n}$. Für großes n wird das beliebig genau angenähert.

2. Für großes n enthält der ϵ-typische Unterraum in etwa $2^{nS(\rho)}$ Elemente.

3. Betrachte einen Projektor, der auf einen Raum abbildet, dessen Dimension kleiner als die des ϵ-typischen Unterraums ist. Dann bildet dieser Projektor für großes n typische Sequenzen auf 0 ab , d.h. $2^{nS(\rho)}$ bildet eine untere Schranke für die Dimension einer "identitätsähnlichen" Abbildung.

3.2.3 Beweis: Satz über typische Unterräume

1) Zu zeigen: tr $(P(n,\epsilon)\,\rho^{\otimes n}) \geq 1-\delta$ Betrachte die linke Seite der zu zeigenden Ungleichung:

$$\operatorname{tr}\left(P(n,\epsilon)\,\rho^{\otimes n}\right) = \sum_{x\in T(n,\epsilon)} p(x_1)\,p(x_2)\cdots p(x_n) \tag{27}$$

Diese entspricht genau der Wahrscheinlichkeit, dass eine beliebige Sequenz ϵ-typisch ist. Nach Definition von ϵ-typisch lässt sich dies formulieren als

$$p\left(\left|\frac{1}{n}\log\left(\frac{1}{p(x_1)\,p(x_2)\cdots p(x_n)}\right) - S(\rho)\right| \leq \epsilon\right) \geq 1-\delta \tag{28}$$

Mit Hilfe der Eigenschaften des Logarithmus und Darstellung von $S(\rho)$ als Erwartungswert erhält man

$$p\left(\left|\sum_{i=1}^{n}\frac{-\log p(x_i)}{n} - E(-\log X)\right| \leq \epsilon\right) \geq 1-\delta \tag{29}$$

was genau dem Gesetz der großen Zahlen mit $\tilde{X} = -\log(X)$ entspricht:

Gesetz der großen Zahlen Seien $X_1, X_2, \ldots$ i.i.d. mit endlichem ersten und zweiten Moment. Dann ist für jedes $\epsilon > 0$

$$\lim_{n\to\infty} p\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E(X)\right| \leq \epsilon\right) = 1 \tag{30}$$

2) Zu zeigen: $(1-\delta)\,2^{n(S(\rho)-\epsilon)} \leq |T(n,\epsilon)| \leq 2^{n(S(\rho)+\epsilon)}$ Aus der Definition von ϵ-typisch gewinnt man eine obere und untere Schranke von $p(x)$:

$$\left|\frac{1}{n}\log\left(\frac{1}{p(x_1)\,p(x_2)\cdots p(x_n)}\right) - S(\rho)\right| \leq \epsilon \tag{31}$$

$$S(\rho) - \epsilon \leq \frac{1}{n}\log\left(\frac{1}{p(x)}\right) \leq S(\rho) + \epsilon \tag{32}$$

$$-n\,(S(\rho)-\epsilon) \geq \log(p(x)) \geq -n\,(S(\rho)+\epsilon) \tag{33}$$

$$2^{-n(S(\rho)-\epsilon)} \geq p(x) \geq 2^{-n(S(\rho)+\epsilon)} \tag{34}$$

Mit diesen Ungleichungen lässt sich die Behauptung zeigen. Durch Aufsummieren aller Wahrscheinlichkeiten von Sequenzen, die innerhalb des ϵ-typischen Unterraums liegen, erhalten wir:

$$1 \geq \sum_{x\in T(n,\epsilon)} p(x) \tag{35}$$

$$\geq \sum_{x\in T(n,\epsilon)} 2^{-n(S(\rho)+\epsilon)} \tag{36}$$

$$= |T(n,\epsilon)|\,2^{-n(S(\rho)+\epsilon)} \tag{37}$$

Daraus folgt eine der beiden zu zeigenden Ungleichungen:

$$|T(n,\epsilon)| \leq 2^{n(S(\rho)+\epsilon)} \tag{38}$$

Analog lässt sich die andere Ungleichung zeigen. Hierzu wenden wir den 1. Teil des Satzes an:

$$\begin{aligned} 1-\delta &\leq \sum_{x\in T(n,\epsilon)} p\left(x\right) && (39)\\ &\leq \sum_{x\in T(n,\epsilon)} 2^{-n(S(\rho)-\epsilon)} && (40)\\ &= \left|T\left(n,\epsilon\right)\right| 2^{-n(S(\rho)-\epsilon)} && (41)\\ \left|T\left(n,\epsilon\right)\right| &\geq \left(1-\delta\right) 2^{n(S(\rho)-\epsilon)} && (42) \end{aligned}$$

3) Zu zeigen: $\left|S\left(n\right)\right| < 2^{nR} < 2^{nS(\rho)} \Rightarrow \mathbf{tr}\left(S\left(n\right)\rho^{\otimes n}\right) \leq \delta$ Hierzu teilen wir die Spur in den typischen Unterraum und den Rest auf, um sie später einzeln abschätzen zu können.

$$\mathrm{tr}\left(S\left(n\right)\rho^{\otimes n}\right) = \mathrm{tr}\left(S\left(n\right)\rho^{\otimes n}P\left(n,\epsilon\right)\right) + \mathrm{tr}\left(S\left(n\right)\rho^{\otimes n}\left(I-P\left(n,\epsilon\right)\right)\right) \quad (43)$$

Da der Projektor $P\left(n,\epsilon\right)$ mit $\rho^{\otimes n}$ kommutiert, gilt:

$$\rho^{\otimes n}P\left(n,\epsilon\right) = P\left(n,\epsilon\right)\rho^{\otimes n}P\left(n,\epsilon\right) \quad (44)$$

Durch die Beschränktheit der Eigenwerte lässt sich eine obere Schranke der Spur angeben, diese geht für $n\to\infty$ gegen 0:

$$\begin{aligned} 0 \leq \mathrm{tr}\left(S\left(n\right)P\left(n,\epsilon\right)\rho^{\otimes n}P\left(n,\epsilon\right)\right) &\leq 2^{nR}2^{-n(S(\rho)+\epsilon)} && (45)\\ &= 2^{-n(S(\rho)-R+\epsilon)} && (46)\\ &\to 0 && (47) \end{aligned}$$

Weiter gilt $S\left(n\right) \leq I$. Da $S\left(n\right)$ und $\rho^{\otimes n}\left(I-P\left(n,\epsilon\right)\right)$ positive Operatoren sind, folgt:

$$\begin{aligned} 0 \leq \mathrm{tr}\left(S\left(n\right)\rho^{\otimes n}\left(I-P\left(n,\epsilon\right)\right)\right) &\leq \mathrm{tr}\left(\rho^{\otimes n}\left(I-P\left(n,\epsilon\right)\right)\right) && (48)\\ &\to 0 && (49) \end{aligned}$$

Denn $I-P\left(n,\epsilon\right)$ ist ein Operator, der auf den zum typischen Unterraum orthogonalen Raum projeziert. Da die entsprechende Spur über $P\left(n,\epsilon\right)$ nach dem ersten Teil des Satzes gegen 1 geht, muss umgekehrt diese Spur gegen 0 gehen.

Damit folgt die Behauptung. □

3.3 Beweis: Satz von Schumacher

Beweis der Existenz, es ist zu zeigen: Wenn $R > S\left(\rho\right)$**, dann** $F\left(\rho,\cdot\right)\to 1$ Hierfür geben wir $\mathcal{C}^n$ und $\mathcal{D}^n$ explizit an. Wähle ein $\epsilon > 0$ so, dass $S\left(\rho\right)+\epsilon \leq R$. Aus dem Theorem über typische Unterräume folgt: Für jedes $\delta > 0$ und genügend große n gilt:

$$\begin{aligned} \mathrm{tr}\left(\rho^{\otimes n}P\left(n,\epsilon\right)\right) &\geq 1-\delta && (50)\\ \left|T\left(n,\epsilon\right)\right| &\leq 2^{nR} && (51) \end{aligned}$$

Sei H_c^n ein 2^{nR}-dimensionaler Hilbertraum, der $T\left(n,\epsilon\right)$ enthält. Komprimiere mit $A_i \equiv \left|0\right\rangle\left\langle i\right|$, mit ortohormaler Basis $\left|i\right\rangle$, die zum typischen Unterraum orthogonal ist, und beliebigem Vektor $\left|0\right\rangle$ aus dem typischen Unterraum. Das Kompressionsschema sei

$$\mathcal{C}^n\left(\sigma\right) = P\left(n,\epsilon\right)\sigma P\left(n,\epsilon\right) + \sum_i A_i\sigma A_i^\dagger \quad (52)$$

Die Dekompression $\mathcal{D}^n$ ist als Identität auf H_c^n definiert. Die Fidelity lässt sich durch $1-2\delta$ nach unten abschätzen. Da δ beliebig klein gewählt werden kann, konvergiert die Fidelity somit gegen 1:

$$\begin{aligned} F\left(\rho^{\otimes n}, \mathcal{D}^n \circ \mathcal{C}^n\right) &= \left|\mathrm{tr}\left(\rho^{\otimes n} P\left(n,\epsilon\right)\right)\right|^2 + \sum_i \left|\mathrm{tr}\left(\rho^{\otimes n} A_i\right)\right|^2 && (53)\\ &\geq \left|\mathrm{tr}\left(\rho^{\otimes n} P\left(n,\epsilon\right)\right)\right|^2 && (54)\\ &\geq \left|1-\delta\right|^2 \geq 1-2\delta && (55)\end{aligned}$$

Beweis der Optimalität, es ist zu zeigen: Wenn $R < S(\rho)$, dann $F(\rho,\cdot) \to 0$ Im Folgenden haben die Projektoren in gewisse Unterräume von $H^{\otimes n}$ die gleiche Bezeichnung wie die Unterräume selbst. OBdA bildet $\mathcal{C}^n$ in einen 2^{nR}-dim Unterraum $S(n)$ von $H^{\otimes n}$ ab. Seien C_j die Operatorelemente von $\mathcal{C}^n$ und D_k die von $\mathcal{D}^n$. Dann lässt sich die Fidelity folgendermaßen schreiben:

$$F\left(\rho^{\otimes n}, \mathcal{D}^n \circ \mathcal{C}^n\right) = \sum_{jk} \left|\mathrm{tr}\left(D_k C_j \rho^{\otimes n}\right)\right|^2 \quad (56)$$

Jeder der Operatoren C_j bildet in den Unterraum von $S(n)$ ab, also

$$C_j = S(n)\, C_j \quad (57)$$

Sei $S^k(n)$ der Projektor auf den Unterraum, auf den D_k den Raum $S(n)$ abbildet. Dann gilt:

$$S^k(n)\, D_k S(n) = D_k S(n) \quad (58)$$

$D_k C_j$ kann somit auch durch einen Projektor ergänzt werden:

$$\begin{aligned} D_k C_j &= D_k S(n)\, C_j && (59)\\ &= S^k(n)\, D_k S(n)\, C_j && (60)\\ &= S^k(n)\, D_k C_j && (61)\end{aligned}$$

Damit ergibt sich (mit Cauchy-Schwarz-Ungleichung) für die Fidelity die Abschätzung

$$\begin{aligned} F\left(\rho^{\otimes n}, \mathcal{D}^n \circ \mathcal{C}^n\right) &= \sum_{jk} \left|\mathrm{tr}\left(D_k C_j \rho^{\otimes n} S^k(n)\right)\right|^2 && (62)\\ &\leq \sum_{jk} \mathrm{tr}\left(D_k C_j \rho^{\otimes n} C_j^\dagger D_k^\dagger\right) \mathrm{tr}\left(S^k(n)\, \rho^{\otimes n} S^k(n)\right) && (63)\end{aligned}$$

Da die Dimension von $S^k(n)$ nach Voraussetzung kleiner ist als die Dimension von $S(\rho)$, lässt sich $\mathrm{tr}\left(S^k(n)\, \rho^{\otimes n} S^k(n)\right)$ nach dem Satz über typische Unterraume mit δ abschätzen. Da C_j, D_k auf $\rho^{\otimes n}$ spurerhaltend sind, gilt damit

$$\begin{aligned} F\left(\rho^{\otimes n}, \mathcal{D}^n \circ \mathcal{C}^n\right) &\leq \delta \sum_{jk} \mathrm{tr}\left(D_k C_j \rho^{\otimes n} C_j^\dagger D_k^\dagger\right) && (64)\\ &= \delta && (65)\end{aligned}$$

Somit wird die Fidelity für genügend lange Sequenzen beliebig klein. Dies zeigt die Optimalität. □

3.4 Beispiel: Schumacher-Kompression

In dem vorigen Beweis wurde konstruktiv gezeigt, wie eine Kompression aussehen kann. Dabei handelte es sich um die sogenannte Schumacher-Kompression, die nun an einem Beispiel veranschaulicht werden soll.

Gegeben sei eine i.i.d. Quantenquelle, die die Zustände

$$|\psi_0\rangle = |0\rangle \tag{66}$$
$$|\psi_1\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right) \tag{67}$$

jeweils mit Wahrscheinlichkeit $\frac{1}{2}$ emittiert. Dies entspricht der Dichtematrix

$$\rho = \frac{1}{2}|\psi_0\rangle\langle\psi_0| + \frac{1}{2}|\psi_1\rangle\langle\psi_1| \tag{68}$$
$$= \frac{1}{4}\begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix} \tag{69}$$

die die folgende orthonormale Dekomposition besitzt:

$$\rho = p\left|\overline{0}\right\rangle\left\langle\overline{0}\right| + (1-p)\left|\overline{1}\right\rangle\left\langle\overline{1}\right| \tag{70}$$
$$\left|\overline{0}\right\rangle = \cos\frac{\pi}{8}|0\rangle + \sin\frac{\pi}{8}|1\rangle \tag{71}$$
$$\left|\overline{1}\right\rangle = -\sin\frac{\pi}{8}|0\rangle + \cos\frac{\pi}{8}|1\rangle \tag{72}$$
$$p = \frac{1}{4}\left(3 + \tan\frac{\pi}{8}\right) \approx 0.85 \tag{73}$$

Bezüglich dieser Basis lässt sich ein Block aus n qubits auch als Linarkombination von Eigenzustandssequenzen schreiben:

$$\sum_{X=\left\{\overline{00}\ldots\overline{0},\overline{0}\ldots\overline{01},\ldots,\overline{11}\ldots\overline{1}\right\}} C_X|X\rangle \tag{74}$$

Da nur typische Sequenzen häufig auftreten, reicht es aus, die typischen Sequenzen $|X\rangle$ zu übertragen. Diese haben die Eigenschaft, dass ihr Hamminggewicht[5] etwa $n(1-p)$ ist. Da $p \gg (1-p)$ gilt, müssen effektiv nur die Sequenzen $|X\rangle$ berücksichtigt werden, die viele $\overline{0}$en enthalten.

Wähle z.B. $n = 3$. Sortiere alle möglichen Ausgaben der Quelle aufsteigend nach Hamminggewicht und ordne ihnen eine Kodierung zu:

$$\begin{array}{ll} 000 \to 000 & 011 \to 100 \\ 001 \to 001 & 101 \to 101 \\ 010 \to 010 & 110 \to 110 \\ 100 \to 011 & 111 \to 111 \end{array} \tag{75}$$

Das erwartete Hamminggewicht liegt hier bei $(1-p)\,3 \approx 0.44$, d.h. wir sollten die Sequenzen berücksichtigen, die ca. 0.44 Einsen besitzen, also die Sequenzen ohne oder mit einer Eins. Diese Aussage wird dadurch bestätigt, dass die erwartete komprimierte Codelänge $3 \cdot S(\rho) =$

[5] Das Hamminggewicht einer Sequenz ist die Anzahl der in ihr enthaltenen 1en.

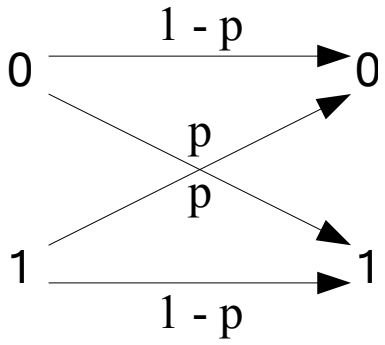

Abbildung 3: Klassischer binärer symmetrisch verrauschter Kanal

$3\left(-p\cdot\log p-(1-p)\cdot\log\left(1-p\right)\right)\approx 1.8$ in etwa bei 2 liegt und somit folgende Zuordnung ausreicht, um zuverlässig zu sein:

$$\begin{array}{ll} 000\rightarrow 00 & 011\rightarrow 11 \\ 001\rightarrow 01 & 101\rightarrow 11 \\ 010\rightarrow 10 & 110\rightarrow 11 \\ 100\rightarrow 11 & 111\rightarrow 11 \end{array} \tag{76}$$

Offensichtlich ist der gesamte rechte Block nicht wieder dekomprimierbar, was aber aufgrund der "geringen" Auftrittswarscheinlichkeit dieser Sequenzen keine Rolle spielt. Das Verfahren wird bei wachsendem n wesentlich besser. Im Grenzwert $n\rightarrow\infty$ wird die Fehlerwahrscheinlichkeit 0.

Der hier vorgestellte Algorithmus ist nicht optimal, veranschaulicht aber den Beweis von Schumachers Theorems sehr schön. Eine optimale Variante wäre das "quantum arithmetic coding".

4 Kodierung unter Rauschen

Bisher haben wir den Kanal immer als Ideal, sozusagen als identische Abbildung zwischen Ein- und Ausgang betrachtet. Dies wollen wir nun fallen lassen und es dem Kanal gestatten stochastische Änderungen an dem Signal vorzunehmen. Ein solcher Kanal wird als *verrauschter Kanal* bezeichnet. Zudem soll der Kanal *gedächnislos* sein, sich also bei jeder Benutzung gleich verhalten!

4.1 Klassische Kodierung unter Rauschen

Einen verrauschten Kanal beschreiben wir im klassischen Fall durch einen Satz von bedingten Wahrscheinlichkeiten $p\left(x|y\right)\geq 0$.

Bei einem binären symmetrisch verrauschten Kanal wird ein Signal am Eingang (0 oder 1) wird mit der Wahrscheinlichkeit $1-p$ unverändert durchgelassen oder mit der Wahrscheinlichkeit p invertiert $(0\mapsto 1, 1\mapsto 0)$. Siehe hierzu auch Abbildung 3.

4.1.1 Shannon: verrauschte Kanalkodierung

Für einen verrauschten Kanal $\mathcal{N}$ ist die Kapazität gegeben durch

$$\mathcal{C}\left(\mathcal{X}\right)=\max_{p(x)}H\left(X:Y\right) \tag{77}$$

wobei das Maximum über alle Eingangsverteilungen $p(x)$ von X gebildet wird und Y die zugehörige induzierte Zufallsvariable am Ausgang des Kanals ist.

4.1.2 Beispiel: Binärer Kanal

Der Satz von Shannon über verrauschte Kanalkodierung lässt sich im Fall eines klassischen binären verrauschten Kanals anwenden:

$$\begin{aligned} \mathcal{C}(\mathcal{X}) &= \max_{p(x)} H(X:Y) && (78)\\ &= \max_{p(x)} \left(H(Y) - H(Y|X)\right) && (79)\\ &= \max_{p(x)} \left(H(Y) - \sum_x p(x) H(Y|X=x)\right) && (80)\\ &= 1 - H_{bin}(p) && (81) \end{aligned}$$

Dabei ist $H_{bin}(p)$ die binäre Entropie eines Zweizustandssystems. Man kann diese Formel auch folgendermaßen auffassen: Für den Fall $p = 0$, also den unverrauschten Kanal, wird die Kapazität 1. Gleiches gilt auch für $p = 1$, da in diesem Fall der Kanal *immer* invertiert und somit am Ausgang problemlos auf das Eingangssignal geschlossen werden kann. Ansonsten wird von der Kapazität genau die Entropie des Rauschens abgezogen, was bei $p = \frac{1}{2}$ dazu führt, dass die Kapazität 0 ist.

Für eine Quelle lässt sich eine sinnvolle Kodierung wie in Abbildung 4 auf der nächsten Seite konstruieren. Jedes Codewort ist ein Punkt im Raum aller möglichen Codewörter. Der gestörte Kanal kann einige der Bits kippen, was zu einem benachbarten Codewort im Codewortraum führt. Wenn wir nun bei der Konstruktion der Kodierung die *typischen Fehlersequenzen* als Punkte von Kugeln um das zu sendende Codewörter interpretieren und dafür sorgen, dass sich diese Kugeln nicht überlappen, haben wir eine *zuverlässige* Kodierung gefunden. Dieso ist umso effizienter, je weniger Raum zwischen den Kugeln ungenutzt bleibt.

4.1.3 Beweis: Shannon - verrauschte Kanalkodierung

Zu zeigen: Für Raten $R < C$ geht die Fehlerwahrscheinlichkeit gegen 0 Sei $p(x)$ Verteilung, mit der wir ein Wörterbuch generieren. Dieses Wörterbuch ist sowohl Sender und als auch Empfänger bekannt und hat die Gestalt

$$\mathcal{C} = \begin{bmatrix} x_1(1) & x_2(1) & \dots & x_n(1) \\ \vdots & \vdots & \ddots & \vdots \\ x_1\left(2^{nR}\right) & x_2\left(2^{nR}\right) & \dots & x_n\left(2^{nR}\right) \end{bmatrix} \tag{82}$$

In der i-ten Spalte steht das für die i-te Sequenz zu übermittelnde Codewort.

Die Wahrscheinlichkeit, ein spezielles Wörterbuch zu generieren ist

$$\Pr(\mathcal{C}) = \prod_{\omega=1}^{2^{nR}} \prod_{i=1}^{n} p(x_i(\omega)) \tag{83}$$

OBdA nehmen wir an, dass die Nachrichten gleichverteilt übertragen werden, denn andernfalls können die Nachrichten zunächst komprimiert werden:

$$\Pr(W = \omega) = 2^{-nR}, \quad \omega = 1, 2, \dots, 2^{nR} \tag{84}$$

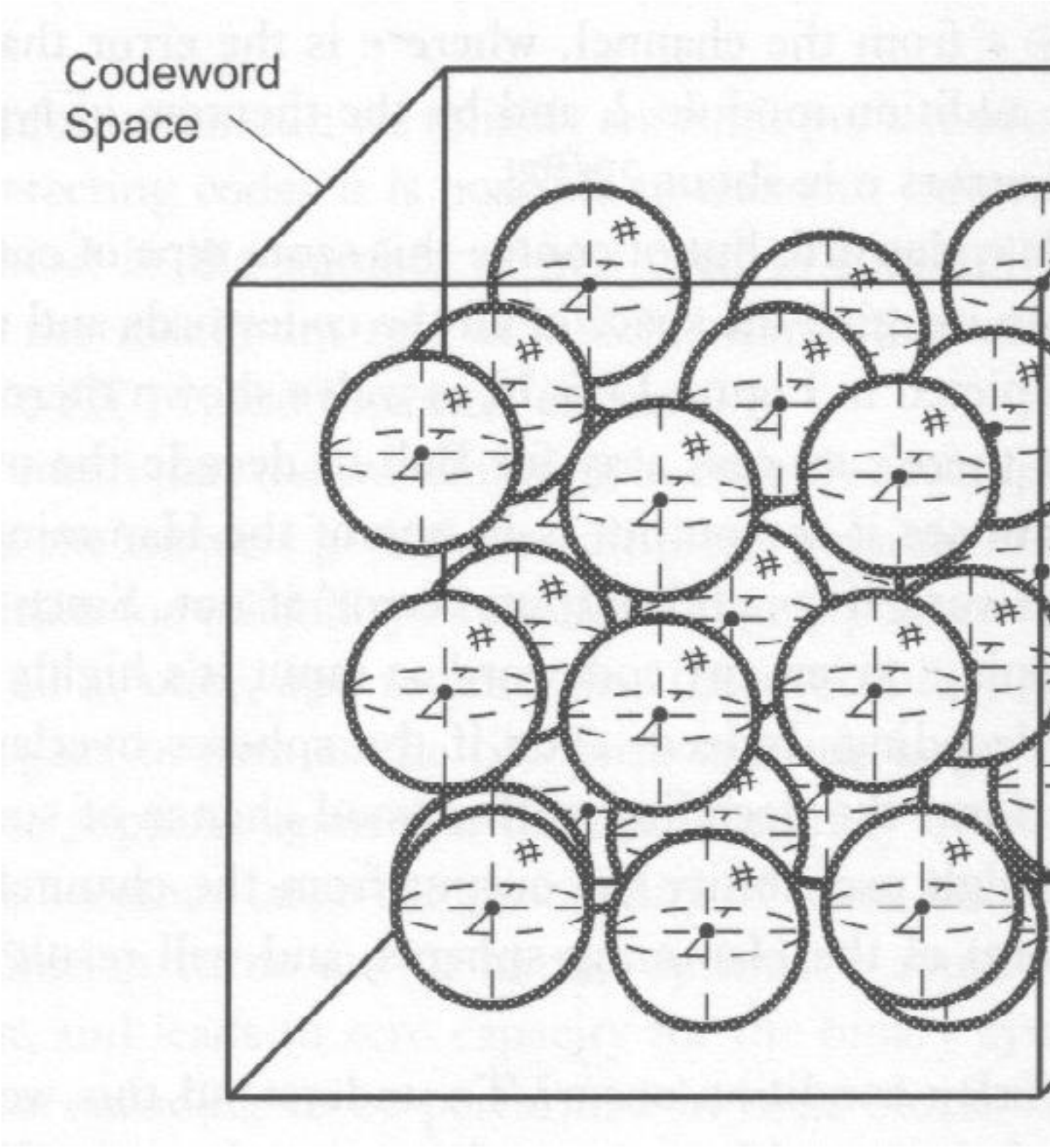

Abbildung 4: Ausnutzung des Coderaums bei einem verrauschten Kanal

Der Empfänger empfängt y^n mit Wahrscheinlichkeit

$$P\left(y^n|x^n\left(\omega\right)\right) = \prod_{i=1}^{n} p\left(y_i|x_i\left(\omega\right)\right) \tag{85}$$

wobei hier die Unabhängigkeit der einzelnen Kanalbenutzungen eingeht.

Beim Dekodieren wird die Nachricht $\hat{W}$ genau dann ausgegeben, wenn

- $\left(X^n\left(\hat{W}\right), Y^n\right)$ gleichzeitig typisch sind
- es keinen anderen Index k gibt, so dass $\left(X^n\left(k\right), Y^n\right)$ gleichzeitig typisch sind

Falls es kein solches $\hat{W}$ gibt, wird ein Standardeintrag ausgegeben. Ein Dekodierfehler, also $W \neq \hat{W}$, soll dem Ereignis $\mathcal{E}$ entsprechen. Unter gleichzeitig typisch versteht man, dass sowohl der erste als auch der zweite Eintrag des Tupels eine typische Sequenz bilden, aber auch die Sequenz der Paare.

Die Fehlerwahrscheinlichkeit $P\left(\varepsilon\right)$ hängt nicht vom Sendewort ab, da kein Sendewort in irgendeiner Art und Weise ausgezeichnet ist.

$$\Pr\left(\mathcal{E}\right) \quad = \quad \Pr\left(\mathcal{E}|W=1\right) \tag{86}$$

Weiter definieren wir uns ein Ereignis $E_i = \{\left(X^n\left(i\right), Y^n\right) | \left(X^n\left(i\right), Y^n\right)$ sind gleichzeitig typisch$\}$. Für genügend große n und $R < H\left(X:Y\right) - 3\epsilon$ gilt für die durchschnittliche Fehlerwahrscheinlichkeit über alle Wörterbücher:

$$\begin{aligned}
\Pr\left(\mathcal{E}|W=1\right) &= P\left(E_1^c \cup E_2 \cup \ldots \cup E_{2^{nR}}\right) && (87)\\
&\leq P\left(E_1^c\right) + \sum_{i=2}^{2^{nR}} P\left(E_i\right) && (88)\\
&\leq \epsilon + \sum_{i=2}^{2^{nR}} 2^{-n(H(X:Y)-3\epsilon)} && (89)\\
&\leq 2\epsilon && (90)
\end{aligned}$$

Hier fließen Sätze über gleichzeitig typische Sequenzen analog zu dem obigen Satz über typische Sequenzen ein.

Nun gibt es mindestens ein Wörterbuch $\mathcal{C}^*$ mit einem durchschnittlichen Fehler $P_e^n\left(\mathcal{C}^*\right) \leq 2\varepsilon$. Die durchschnittliche Fehlerwahrscheinlichkeit in diesem Wörterbuch ist also:

$$P_e^n\left(\mathcal{C}^*\right) = \frac{1}{2^{nR}} \sum_i \lambda_i\left(\mathcal{C}^*\right) \leq 2\epsilon \tag{91}$$

wobei $\lambda_i\left(\mathcal{C}^*\right)$ die Fehlerwahrscheinlichkeit der Erkennung der i-ten Sequenz im Wörterbuch $\mathcal{C}^*$ ist. Nun verwerfen wir in diesem Wörterbuch die schlechtere Hälfte der Zuordnungen, also diejenigen mit der größten Fehlerwahrscheinlichkeit, d.h. wir nutzen sie nicht beim Senden. Das bedeutet für alle übrigen Zuordnungen

$$\lambda_i\left(\mathcal{C}^*\right) \leq 4\epsilon \tag{92}$$

Denn sonst würden sie bereits zu einem höheren durchschnittlichen Fehler führen. Somit haben wir nur noch 2^{nR-1} Wörterbucheinträge, womit die Rate von R auf $R - \frac{1}{n}$ gesunken ist, aber jedes einzelne Codewort mit einer beliebig hohen Genauigkeit erkannt werden kann. Insgesamt haben wir einen Code mit $R' = R - \frac{1}{n}$ und Fehlerwahrscheinlichkeit $\lambda^{(n)} \leq 4\epsilon$ konstruiert. Für große n ist also gezeigt, dass man mit der im Satz behaupteten Rate Daten übertragen kann. In der Praxis wird ein solcher Zufallscode jedoch nicht das Mittel der Wahl sein!

Zu zeigen: Für eine gegen 0 gehende Fehlerwahrscheinlichkeit gilt $R > C$ OBdA senden wir nur typische Sequenzen, also entspricht R der Entropie der Sendesequenzen

$$\begin{aligned} nR &= H(W) = H(W|Y^n) + H(W:Y^n) && (93)\\ &\leq H(W|Y^n) + H(X^n(W)|Y^n) && (94)\end{aligned}$$

Durch Anwenden von Fanos Ungleichung erhalten wir

$$\begin{aligned} nR &\leq 1 + P_e^{(n)} nR + H(X^n(W)|Y^n) && (95)\\ &\leq 1 + P_e^{(n)} nR + nC && (96)\end{aligned}$$

$P_e^{(n)}$ ist die Fehlerwahrscheinlichkeit für eine Sequenz der Länge n. Dies führt zur Behauptung:

$$P_e^{(n)} \geq 1 - \frac{1}{nR} - \frac{C}{R} \geq 0 \Rightarrow R > C \quad (97)$$

□

4.2 Quantenmechanisch

Einen verrauschten Kanal beschreiben wir im quantenmechanischen Fall durch eine spurerhaltende Quantenoperation ε, die auf den Ausgaben der Quelle ρ wirkt. Hier soll nun kein Zustand komprimiert, sondern nach dem Einwirken von ε wiedererkannt werden. Der Sender kodiert hierzu seine Information "semiklassisch" insofern, als dass er eine Sequenz von (klassischen) Symbolen $M = \{i\}$ durch eine Sequenz von (quantenmechanischen) Dichteoperatoren $\rho_M = \{\rho_i\}$ übermittelt. Der Empfänger muss seine Fehlerkorrekturmaßnahmen zudem *nicht* auf einzelnen Zuständen durchführen, sondern kann auch auf dem Produktzustand, also der kompletten übermittelten Sequenz

$$\rho_M = \rho_{M_1} \otimes \rho_{M_2} \otimes \ldots \otimes \rho_{M_n} \quad (98)$$

arbeiten.

4.2.1 Theorem Holevo-Schumacher-Westmoreland (HSW)

Sei ε eine spurerhaltende Quantenoperation. Dann ist die Produktzustandskapazität

$$C^{(1)}(\varepsilon) = \max_{\{p_j,\rho_j\}} \left[S\left(\varepsilon\left(\sum_j p_j \rho_j\right)\right) - \sum_j p_j S(\varepsilon(\rho_j)) \right] \quad (99)$$

wobei das Maximum über alle Ensemble $\{p_j, \rho_j\}$ von möglichen Eingangszuständen ρ_j des Kanals gebildet wird.

Diesen Satz werde ich hier nicht beweisen. Bei Interesse sei auf die Literatur verwiesen.

Literatur

[1] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, October 2000.