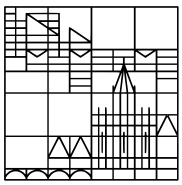


Universität Konstanz
Fachbereich Physik

Quanteninformationstheorie

Sommersemester 2013

Prof. Dr. G. Burkard

TEX: Markus Gruber

Herausgeber: Fachbereich Physik, Universität Konstanz
Fach 627, 78457 Konstanz
Tel. 07531/88-2413
fachbereich.physik@uni-konstanz.de

Stand: 20. Juli 2013

Inhaltsverzeichnis

Vorwort

Dieses Skript entsteht aus der Mitschrift der Vorlesung von Guido Burkard über den Quanteninformationstheorie im Sommersemester 2013. Übertragungsfehler könnten dazu geführt haben, dass vereinzelt Formeln von ihrem wahren Wert abweichen. Deshalb übernehme ich keine Gewähr für die Richtigkeit.
Veröffentlichung und Weitergabe nur mit Zustimmung der Dozenten und Autoren. Über Rückmeldungen zu Fehlern/Anregungen/Kritik würden wir uns freuen.

Markus Gruber, markus.gruber@uni-konstanz.de

1 Einleitung

1.1 Fragestellungen

1) Was ist Information?

2) Welches sind die Gesetze, denen Information unterworfen ist?

3) Anwendungen ("Theorie")
   (a) Übertragung/Speicherung von Information (Kommunikation) → Datenkompression, Fehlerkorrektur, Kryptologie
   (b) Verarbeitung von Information (Rechnen) → Algorithmen, Komplexitätstheorie, Fehlerkorrektur

Wieso ist dieses Thema interessant für Physiker? Die Antworten auf diese Fragen hängen von den geltenden physikalischen Gesetzen ab.
Rolf Landauer sagte "Information is physical", z.B. Landauer-Prinzip: Löschen eines Bits gibt die Wärme $Q \geq k_B T \ln 2$ frei.
In der traditionellen Behandlung von 1)-3) in der Informatik und E-Technik werden die Gesetze der klassischen Physik zu Grund gelegt. Dies ist ausreichend für die Situationen im Alltag und erhältliche technische Geräte (Hardware).
Aber: die fundamentalere Beschreibung ist die Quantenmechanik:

- mikroskopische/atomare Systeme, empirisch überprüft
- mesoskopische Systeme: quantenmechanische Effekte (Interferenz) auf der Mikrometer-Skala meist bei tiefen Temperaturen ($\lesssim 1\,\mathrm{K}$). Zum Beispiel:
  - Supraleitende elektrische Schaltkreise ($\approx$ µm)
  - Quantendots $\approx 100\,\mathrm{nm} - 1\,\mu\mathrm{m}$, Elektronen-Spins
- makroskopische Systeme: Gibt es eine Grenze für die Anwendbarkeit der Quantenmechanik? Diese Frage ist noch offen. Bislang gibt es jedoch keine Evidenz dafür. Falls ja, müsste "neue Physik" ins Spiel kommen, falls nein wäre die Quanteninformationstheorie universell anwendbar und Quantenmaschinen wären im Prinzip skalierbar.

Die Skizze in Abbildung 1.1 nach Benett (IBM) veranschaulicht das Gebiet der Quanteninformationstheorie.
Fragen dieser Vorlesung:

1') Was ist Quanteninformation?

2') Welchen Gesetzen unterliegt Quanteninformation? (Welchen Gesetzen unterliegt Information, wenn quantenmechanische Regeln gelten?) Insbesondere gilt 2)$\neq$ 2')?

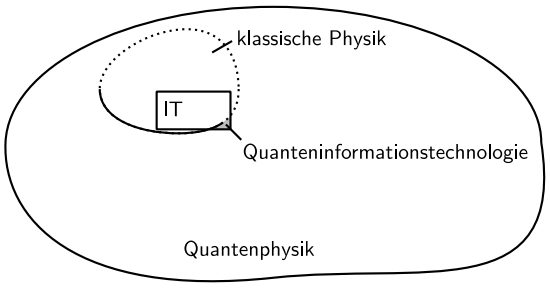


Abbildung 1.1: Quanteninformationstheorie nach Bennett (IBM)

3') (a) Übertragung von Quanteninformation → Quantenkommunikation
(b) Verarbeitung von Quanteninformation → Quantencomputer

Bemerkung:

- Diese Vorlesung könnte auch *Quanteninformation und Quantencomputing* heißen.
- Quanteninformation bezeichnet oft nur 2') und 3'a), nicht 3'b).

Worum es in der Vorlesung geht:

- Einführung in die Quanteninformationstheorie
- wichtigste Grundlagen / Verständnis
- Formalismus
- Beispiele

Worum es hier *nicht* geht:

- Grundlagen der Quantenmechanik (bis auf wenige Ausnahmen)
- Interpretation der Quantenmechanik
- kompletter Überblick der Quanteninformationstheorie
- physikalische Realisierungen (bis auf wenige Ausnahmen)

1.2 Klassische Informationstheorie

Beispiel: Kaffee bestellen (vereinfacht). Es gibt folgende Wahlmöglichkeiten:

- groß oder klein,
- heiß oder Eis,
- mit oder ohne Koffein.

Dies sind insgesamt $2^3 = 8$ Kombinationsmöglichkeiten und drei unabhängige Einheiten von Information, Antworten auf ja/nein. Der Informationsgehalt der Bestellung beträgt 3 Bits. *Bit* ist die englische Kurzform für *binary digit* und hat die Werte 0 oder 1 (Veröffentlicht von C. Shannon 1948, erfunden von John Tukey).

Bemerkung:

1) Der Informationsgehalt stimmt nur, wenn der Kellner keine „Erwartung“ hat (a priori Wahrscheinlichkeiten $p = 1/2$)

2) Unabhängigkeit $I = I_{\text{Größe}} + I_{\text{Temperatur}} + I_{\text{Koffein}} = 1 + 1 + 1 = 3$

Gegenbeispiel zu 1): Der Kellner kennt seinen Kunden, d.h. er weiß, was er/sie *immer* bestellt. Dann ist der Informationsgehalt $I = 0$. Im Allgemeinen ist die Wahrscheinlichkeit $0 \leq p \leq 1$, z.B. weiß der Kellner $p_{\text{groß}} = 0.5$, $p_{\text{heiß}} = 0.9$ und $p_{\text{Koffein}} = 0.8$.
Wie kann man die Information messen? Wir wünschen $I_{\text{groß}} = 1$ (in Bits), $0 < I_{\text{heiß}}, I_{\text{Koffein}} < 1$. Wie sieht die Funktion $I(p)$ aus?
Forderungen:

(i) $I(p) \geq 0$

(ii) $I(1) = 0$

(iii) Für unabhängige Ereignisse mit Wahrscheinlichkeiten p_1 und p_2 gilt $p = p_1 p_2$. Damit $I(p) = I(p_1 p_2) = I(p_1) + I(p_2)$ (Additivität)

(iv) $I(p)$ sei stetig (und differenzierbar zur Vereinfachung)

Daraus ergeben sich folgende Beziehungen:

$$\begin{aligned} I(p^2) &= 2I(p) \text{ nach (iii)} \\ I(p^n) &= nI(p) \\ I(p) &= I((p^{1/m})^m) = mI(p^{1/m}) \\ \Rightarrow I(p^{n/m}) &= \frac{n}{m} I(p). \end{aligned}$$

Aufgrund der Stetigkeit (iv) folgt $I(p^a) = aI(p)$ für $0 \leq p \leq 1$ und $a > 0$. Ableiten dieser Beziehung nach a ergibt

$$I'(p^a) p^a \ln p = I(p).$$

Für $a = 1$ erhält man die Bedingung

$$\begin{aligned} I'(p) p \ln p &= I(p) \\ \Leftrightarrow \frac{I'(p)}{I(p)} &= \frac{\mathrm{d}}{\mathrm{d}p} \ln I(p) = \frac{1}{p \ln p} = -\frac{1}{p|\ln p|}. \end{aligned}$$

Integration auf beiden Seiten liefert

$$\begin{aligned} \ln I(p) &= \ln(|\ln(p)|) + \text{const} \\ \Leftrightarrow I(p) &= -d \ln p. \end{aligned}$$

Wähle nun d so, dass

$$I(p) = -\log_2 p = -\log p. \tag{1.1}$$

Dabei bezeichnet log immer den Logarithmus zur Basis 2.

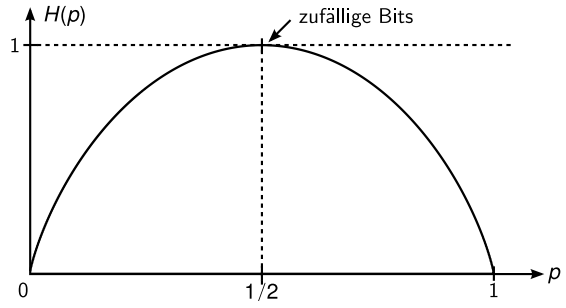


Abbildung 1.2: Shannon-Entropie für ein Bit als Funktion der Wahrscheinlichkeit p, dass $x = 1$

Bemerkung:

(i) $I(p)$ heißt *Informationsgehalt* oder *Überraschungswert* einer Nachricht (Shannon).

(ii) $I(p) > 1$, falls $p < 1/2$.

(iii) Aber: mittlerer Informationsgehalt $\langle I \rangle$ ist beschränkt, z.B. ist für Bits $0 \leq \langle I \rangle \leq 1$. Konkret ist

$$H(p) = \langle I(p) \rangle = \sum_x p(x) I(p(x)) = -\sum_x p(x) \log p(x)$$

die *Shannon-Entropie* (s.u.).
Für Bits gilt $x \in \{0, 1\}$, d.h. $H(p) = -p(0)\log(p(0)) - p(1)\log(p(1))$. Wegen $p(0) + p(1) = 1$ definiere $p := p(0)$ und erhalte

$$H(p) = -p \log p - (1-p)\log(1-p).$$

Es gilt $H(1-p) = H(p)$, d.h. H ist symmetrisch bezüglich $p = 1/2$ und ein Plot findet sich in Abbildung 1.2.

Beispiel: Kaffee-Bestellung
Bemerkung:

1) Der Informationsgehalt hängt von der Bestellung ab.

2) Der Informationsgehalt kann größer als 3 Bits sein.

Für die Bestellung „großer, heißer, entkoffeinierter Kaffe“ gilt

$$\begin{aligned} I &= I(p_{\text{groß}}) + I(p_{\text{heiß}}) + I(1 - p_{\text{Koffein}}) \\ &= 1 + \log(0.9) + \log(0.2) = 3.47. \end{aligned}$$

Im Mittel gilt

$$\begin{aligned} H &= H(p_{\text{groß}}) + H(p_{\text{heiß}}) + H(p_{\text{Koffein}}) \\ &= 1 + 0.47 + 0.72 = 2.19 < 3 \end{aligned}$$

Zwei Theoreme von Shannon (1948)

(I) Wie stark kann Information komprimiert werden? → *noiseless coding theorem*

(II) Mit welcher Rate kann man durch einen fehlerhaften Kanal verlässlich kommunizieren? → *noisy channel coding theorem*

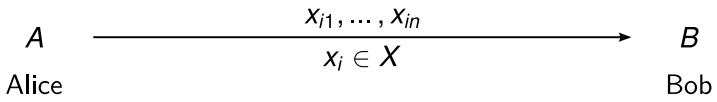
Shannon-Theorem I

(i)

Definition: Ein Alphabet X ist eine endliche Menge $X = \{x_1, \ldots, x_K\}$, wobei die Anzahl der Elemente $K > 1$ ist. Z.B. Bits: $X = \{0, 1\}$, $K = 2$.

(ii)

Definition: Eine *Nachricht* (*Botschaft*) der Länge n ist eine Sequenz von n Symbolen aus einem Alphabet X.



Zweck ist die Übertragung oder Speicherung von Information.

(iii) Die Symbole in der Nachricht seinen *i.i.d.* (*independently and identically distributed*, also unabhängig und identisch verteilt), d.h. für die gemeinsame Wahrscheinlichkeitsdichte gilt

$$p(x_{i1}, \ldots, x_{in}) = \prod_{k=1}^{n} p(x_{ik})$$

mit der Wahrscheinlichkeitsverteilung $p : X \to [0, 1]$, $\sum_x p(x) = 1$.

(iv) betrachte den Grenzfall $n \to \infty$ (Asymptotik)

(v) Typische Botschaften der Länge n:

- Es gibt insgesamt K^n Botschaften der Länge n.
- Nur typische Botschaften haben eine endliche Wahrscheinlichkeit für $n \to \infty$.
- In einer typischen Botschaft kommt das Symbol x $p(x)n$-mal vor (gilt für alle x). Die Häufigkeit $p(x)n$ kann gerundet werden, der Rundungsfehler geht gegen 0 für $n \to \infty$.
- Nicht typische Botschaften sind unwahrscheinlich ($p \to 0$ für $n \to \infty$).
- Typische Botschaften unterscheiden sich nur durch die *Reihenfolge* der Symbole.
- Wie viele typische Botschaften gibt es?
  Anzahl von Permutationen von n Objekten mit Wiederholungen $p(x_i)n$, $i = 1, 2, \ldots, K$ ($\to$Kombinatorik) ist gegeben durch

$$N = \frac{n!}{(p(x_1)n)! \cdots (p(x_K)n)!} = \frac{n!}{\prod_{i=1}^{K}(p(x_i)n)!}.$$

  Mit der Stirling-Approximation des Logarithmus

$$\ln n! \approx n \ln n - n + \mathcal{O}(\ln n)$$

  und $\log x = \log_2 x = (\log_2 e) \ln x$ folgt

$$\log n! \approx \log n - n \log e$$

und daraus

$$\begin{aligned}
\log N &= \log n! - \sum_{i=1}^{K} \log((p(x_i)n)!) \\
&\approx n\log n - n\log e - \sum_{i=1}^{K} p(x_i)n\log(p(x_i)n) + \sum_{i=1}^{K} np(x_i)\log e \\
&= n\log n - n\log e - n\sum_{i=1}^{K} p(x_i)(\log p(x_i) + \log n) + n\log e \underbrace{\sum_{i=1}^{K} p(x_i)}_{=1} \\
&= n\log n - n\log n\sum_{i=1}^{K} p(x_i) - n\sum_{i=1}^{K} p(x_i)\log p(x_i) \\
&= -n\sum_{i=1}^{K} p(x_i)\log p(x_i) \\
&= nH(p)
\end{aligned}$$

mit der *Shannon-Entropie*

$$H(p) := -\sum_{i=1}^{K} p(x_i)\log p(x_i).$$

Damit ist die Anzahl der typischen Botschaften $N \approx 2^{nH(p)}$.

Komprimierung:

- Block-Code (Blocks der Länge $n \gg 1$)
- benötige $nH(p)\log K$ Bits
- mit Bits ($K = 2$) benötigt man $nH(p)$ Bits
- $nH(p) \leq n$
- falls $H(p) < n$ spricht man von *Kompression*.

Beispiel: $X = \{0,1\}, p = p(0), 1-p = p(1)$, $H(p) = p\log p - (1-p)\log(1-p)$, siehe Abbildung 1.2. Für $p \neq 1/2$ ist eine Kompression möglich.

Satz (Shannon-Theorem I (noiseless coding))**.** *Eine Botschaft (in einem Alphabet X mit K Symbolen) kann von $n\log K$ Bits auf $nH(p)\log K$ Bits komprimiert werden (für Bits, $K = 2$, von n auf $nH(p)$ Bits), wobei $H(p) = \sum_{x\in X} p(x)\log p(x)$ die Shannon-Entropie ist.*

Korollar. *Zufällige Bits ($p = 1/2$) können nicht komprimiert werden*

Bemerkung: Die Shannon-Entropie ist der Erwartungswert der übertragenen Information: $H(p) = \langle -\log p\rangle$.
Beispiel: (i) $p(x_1) = 1$, $p(x_{i\neq 1}) = 0$, daraus folgt $H(p) = 0$

(ii) $K = 2$ (Bits): $H(p) = -p \log p - (1-p)\log(1-p)$

Eigenschaften von $H(p)$:

(i) Seien $p_1, p_2 : X \to [0,1]$ Wahrscheinlichkeitsverteilungen über X. Definiert man eine neue Wahrscheinlichkeitsverteilung p durch „mischen“

$$p(x) := wp_1(x) + (1-w)p_2(x),$$

mit $0 \leq w \leq 1$, so gilt

$$H(p) = H(wp_1 + (1-w)p_2) \geq wH(p_1) + (1-w)H(p_2).$$

Diese Eigenschaft bezeichnet man als *Konkavität*.

(ii) Zu $p : X \times Y \to [0,1], (x,y) \mapsto p(x,y)$ kann man die *Randverteilungen*

$$p_x : X \to [0,1] \quad x \mapsto \sum_{y \in Y} p(x,y)$$
$$p_y : Y \to [0,1] \quad y \mapsto \sum_{x \in X} p(x,y)$$

definieren. Für diese gilt

$$H(p) \leq H(p_x) + H(p_y).$$

Gleichheit gilt genau dann, wenn $p(x,y) = p(x)p(y)$ also keine Korrelation vorliegt. Die obige Ungleichung wird *Subadditivität* genannt.

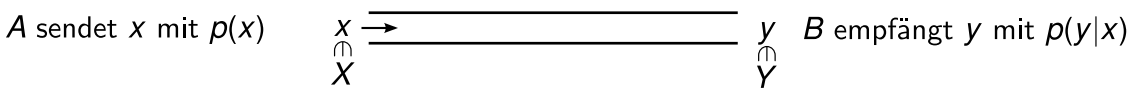
(iii) Sei $p : X \times Y \to [0,1], (x,y) \mapsto p(x,y)$. Dann gilt auch

$$H(p) \geq H(p_x) \text{ und } H(p) \geq H(p_y).$$

Dies ist intuitiv klar, denn nach dieser Ungleichung enthält die gesamte Botschaft mindestens soviel Information wie jedes ihrer Teile. Allerdings stimmt diese *nicht* in der Quanteninformationstheorie! Der Grund dafür ist Verschränkung (vgl. Kapitel **??**).

Shannon-Theorem II

Wir betrachten Kanäle mit Rauschen



Die *bedingte Wahrscheinlichkeit* $p(y|x)$ ist eine Eigenschaft des Kanals.

Bemerkung:

- Beide Seiten können das selbe Alphabet verwenden ($X = Y$), müssen es aber nicht zwingend.
- Der Kanal wird charakterisiert durch $p(y|x)$.

Beispiel: Für Bits ($X = Y = \{0,1\}$) betrachten wir die *stochastische Matrix* für einen symmetrischen binären Kanal:

		x	
	$p(y\|x)$	0	1
y	0	$1-\varepsilon$	ε
	1	ε	$1-\varepsilon$

Es gilt $\sum_y p(y|x) = 1$. Für $\varepsilon = 0, 1$ ist der Kanal perfekt, für $\varepsilon = 1/2$ gibt es nur Rauschen.
Frage: Wieviel Information kann man (pro Benutzung des Kanals) fehlerfrei (für $n \to \infty$) transportieren? Die Antwort liefert die Kapazität des Kanals und das Shannon-Theorem II.
Aufgabe für B: bestimme x aus y, d.h. die Wahrscheinlichkeit $p(x|y)$. Benutze dazu $p(x,y) = p(y|x)p_x(x) = p(x|y)p_y(y)$. Daraus erhält man

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(y|x)p(x)}{\sum_{x'} p(y|x')p(x')},$$

wobei für das zweite Gleichheitszeichen die Beziehung $p(y) = \sum_{x'} p(x', y)$ verwendet wurde. Diese Beziehung nennt man *Satz von Bayes* .
Gesucht ist der Informationsgewinn von B pro gesendetem Symbol. Dazu betrachten wir die Bilanz zwischen Wissen von B vor und nach der Übertragung eines Symbols. Vor der Übertragung kennt B nur die a-priori-Wahrscheinlichkeit $p(x)$ für die Übertragung des Symbols x. Nach dem Empfang von y ist das Wissen von B über x gegeben durch $p(x|y)$. Die fehlende Information pro Symbol ist gegeben durch die Shannon-Entropie.

$$\text{vorher:} \quad H(X) = -\sum_x p(x) \log p(x)$$

$$\text{nachher:} \quad H(X|Y) := \langle -\log p(x|y) \rangle = -\sum_{x,y} p(x,y) \log p(x|y)$$

$H(X|Y)$ heißt *bedingte Shannon-Entropie* oder *Äquivokation*.
Im Mittel beträgt der Informationsgewinn nach dem Empfang von y

$$I(X;Y) := H(X) - H(X|Y).$$

$I(X;Y)$ heißt *gegenseitige Information* oder *Transinformation*, auf Englisch *mutual information*.
Mit Hilfe von $p(x|y) = p(x,y)/p_y(y)$ kann man $H(X|Y)$ konkret ausrechnen:

$$H(X|Y) = -\sum_{x,y} p(x,y) \log p(x|y) = -\sum_{x,y} p(x,y) \log p(x,y) + \sum_y \underbrace{\sum_x p(x,y)}_{p_y(y)} \log p_y(y)$$

$$= H(X,Y) - H(Y).$$

Analog erhält man $H(Y|X) = H(X,Y) - H(X)$. Damit ergibt sich für die gegenseitige Information

$$I(X;Y) = \underbrace{-H(X|Y)}_{\text{Information nachher}} - \big(\underbrace{-H(X)}_{\text{Information vorher}} \big)$$
$$= H(X) + H(Y) - H(X,Y)$$
$$= H(Y) - H(Y|X) = I(Y;X) \geq 0 \quad \text{(Subadditivität)}.$$

Betrachtung der Extremfälle:

(i) „nur Rauschen“, y ist unabhängig von x, d.h. $p(y|x) = p_y(y)$. Damit ist $p(x,y) = p(y|x)p(x) = p_x(x)p_y(y)$, woraus sich für die Entropie

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

und die Transinformation

$$I(X,Y) = -H(X,Y) + H(X) + H(Y) = 0$$

ergibt.

(ii) „perfekte Übertragung“. Wir nehmen $X = Y$ an und können jedes $y \in Y$ mit einem $x \in X$ identifizieren. Dann gilt für die Wahrscheinlichkeiten $p(y|x) = \delta_{xy}$, $p_x(x) = p_y(x)$ und damit $p(x,y) = \delta_{xy}p_x(x) = \delta_{xy}p_y(y)$. Für die Entropie ergibt sich

$$H(X,Y) := -\sum_{x,y} p(x,y)\log p(x,y) = -\sum_x p_x(x)\log p_x(x) = H(x) = H(y)$$

und für die Transinformation

$$I(X;Y) = 2H(X) - H(X) = H(X).$$

Wir suchen nun eine Definition für die Kapazität eines Kanals (klassisch). Die gegenseitige Information $I(X;Y)$ hängt ab von

1) Kanal über $p(y|x)$,

2) Quelle/Sender über $p_x(x)$.

Wir wollen die bestmögliche Informationsübertragung durch den Kanal betrachten. Daher definieren wir

$$C := \max_{\{p_x\}} I(X;Y)$$

als klassiche *Kapazität* des Kanals. C hängt damit nur noch vom Kanal ab (über $p(y|x)$).

Beispiel: Symmetrischer binärer Kanal: $X = Y = \{0,1\}$ Setze $p_x : X \to [0,1], 0 \mapsto p, 1 \mapsto 1-p$ für ein $p \in [0,1]$. Damit ergibt sich

$$H(X) = -p\log p - (1-p)\log(1-p).$$

Für den symmetrischen binären Kanal gelten die Übertragungswahrscheinlichkeiten

$$p(0|0) = p(1|1) = 1-\varepsilon$$
$$p(1|0) = p(0|1) = \varepsilon$$

für ein $\varepsilon \in [0,1]$.
Für $p_y(y) := \sum_{x\in X} p(y|x)p_x(x)$ ergibt sich damit

$$p_y : Y \to [0,1], \begin{cases} 0 \mapsto (1-\varepsilon)p + \varepsilon(1-p) = p + \varepsilon - 2\varepsilon p \\ 1 \mapsto (1-\varepsilon)(1-p) + \varepsilon p = 1 - p - \varepsilon + 2\varepsilon p. \end{cases}$$

Für die Shannon-Entropie erhalten wir daraus

$$H(Y) = -(p+\varepsilon+2\varepsilon p)\log(p+\varepsilon+2\varepsilon p) - (1-p-\varepsilon-2\varepsilon p)\log(1-p-\varepsilon-2\varepsilon p)$$

Mit $p(x,y) = p(y|x)p_x(x)$ ergibt sich die gemeinsame Verteilungsfunktion

$$p : X \times Y \to [0,1] \quad \begin{cases} (0,0) \mapsto (1-\varepsilon)p \\ (1,0) \mapsto \varepsilon p \\ (0,1) \mapsto \varepsilon(1-p) \\ (1,1) \mapsto (1-\varepsilon)(1-p) \end{cases} .$$

Deren Entropie beträgt

$$H(X,Y) = -(1-\varepsilon)p\log(1-\varepsilon p) - \varepsilon p\log\varepsilon p - \varepsilon(1-p)\log\varepsilon(1-p) - (1-\varepsilon)(1-p)\log(1-\varepsilon)(1-p).$$

Damit können wir nun die gegenseitige Information berechnen

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y) \\ &= \varepsilon\log\varepsilon + (1-\varepsilon)\log(1-\varepsilon) - \\ &\quad (p+\varepsilon+2\varepsilon p)\log(p+\varepsilon+2\varepsilon p) + (1-p-\varepsilon+2\varepsilon p)\log(1-p-\varepsilon+2\varepsilon p) \\ &= -H(\varepsilon) + H(p+\varepsilon+2\varepsilon p), \end{aligned}$$

wobei $H(z) = -z\log z - (1-z)\log(1-z)$ (für ein $z \in [0,1]$) als Abkürzung verwendet wurde. Für die Kapazität müssen wir nun $I(X;Y)$ als Funktion von p maximieren. Da p nur im zweiten Summanden auftaucht, genügt es, $H(p+\varepsilon+2\varepsilon p)$ zu maximieren. Wir wissen bereits, dass $H(z)$ maximal ist für $z = 1/2$, daher fordern wir

$$p_{\max} + \varepsilon - 2\varepsilon p_{\max} \stackrel{!}{=} \frac{1}{2}$$

und erhalten

$$p_{\max} = \frac{1/2-\varepsilon}{1-2\varepsilon} = \frac{1}{2}.$$

Damit ist die Kapazität des Kanals

$$C = \max_p I(p,\varepsilon) = I(1/2,\varepsilon) = 1 - H(\varepsilon).$$

In Abbildung 1.3 ist die Kapazität in Abhängigkeit von der Rauschwahrscheinlichkeit ε gezeigt. Für $\varepsilon = 1/2$ (nur Rauschen) hat der Kanal überhaupt keine Kapazität. Die Funktion ist symmetrisch bezüglich $\varepsilon = 1/2$. Für $\varepsilon > 1/2$ werden die Bits gedreht, was aber für den Informationsgehalt keinen Unterschied macht.

Wie kann man über einen fehlerhaften Kanal verlässlich Daten übertragen? Die Antwort liefert das Shannon-Theorem II.

Satz (Shannon-Theorem II (noisy channel coding theorem))**.** *Durch* Codierung *einer Nachricht der Länge nR (in Bits) in 2^{nR} Codeworte der Länge n kann die Nachricht asymptotisch ($n \to \infty$) fehlerfrei übertragen werden, falls die Rate $0 \le R \le 1$ die Kapazität C nicht übersteigt: $R < C$.*

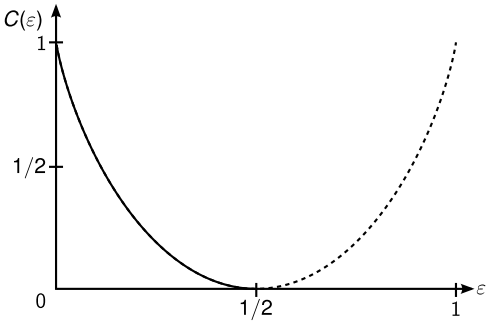


Abbildung 1.3: Kapazität eines symmetrischen binären Kanals mit Rauschwahrscheinlichkeit ε

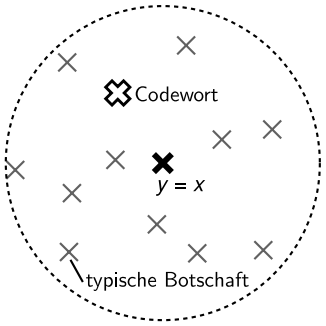


Abbildung 1.4: Visualisierung von empfangener Botschaft y und möglichen gesendeten typischen Botschaften. Falls in der Kugel genau ein Codewort vorhanden ist, ist eine Dekodierung möglich.

Beweisskizze. 1) Wähle 2^{nR} Codeworte aus 2^n möglichen n-Bit Nachrichten (Bitketten).

2) Idee: Codeworte solle möglichst „weit auseinander“ liegen (→*Hamming-Distanz*) so, dass fehlerhafte Codeworte noch unterscheidbar sind.

   Beispiel: $R = 1/3$, $0 \to 000$ und $1 \to 111$. Dann wird aus dem fehlerhaften Codewort 010 wieder 0. Damit können Fehler, die durch ein fehlerhaft übertragenes Symbol entstanden sind, korrigiert werden.

3) Idee (Shannon): Wähle Codeworte zufällig, d.h. mit Wahrscheinlichkeit $p(x)$ (random coding). Dies ist gut für den Beweis aber nicht sehr praktikabel.

4) Mit hoher Wahrscheinlichkeit sind so gewählte Codeworte *typische* Botschaften. Für diese gibt es $2^{nH(x)}$ Möglichkeiten.

5) typische empfangene Botschaft $y \to x$ $(X = Y)$ mit bedingter Wahrscheinlichkeit $p(x|y)$. Für diese gibt es $2^{nH(X|Y)}$ mögliche gesendete Botschaften.

6) Betrachte Kugel um y, welche $2^{n(H(X|Y)+\delta)}$, $\delta > 0$ Botschaften enthält (vgl. Abbildung 1.4).

7) Falls genau ein Codewort in der Kugel ist, kann die Nachricht dekodiert werden.

8) Der Anteil der typischen Botschaften in der Kugel ist

$$\frac{2^{n(H(X|Y)+\delta)}}{2^{nH(X)}} = 2^{-n(H(X)-H(X|Y)-\delta)} = 2^{-n(I(X;Y)-\delta)}.$$

9) Die Wahrscheinlichkeit, dass ein Codewort zufällig in der Kugel liegt, beträgt

$$\overline{p} = 2^{nR}2^{-n(I(X;Y)-\delta)} = 2^{-n(I(X;Y)-R-\delta)}.$$

Um fehlerfrei Dekodieren zu können muss $\overline{p} \to 0$ für $n \to \infty$ gelten. $\delta > 0$ kann beliebig klein gewählt werden und damit ist $R < I$ beliebig nahe an I wählbar (aber nicht $\geq I$).

10) Man kann zeigen, dass Codes existieren mit einer beliebig kleinen Fehlerwahrscheinlichkeit für jedes Codewort, nicht nur im Mittel.

11) Das Maximum über alle Wahrscheinlichkeitsverteilungen $p(x)$ liefert

$$R < I(X,Y) \leq \max_{\{p(x)\}} I(X;Y) \equiv C.$$

Somit muss die Rate kleiner als die Kanalkapazität sein.

□

1.3 Quanteninformation

Wir verwenden zur Informationsübertragung und -speicherung statt klassischer Systeme (z.B. Bits) nun Quantensysteme. Das einfachste Quantensystem ist ein quantenmechanisches zwei-Zustandssystem.
Beispiele:

(i) Photonen mit Polarisation $|H\rangle$ (horizontal) und $|V\rangle$ (vertikal). Man kann auch zirkular polarisierte Photonen betrachten mit $|\sigma_\pm\rangle \sim |H\rangle \pm i\,|V\rangle$.

(ii) Abstrakt können wir definieren $|H\rangle =: |0\rangle$, $|V\rangle =: |1\rangle$. Dann gilt $\langle 0|\,1\rangle = \langle H|\,V\rangle = 0$.

(iii) Elektronen mit Spin 1/2. Bezüglich einer beliebigen aber fest gewählten Quantisierungsrichtung (z.b. externes Magnetfeld) definiere $|\uparrow\rangle = |0\rangle$, $|\downarrow\rangle = |1\rangle$.

Allgemein bzw. abstrakt bezeichnen wir dieses System als *Quantenbit* oder *Qubit*. Die (Rechen-)Basis ist $|0\rangle, |1\rangle$. Allgemeine Zustände können durch Linearkombinationen dargestellt werden

$$\alpha\,|0\rangle + \beta\,|1\rangle \;\text{ mit } \alpha, \beta \in \mathbb{C}\,.$$

Aufgrund der Normierung muss $|\alpha|^2 + |\beta|^2 = 1$ gefordert werden und da die globale Phase irrelevant ist, kann ohne Einschränkung $\alpha \in \mathbb{R}$ gewählt werden. Damit bleiben zwei Freiheitsgrade für die Wahl von α und β. Diese Zustände können durch die Bloch-Kugel (bzw. Poincaré-Kugel für Photonen) veranschaulicht werden.
System zur Übertragung von Quantensystemen (z.B. Qubits). Ein *Quantenkanal* benötigt ein *Quantenalphabet* $X \subset \mathcal{H}$ aus einem Hilbertraum $\mathcal{H}$. Für die Symbole schreiben wir $|\varphi_x\rangle \in X$ mit $x = 1, 2, \ldots, K$. Die Wahrscheinlichkeitsverteilung sei gegeben durch $p(x)$ ($\sum_{x\in X} p(x) = 1$).
Allgemeiner ist das Alphabet für gemischte Zustände gegeben durch $X = \{\rho_1, \ldots, \rho_K\}$ mit den Dichtematrizen ρ_i. Wir notieren dies als *Ensemble*

$$\mathcal{E} = (\{\rho_x : x = 1, \ldots, K\}, p).$$

Für reine Zustände ist $\rho_x = |\varphi_x\rangle\langle\varphi_x|$, d.h. das Alphabet ist das selbe wie oben.
Fragestellungen:

1) Wieviel klassische Information kann mit einer Quantenquelle durch einen Quantenkanal übertragen werden?
   Die Antwort gab Holevo schon 1973 (vgl. **??**). Die *zugängliche Information* (*accessible information*) ist beschränkt durch die Holevo-Schranke $\chi(\mathcal{E})$:

$$\mathrm{Acc}(\mathcal{E}) \leq \chi(\mathcal{E})$$

   Im „besten Fall“, d.h. für reine, orthogonale Zustände kann man höchstens n Bits pro n Qubits übertragen.

2) Analog zu Shannon I: Wie stark lässt sich eine Botschaft aus einer Quantenquelle komprimieren?

   (a) Alphabet aus reinen Zuständen, d.h. $\mathcal{E} = (\{|\varphi_x\rangle\}, p)$
   Die Antwort liefert Benjamin Schumacher 1995 durch die Schumacher-Codierung. Gleichzeitig erfand er das Wort *Qubit*. Dazu definiert man $\rho = \sum_x p(x) |\varphi_x\rangle \langle\varphi_x|$ und die *von-Neumann-Entropie*

$$S(\rho) = -\operatorname{tr}(\rho \log \rho),$$

   wobei tr die Spur bezeichnet und der Logarithmus sinnvoll ist, da ρ positiv definit ist. Nach Schumacher benötigt man $nS(\rho)$ Qubits, um n Qubits zu übertragen bzw. speichern.

   Im Spezialfall $\langle\varphi_x|\,\varphi_{x'}\rangle = \delta_{xx'}$, d.h. für orthogonale also unterscheidbare Zustände gilt $S(\rho) = H(p)$. Wir erhalten also für unterscheidbare (reine) Quantensysteme den klassischen Fall!

   (b) Alphabet aus gemischten Zuständen $\{\rho_x\}$
   Die Antwort liefert die *Holevo-Information*. Definiert man $\rho = \sum_x p(x)\rho_x$ und

$$\chi(\rho) := S(\rho) - \sum_x p(x) S(\rho_x),$$

   so kann man n Qubits auf $\chi(\rho)n$ Qubits komprimieren. Für reine Zustände gilt

$$S(|\varphi_x\rangle \langle\varphi_x|) = S\left(\begin{pmatrix} 0 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \ddots \\ & & & & 0 \end{pmatrix}\right) = 0,$$

   wir haben also wieder Fall (a).

3) Analogon zu Shannon II: Wieviel Quanteninformation lässt sich durch einen fehlerhaften Quantenkanal transportieren?
   $\rightarrow$ *Quanten-Kapazität* C_Q

Bemerkungen:

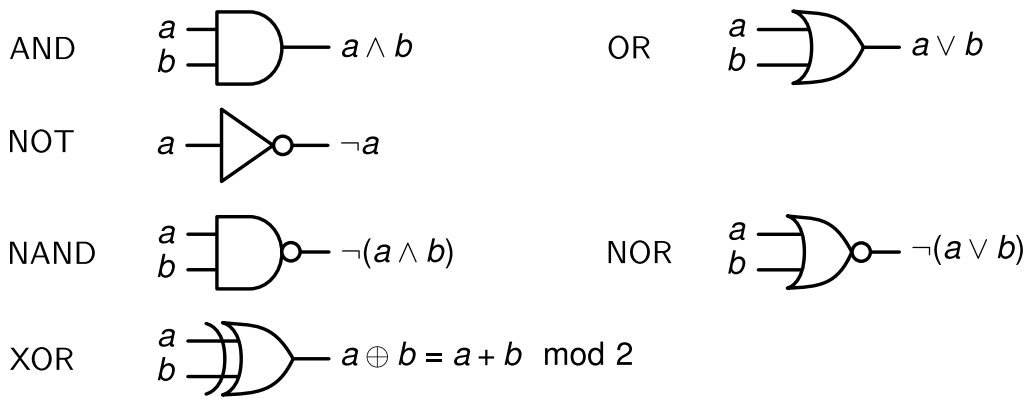
- Übergang von klassischer Information zu Quanteninformation
  projektive Messung $\rightarrow$ verallgemeinerte Messung (POVM)
- benötige Quantenmechanik der gemischten Zustände, Dichtematrizen, ...
  Kanal bildet Dichteoperatoren auf Dichteoperatoren ab, $\rho \mapsto \rho' = \Lambda\rho$, Λ wird deshalb auch Superoperator genannt (vgl. Kapitel 2.4, Quanteninformation Kapitel **??**)

1.4 Klassische Rechner

Es gibt die beiden wichtigen Modelle der Turing-Maschine und der Schaltungen (circuits). Wir wollen hir nur Schaltungen betrachten. Die Grundlage jedes Computers ist die Berechnung einer Bool'schen Funktion $f : \{0,1\}^n \to \{0,1\}^m$.

Bemerkung: Falls nötig können wir ohne Beschränkung der Allgemeinheit $m = 1$ annehmen (Entscheidungsproblem). Der Grund hierfür ist, dass man f in m Entscheidungsprobleme aufteilen kann und eine 1-Bit-Ausgabe oft ausreichend ist.

Man kann jede Boolsche Funktion f durch elementare logische Gatter (Operatoren) ausdrücken:



Unter Verwendung der Gatter AND, OR und NOT kann man f darstellen in der *konjunktiven Normalform*

$$f(x) = \bigwedge_i \bigvee_j (\neg)^{n_{ij}} x_{R_{ij}},$$

wobei $x = (x_1, \ldots, x_n)$, $R_{ij} \in \{1, \ldots, n\}$, $n_{ij} \in \{0,1\}$ und $(\neg)^0 := 1$, $(\neg)^1 := \neg$. Alternativ gibt es mit diesen Gattern auch die *disjunktive Normalform*

$$f(x) = \bigvee_i \bigwedge_j (\neg)^{n_{ij}} x_{R_{ij}}.$$

Mit Hilfe der de-Morganschen Regeln kann man aus AND und NOT ein OR-Gatter bauen, oder aus OR und NOT ein AND-Gatter, z.B. ist $a \wedge b = \neg(\neg a \vee \neg b)$. Man kann sogar zeigen, dass ein Gatter (z.B. NAND) ausreichen ist. In den bisherigen Schlüssen haben wir implizit angenommen, dass COPY oder FANOUT, also ein Kopieren der Bits (), möglich ist sowie Hilfsbits verwendet werde können ().

Bemerkung: AND, OR und NAND sind irreversible Operationen. Nach Landauer wird dabei die Wärme $\Delta Q = k_B T \ln 2$ frei.
Diesen Zusammenhang kann man sich anhand eines Bits, das durch ein ideales Gas gegeben wird, klarmachen. Dazu betrachten wir ein ideales Gas mit einem Molekül in einem Behälter mit zwei Kammern (vgl Abbildung 1.5). Ist das Molekül in der linken Kammer, so hat das Bit den Wert 0, sonst 1. Das Bit kann gelöscht, d.h. in einen definierten Zustand gebracht werden, in dem das Gas soweit komprimiert wird, bis es nur noch die linke Kammer füllt. Die Entropieänderung beträgt dabei $\Delta S = k_B \ln V_2/V_1 = k_B \ln 2$.

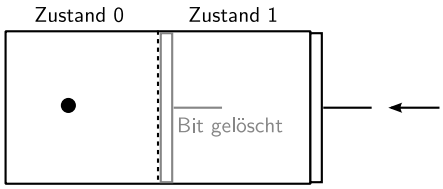
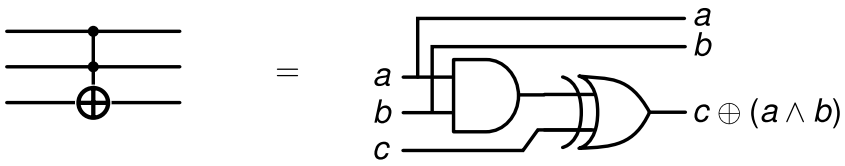
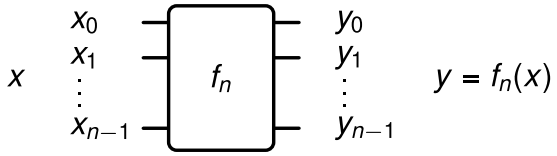


Abbildung 1.5: Entropieänderung durch Löschen eines Bits

Reversibles Rechnen ist immer möglich (Bennett 1973), in dem genügend Information über die Eingabe behalten wird. Eleganter ist das Toffoli-Gatter:



Dieses arbeitet reversibel und arbeitet für $c = 0$ für a, b wie ein AND, für $c = 1$ für a, b wie ein NAND und für $a = b = 1$ für c wie ein NOT-Gatter und ist damit universell.
Was passiert mit der Eingabe a, b? Muss man diese irgendwann doch löschen? Nein, denn man kann rechnen, das Ergebnis kopieren (ist reversibel) und anschließend die Rechnungen umkehren (rückwärts rechnen). Da alle Prozesse reversibel sind, erhält man daraus wieder den Anfangszustand.
Zu einem Problem (Multiplikation, Faktorisierung, ...) gehört nicht nur eine Funktion f, sondern eine ganze Familie $\{f_n\}_n$ $(n = 1, 2, \ldots)$ zu jeder Problemgröße n.



Hinter der *Komplexität* eines Problems steht die Frage: Wie groß bzw. komplex ist die Schaltung C_n für f_n als Funktion von n? Hierfür gibt es unterschiedliche Begriffe:

Größe (engl. size)	Größe$(C_n) =$	Anzahl der Gatter
Breite (engl. width)	Breite$(C_n) =$	maximale Anzahl an Gattern, die gleichzeitig ausgeführt werden müssen
Tiefe (engl. depth)	Tiefe$(C_n) =$	minimale Anzahl an Rechenschritten, wenn C_n parallel läuft

Bemerkung: Die Funktionsvorschrift $f_n \mapsto c_n$ muss „leicht“ zu berechnen sein, sonst kann die Komplexität des Problems unter Umständen in dieser Abbildung „versteckt“ werden $\rightarrow$ *Uniformität*.

Frage: Wie unterscheidet man „leichte“ von „schweren“ Problemen?

Dies ist keine leichte Frage. Die Standardantwort aus der Informatik ist folgende:

leicht die Größe des Problems ist polynomial beschränkt, d.h. es gibt $k \geq 1$ und $c > 0$,

sodass

$$\text{Größe}(C_n) \leq cn^k$$

für alle $n \in \mathbb{N}$ gilt. Ein Beispiel hierfür ist die Multiplikation. Diese Komplexitätsklasse bezeichnet man mit P (für polynomial).

$$\text{P} := \{\text{Entscheidungsprobleme } f_n : \{0,1\}^n \to \{0,1\} \text{ für welche Größe}(C_n) \leq cn^k\}$$

schwer sind alle anderen Probleme. Dies ist der generische Fall, da immer alle 2^n Werte von f_n tabelliert werden können. Dann gilt Größe$(C_n) \sim 2^n > cn^k$, da 2^n exponentiell wächst.

Bemerkung:

(i) Die Komplexitätsklasse hängt *nicht* vom gewählten (universellen) Satz von logischen Gattern ab, denn diese beeinflussen die Größe nur über einen konstanten Faktor, da jedes logische Gatter durch einen endlichen Satz an anderen Gattern ersetzt werden kann.

(ii) Die Komplexität ist unabhängig vom gewählten Modell (z. B. Schaltung oder Turing-Maschine) Die *Church-Turing-Hypothese* besagt: alle Probleme, die „irgendwie“ in der Natur berechenbar sind, können durch eine universelle Turing-Maschine berechnet werden. Die starke Church-Turing-Hypothese besagt, dass ein Problem, das auf einer Maschine effizient (polynomial) berechnet werden kann, auch auf allen anderen Maschinen effizient gelöst werden kann.

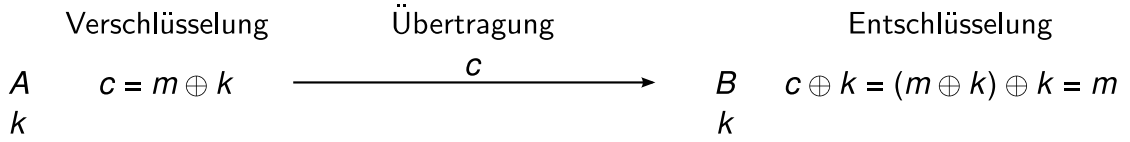
Weitere Komplexitätsklassen sind z.B.

$$\text{NP} := \{\text{Entscheidungsprobleme, deren Antwort leicht überprüfbar ist,}$$
$$\text{wenn ein Hinweis (}\textit{witness}\text{) gegeben ist}\},$$

wobei NP für *non-deterministic polynomial* steht. Ein Beispiel hierfür ist die Faktorisierung F, d.h. die Aufgabe „bestimme die Primfaktoren“. Als Entscheidungsproblem kann das Problem so formuliert werden: Gegeben ein $x \in \mathbb{N}$ und ein $y \in \mathbb{N}$ mit $y < x$. Gibt es einen Faktor von x, der kleiner ist als y? Bislang ist kein effizienter (polynomial großer) Algorithmus $(C_n\})$ bekannt (klassisch). Man vermutet $F \notin$ P. Aber wenn ein $z < y$ gegeben ist, das ein Faktor von x sein soll, kann diese Lösung mittels eines Algorithmus aus P überprüft werden: Division mit Rest von x durch z. Daher $F \in$ NP. Es ist klar, dass P $\subseteq$ NP. Für die Antwort auf NP = P? gibt es vom Clay Mathematics Institute 1×10^6 US$.
Alle heutigen Kryptographieverfahren basieren darauf, dass bestimmte Operationen mit Hinweisen (Schlüssel) polynomiell, aber ohne nur exponentiell ausgeführt werden können.

Kryptographie

Shannon zeigte: Das einzig (mathematisch, informationstheoretisch) sichere Verfahren ist das *one time pad* (Vernam Verschlüsselung). Gegeben ist eine Botschaft $m := x_0x_1 \cdots x_{n-1}$ aus n Bits. Mit Hilfe eines gleich langen geheimen zufälligen Schlüssels k, der nur einmal (!) verwendet werden darf, verschlüsselt man die Botschaft durch die Operation $c = m \oplus k$.



Bei mehrmaliger Verwendung z.B. $c_1 = m_1 \oplus k$, $c_2 = m_2 \oplus k$ kann man den Schlüssel durch $c_1 \oplus c_2 = m_1 \oplus m_2$ eliminieren und so versuchen, die Botschaften zu entschlüsseln. Dieses Verfahren ist unpraktisch, aber sicher (Shannon).
Eine Lösung sind Verfahren mit öffentlichen Schlüsseln (*public key*), z.B.

1) Diffie-Hellmann 1970er Jahre

2) RSA 1976 Rivest, Shamir und Adleman. Dieses Verfahren wurde bereits vorher von Ellis, Cocks und Williams für den Britischen Geheimdienst GHCQ entwickelt und erst 1997 veröffentlicht (vgl. Abbildung 1.6)

Bemerkung: Peter Shor hat 1994 gezeigt, dass es einen effizienten Quantenalgorithmus zur Faktorisierung gibt.
Weitere Komplexitätsklassen sind hier nur noch kurz erwähnt, z.B.

$$\begin{aligned}\text{BPP} = \{&\text{Probleme } \{f_n\}, \text{ für welche ein } \{C_n\} \text{ mit Zufallselementen existiert,}\\ &\text{wobei Größe}(C_n) \leq cn^k, \text{ sodass mit Wahrscheinlichkeit } \frac{1}{2} + \varepsilon, (\varepsilon > 0)\\ &\text{die korrekte Antwort erzeugt wird}\},\end{aligned}$$

BPP steht für *bounded error probabilistic polynomial*. Später lernen wir BQP kennen, die Klasse der mit einem Quantenrechner in polynomialer Zeit lösbaren Probleme (vgl. Kapitel **??**).
Bemerkungen:

1) Der Wert von $\varepsilon > 0$ ist egal, da die Wahrscheinlichkeit durch Wiederholungen beliebig nahe an 1 kommt

2) Es ist klar, dass P $\subseteq$ BPP. Falls BPP $\neq$ P, muss die starke Church-Turing-Hypothese revidiert werden.

1.5 Quantenrechner

Peter Shor zeigte 1994, dass ein Quantenalgorithmus existiert, welcher eine Zahl aus n Bit mit polynomialem Aufwand in n faktorisiert. Damit gilt $F \in$ BQP. Es wird vermutet, dass

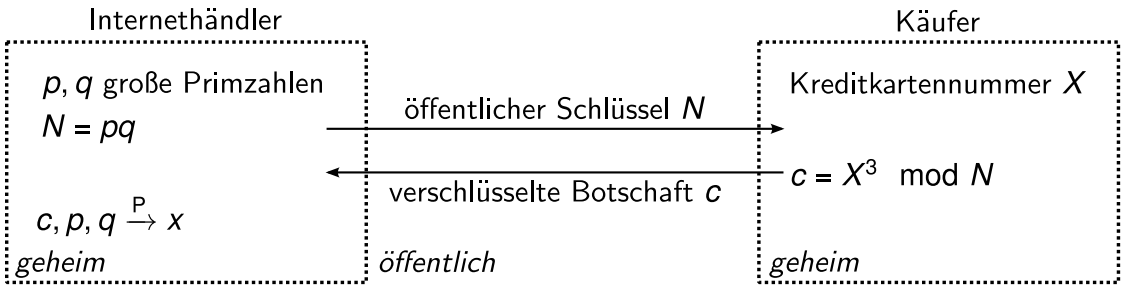
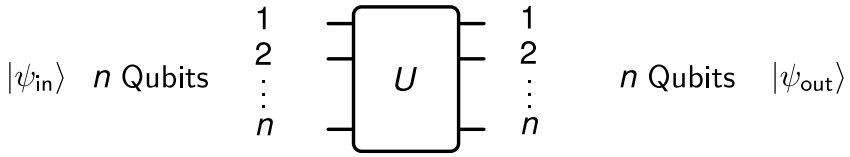


Abbildung 1.6: Ablauf der Verschlüsselung mit RSA. Mit Hilfe der beiden Primzahlen p und q ist die Entschlüsselung mit polynomialem Aufwand möglich. Sonst kann man x nur mit exponentiellem Aufwand herausfinden.

$F \notin$ BPP. Falls tatsächlich $F \in$ BQP $\setminus$ BPP gelten würde, wäre die (revidierte) starke Church-Turing-Hypothese falsch (und müsste nochmals revidiert werden).
Für Quantenrechner gibt es analoge Modelle, die Quanten-Turing-Maschine (Deutsch 1980er) und die *Quantenschaltung* (*quantum circuit*). Die Quantenschaltung werden wir im Folgenden genauer untersuchen.
Die Ausgangswellenfunktion $|\psi_{\text{out}}\rangle$ geht durch eine unitäre Abbildung U aus der Eingangswellenfunktion $|\psi_{\text{in}}\rangle$ hervor, d.h.

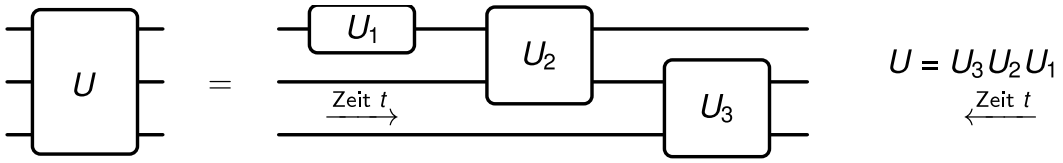
$$|\psi_{\text{out}}\rangle = U\,|\psi_{\text{in}}\rangle\,,$$

da die zeitliche Entwicklung von abgeschlossenen Systemen durch die Schrödingergleichung durch unitäre Operatoren gegeben ist.



Eine Quantenschaltung besitzt aufgrund der Reversibilität immer genau so viele Eingänge wie Ausgänge. Deshalb ist U eine $2^n \times 2^n$-Matrix. Die Wellenfunktion für die Schaltung stammt aus dem Hilbertraum $\mathcal{H} = \mathcal{H}_2^{\otimes n}$ mit $\mathcal{H}_2 = \{\alpha\,|0\rangle + \beta\,|1\rangle\}$, welcher die Dimension $\dim \mathcal{H} = 2^n$ besitzt.

Quantengatter



Di Vincenzo zeigte 1995, dass Ein- und Zwei-Qubit-Gatter ausreichend sind, um einen beliebigen Quantenalgorithmus U zu erzeugen. Zum Beispiel SU(2) (Spezielle unitäre Gruppe der Dimension 2, unitäre 2×2-Matrizen mit Determinante 1) zusammen mit dem Controlled-NOT-Gatter (CNOT) (s. unten).
$U \in$ SU(2) ist ein Beispiel für *Ein-Qubit-Operationen* (-Gatter), z.B. die *Pauli-Matrizen/-Gatter*

$$X = \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad Y = \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad Z = \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

die Hadamard-Matrix bzw. das *Hadamard-Gatter*

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \text{ d.h. } \begin{array}{lcl} |0\rangle & \mapsto & \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \\ |1\rangle & \mapsto & \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle), \end{array}$$

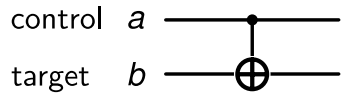
das *Phasengatter*

$$S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$$

und das $\pi/8$-*Gatter*

$$T := \begin{pmatrix} 1 & 0 \\ 0 & \exp(i\pi/4) \end{pmatrix}.$$

Ein Beispiel für *Zwei-Qubit-Gatter* ist das CNOT (controlled NOT, XOR, Quanten-XOR).



Wirkung auf Produktbasis:

$$\begin{array}{lcl} |0\rangle_a |0\rangle_b & \mapsto & |0\rangle_a |0\rangle_b \\ |0\rangle_a |1\rangle_b & \mapsto & |0\rangle_a |1\rangle_b \\ |1\rangle_a |0\rangle_b & \mapsto & |1\rangle_a |1\rangle_b \\ |1\rangle_a |1\rangle_b & \mapsto & |1\rangle_a |0\rangle_b \end{array}$$

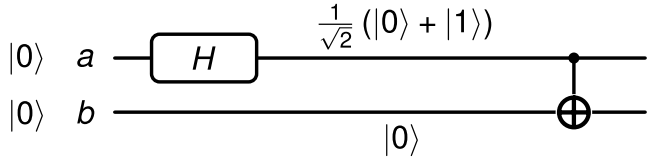
allgemein: $|x\rangle_a |y\rangle_b \mapsto |x\rangle_a |x \oplus y\rangle_b$ und linear fortgesetzt

$$U_{\text{CNOT}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Für beliebige Anfangszustände $|\psi\rangle = \sum_{x,y=0,1} c_{xy} |x\rangle_a |y\rangle_b \in \mathcal{H}_2 \otimes \mathcal{H}_2$ ($\mathcal{H}_2 = \{\alpha |0\rangle + \beta |1\rangle\}$) kann man U_{CNOT} ebenfalls linear fortsetzen durch

$$U_{\text{CNOT}} |\psi\rangle = \sum_{x,y=0,1} c_{xy} U_{\text{CNOT}} |x\rangle_a |y\rangle_b = \sum_{x,y=0,1} c_{xy} |x\rangle_a |x \oplus y\rangle_b .$$

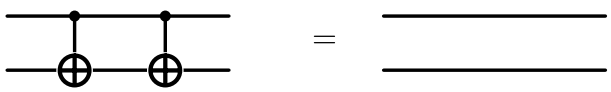
Um die Wirkung von U_{CNOT} zu verstehen betrachten wir folgendes Beispiel:



$$\frac{1}{\sqrt{2}} \left(|0\rangle_a |0\rangle_b + |1\rangle_a |1\rangle_b \right)$$

Der Endzustand lässt sich nicht mehr auf die beiden Kanäle aufteilen. In diesem Fall spricht man von *Verschränkung*!

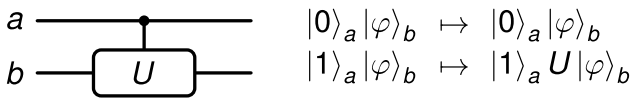
Bemerkung: Das Quantengatter ist reversibel, weil $U^{-1} = U^\dagger$, da U unitär ist, z.B.



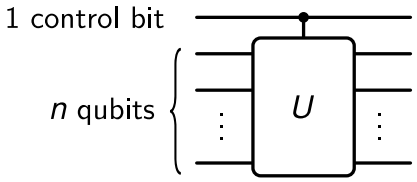
$$U_{\text{CNOT}}^2 = \mathbb{1}$$

Allgemeinere Quantengatter sind

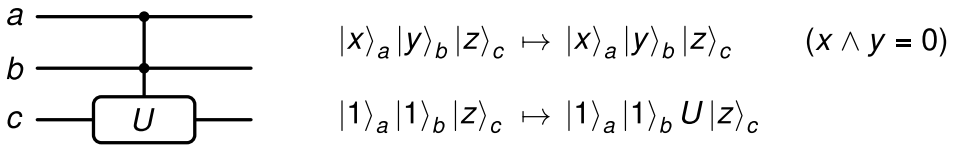
1) *controlled-U* für $U \in \text{SU}(2)$. Im Spezialfall $U = \sigma_x$ erhalten wir das CNOT-Gatter. Die Operation ist gegeben durch



2) controlled-U mit $U \in \mathrm{SU}(n)$ für n Qubits:

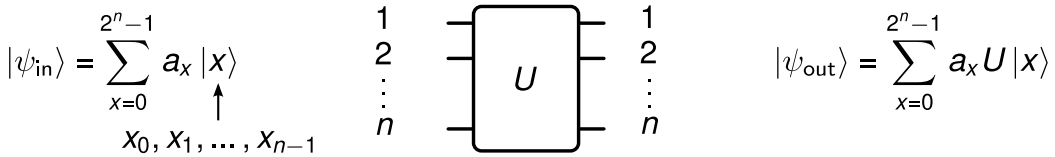


3) controlled-controlled U:

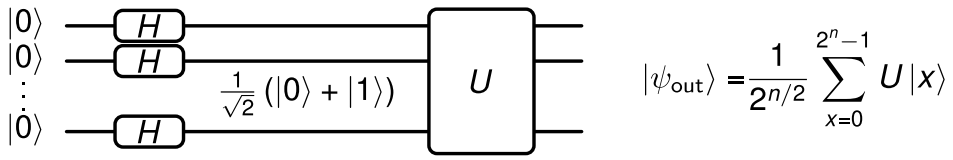


Für $U = \sigma_x$ erhalten wir das Quanten-Toffoli-Gatter.

Bei n-Bit-Gattern ergibt sich ein so genannter *Quantenparallelismus*:



In der Produktbasis könnte man auch $|\psi_{\text{in}}\rangle == \sum_{x_i=0,1} a_{(x_0,\ldots,x_{n-1})} |x_0\rangle \cdots |x_{n-1}\rangle$ schreiben mit komplexen Gewichtungskoeffizienten $a_{(x_0,\ldots,x_{n-1})}$. Dabei hat U auf 2^n Eingaben gleichzeitig gewirkt! Versuche z.B.



Das Problem dabei ist, dass eine Messung die Überlagerung in $|\psi_{\text{out}}\rangle$ zerstört und man nur ein Ergebnis erhält. Daher benötigt man weitere Tricks, bzw. gute Algorithmen, um *globale* Eigenschaften von $|\psi_{\text{in}}\rangle$ herauszufinden.

Beispiel: Deutsch-Algorithmus (David Deutsch, 1985)
Gegeben sei eine unbekannte Boolsche Funktion (klassich) $f : \{0,1\} \to \{0,1\}$.
Frage: Ist f konstant, d.h. $f(0) = f(1)$ oder nicht ($f(0) \neq f(1)$)?

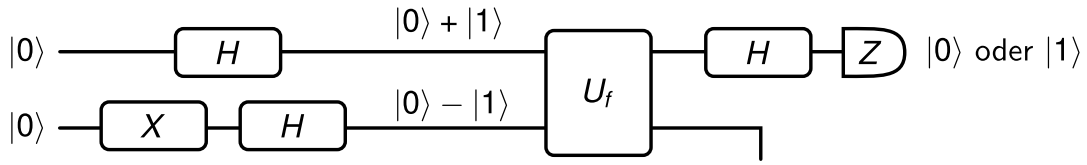


Abbildung 1.7: Quantenschaltkreis des Deutsch-Algorithmus

Zusätzlich sei $U_f : |x\rangle\,|y\rangle \mapsto |x\rangle\,|y \oplus f(x)\rangle$ als reversible Version von f (mit $U_f^2 = \mathbb{1}$) gegeben. Die Quantenmechanik ergibt nun

$$\begin{aligned}(|0\rangle + |1\rangle)_a(|0\rangle - |1\rangle)_b \overset{U_f}{\mapsto} |0\rangle_a \left(|f(0)\rangle - \underbrace{|1 \oplus f(0)\rangle}_{=|\neg f(0)\rangle}\right)_b + |1\rangle_a \left(|f(1)\rangle - \underbrace{|1 \oplus f(1)\rangle}_{=|\neg f(1)\rangle}\right)_b \\ = |0\rangle_a (-1)^{f(0)} (|0\rangle - |1\rangle)_b + |1\rangle_a (-1)^{f(1)} (|0\rangle - |1\rangle)_b \\ = (-1)^{f(0)} (|0\rangle + (-1)^{f(1)-f(0)} |1\rangle)_a (|0\rangle - |1\rangle)_b\end{aligned}$$

Hierbei wurden die Normierungskonstanten aus Gründen der Übersichtlichkeit weggelassen. Misst man nun das erste Qubit in der Basis $|\pm\rangle := |0\rangle \pm |1\rangle$, so wissen wir beim Ergebnis $|+\rangle$, dass $f(0) = f(1)$ und beim Ergebnis $|-\rangle$, dass $f(0) \neq f(1)$. Dafür ist nur eine einzige Abfrage nötig. Der zu diesem Algorithmus gehörende Quantenschaltkreis ist in Abbildung 1.7 gezeigt.

Bemerkung:

1) Es gibt eine Verallgemeinerung (Deutsch-Josza) für n Bits: $f : \{0,1\}^n \to \{0,1\}$ sei unbekannt, aber wir wissen f ist entweder konstant oder ausgeglichen (balanciert), d.h. 2^{n-1}-mal gilt $f(x) = 0$ und 2^{n-1}-mal gilt $f(x) = 1$. Klassisch benötigt man im schlimmsten Fall $\lesssim 2^n/2$ Abfragen. Quantenmechanisch kommt man mit einer Abfrage aus. Der quantenmechanische Algorithmus ist also exponentiell schneller.

2) Dies sind sogenannte *Orakelprobleme*. Da man annimmt, dass es eine quantenmechanische Implementierung der Funktion f gibt, kann man die Komplexität zur „relativ zum Orakel f“ angeben. Im Gegensatz dazu kommt der oben angesprochene Faktorisierungsalgorithmus ohne ein solches Orakel aus.

2 Allgemeine Quantenzustände und -operatoren

2.1 Axiome der Quantenmechanik

1. Zustände

Zustände beschreiben ein physikalisches System vollständig. In der Quantenmechanik wird ein Zustand durch einen Vektor $|\psi\rangle$ aus einem Hilbertraum $\mathcal{H}$ beschrieben. Genauer beschreibt dieser Vektor einen Zustand eines *abgeschlossenen* Systems. Außerdem beschreiben Vektoren, die sich nur um eine skalare Multiplikation unterscheiden den selben Zustand. Man betrachtet daher *Strahlen* , die als Äquivalenzklasse in $\mathcal{H}$ bezüglich skalarer Multiplikation definiert sind, d.h. $|\psi\rangle \cong |\varphi\rangle \Leftrightarrow \exists c \in \mathbb{C}\setminus\{0\} : |\psi\rangle = c\,|\varphi\rangle$.
Ein Hilbertraum $\mathcal{H}$

(a) ist ein Vektorraum über $\mathbb{C}$

(b) mit einem Skalarprodukt $\mathcal{H} \times \mathcal{H} \to \mathbb{C}, (|\psi\rangle, |\varphi\rangle) \mapsto \langle\varphi|\,\psi\rangle$, das die Eigenschaften

(i) $\langle\psi|\,\psi\rangle > 0$ für alle $|\psi\rangle \neq 0$ und $\langle\psi|\,\psi\rangle = 0 \Leftrightarrow |\psi\rangle = 0$

(ii) $\langle\varphi|\,(a\,|\psi_1\rangle + b\,|\psi_2\rangle) = a\langle\varphi|\,\psi\rangle + b\langle\varphi|\,\psi\rangle$ für alle $\varphi, \psi \in \mathcal{H}$, $a, b \in \mathbb{C}$

(iii) $\langle\varphi|\,\psi\rangle = \langle\psi|\,\varphi\rangle^*$

erfüllt.

(c) und vollständig ist bezüglich der Norm $\|\psi\| = \sqrt{\langle\psi|\,\psi\rangle}$.

Die letzte Eigenschaft ist hier nicht so wichtig, da wir nur endlichdimensionale Hilberträume betrachten.

2. Observablen

Observablen sind messbare Größen. In der Quantenmechanik werden Observablen durch hermitesche Operatoren A auf $\mathcal{H}$ beschrieben: $A : \mathcal{H} \to \mathcal{H}, |\psi\rangle \mapsto A\,|\psi\rangle$. Diese besitzen folgende Eigenschaften:

(i) Linearität: $A(a\,|\psi_1\rangle + b\,|\psi_2\rangle) = aA\,|\psi_1\rangle + bA\,|\psi_2\rangle$ für $|\psi_1\rangle, |\psi_2\rangle \in \mathcal{H}$, $a, b \in \mathbb{C}$

(ii) Hermitizität: $A = A^\dagger$, wobei $A^\dagger$ der adjungierte Operator zu A ist, der durch die Beziehung $\langle\varphi|\,A\psi\rangle = \langle A^\dagger\varphi|\,\psi\rangle$ für alle $|\varphi\rangle, |\psi\rangle \in \mathcal{H}$ definiert ist. Dabei wurde die Schreibweise $|A\psi\rangle := A\,|\psi\rangle$ verwendet.

Hermitesche Operatoren besitzen eine Spektraldarstellung $A = \sum_n a_n P_n$, wobei $a_n \in \mathbb{R}$ ein Eigenwert von A und P_n der Projektor auf den Eigenraum zu a_n ist. Falls a_n nicht entartet ist, gilt $P_n = |\psi_n\rangle\langle\psi_n|$ mit dem zu a_n gehörenden Eigenvektor $|\psi_n\rangle$. Die Projektoren sind hermitesch $P_n^\dagger = P_n$ und es gilt $P_n P_{n'} = \delta_{nn'} P_n$.

3. Messung

Eine Messung wird beschrieben durch einen hermiteschen Operator $A = A^\dagger$.

(i) Der Ausgang der Messung zu A ist einer der Eigenwerte a_n von A.

(ii) Der Zustand nach der Messung ist gegeben durch

$$\left|\psi'\right\rangle = \frac{P_n \left|\psi\right\rangle}{\sqrt{\left\langle\psi\right| P_n \left|\psi\right\rangle}},$$

wenn der Zustand vor der Messung durch $|\psi\rangle$ gegeben war. Falls a_n nicht entartet ist, gilt $|\psi'\rangle = |\psi_n\rangle$.

(iii) Die Wahrscheinlichkeit für das Messergebnis a_n ist

$$p_n = \|P_n \left|\psi\right\rangle\|^2 = \left\langle\psi\right| P_n \left|\psi\right\rangle$$

und ohne Entartung $p_n = |\langle\psi_n|\psi\rangle|^2$.

Korollar. *Der Erwartungswert einer Observablen ist gegeben durch*

$$\langle A\rangle = \sum_n p_n a_n = \langle\psi| \underbrace{\sum_n a_n P_n}_{A} |\psi\rangle = \left\langle\psi\right| A \left|\psi\right\rangle.$$

4. Zeitentwicklung

Die *Zeitentwicklung* ist unitär und wird erzeugt durch einen hermiteschen Operator, den *Hamilton-Operator* H. Die *Schrödingergleichung* lautet ($\hbar = 1$)

$$\frac{\mathrm{d}}{\mathrm{d}t} \left|\psi(t)\right\rangle = -iH(t) \left|\psi(t)\right\rangle.$$

Aufgrund der Linearität der Schrödingergleichung verwenden wir den Ansatz

$$\left|\psi(t)\right\rangle = U(t) \left|\psi(0)\right\rangle.$$

Dann ergibt sich

$$U(t) = T \exp\left(-i \int_0^t H(t')\mathrm{d}t'\right),$$

wobei T der Zeitordungsoperator ist. Falls $\partial_t H = 0$, so folgt

$$U(t) = \exp\left(-itH\right).$$

Beispiel: Betrachten wir ein Qubit. Der Hilbertraum $\mathcal{H}$ hat Dimension 2 und wird aufgespannt durch $|0\rangle$ und $|1\rangle$. Ein allgemeiner hermitescher Operator ist gegeben durch

$$A = \begin{pmatrix} a & b \\ b^* & c \end{pmatrix}, \quad a, c \in \mathbb{R}, b \in \mathbb{C}.$$

Dieser lässt sich in einer Basis aus hermiteschen Matrizen entwickeln

$$A = \frac{a+c}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{a-c}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + \frac{b+b^*}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + i\frac{b-b^*}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

Die Koeffizienten sind alle reell. Die hermiteschen Matrizen sind gerade die Einheitsmatrix sowie die Pauli-Matrizen $\sigma_z, \sigma_x, \sigma_y$. Damit können wir auch schreiben

$$A = c_0\mathbb{1} + c_1\sigma_x + c_2\sigma_y + c_3\sigma_z = c_0\sigma_0 + c_1\sigma_1 + c_2\sigma_2 + c_3\sigma_3 = \sum_{i=0}^{3} c_i\sigma_i$$
$$= c_0 + \vec{c}\cdot\vec{\sigma},$$

wobei $c_i \in \mathbb{R}$ $(i = 0, 1, 2, 3)$, $\vec{c} = (c_1, c_2, c_3)^T, \vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3) = (\sigma_x, \sigma_y, \sigma_z)$.
Die Spur von A ist gegeben durch $\operatorname{tr} A = 2c_0$, da $\operatorname{tr}\sigma_i = 0$ $(i = 1, 2, 3)$.
Da der Hamiltonoperator ebenfalls hermitesch sein muss, kann man ihn also auch in der Form $H = \vec{n}\cdot\vec{\sigma} + c$ schreiben, wobei $\vec{n}$ zeitabhängig sein kann und ohne Einschränkung $c = 0$ gesetzt werden kann, da diese Konstante nur als globale Phase beiträgt. Mit der Definition $\vec{S} = \hbar\vec{\sigma}/2$ (wobei wir im Folgenden $\hbar = 1$ setzen) gilt für den Hamiltonoperator

$$H = \vec{b}\cdot\vec{S} = \frac{\vec{b}}{2}\cdot\vec{\sigma},$$

sodass wir uns ein Qubit immer als Spin in einem Magnetfeld $\vec{B}$ vorstellen können. Dabei ist $\vec{b} = g\mu_{\mathrm{B}}\vec{B}$. Falls $\vec{b}$ zeitunabhängig ist, ist auch H zeitunabhängig, und es gilt für den Zeitentwicklungsoperator

$$U(t) = \exp\left(-i\frac{\vec{b}\cdot\sigma}{2}t\right) = \mathbb{1}\cos\left(\frac{b}{2}t\right) - i\sin\left(\frac{b}{2}t\right)\frac{\vec{b}}{|\vec{b}|}\cdot\sigma.$$

Die Erweiterung auf offene Systeme kann in folgender Tabelle zusammengefasst werden:

	abgeschlossenes System	offenes System (Teilsystem)
Zustände	Strahlen im Hilbertraum $\mathcal{H}$	Dichtematrizen (gemischte Zustände) (keine Strahlen im Hilbertraum $\mathcal{H}$)
Observablen:	$A = A^\dagger = \sum_i \lambda_i P_i$	nicht orthogonale Projektoren, verall-
Messungen:	orthogonale Projektoren P_i	gemeinerte Messung: POVM = positive operator valued measure
Zeitentwicklung	unitär $U(t)$	nicht unitär

Ein offenes System verhält sich in Spezialfällen wie ein abgeschlossenes System.

2.2 Zustände (offenes System)

Betrachte das abgeschlossene Gesamtsystem $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, zusammengesetzt aus dem Teilsystem $\mathcal{H}_A$ mit der Orthonormalbasis $\{|i\rangle_A\}$ und dem Rest $\mathcal{H}_B$ mit der Orthonormalbasis $\{|j\rangle_B\}$. Die Zustände sind gegeben durch $|\psi\rangle_{AB} = \sum_{i,j} c_{ij}\,|i\rangle_A\,|j\rangle_B$ mit $\sum_{i,j}|c_{ij}|^2 = 1$.
Beispiel: zwei Qubits: $|\psi_{AB}\rangle = c_{00}\,|0\rangle_A\,|0\rangle_B + c_{01}\,|0\rangle_A\,|1\rangle_B + c_{10}\,|1\rangle_A\,|0\rangle_B + c_{11}\,|1\rangle_A\,|1\rangle_B$.
Sei O_A eine Observable auf $\mathcal{H}_A$. Durch $O = O_A \otimes \mathbb{1}_B$ kann O_A auf Zustände aus dem

Gesamtsystem $\mathcal{H}$ angewandt werden. Für den Erwartungswert gilt

$$\begin{aligned}\langle O\rangle &= \langle\psi_{AB}\,|O|\,\psi_{AB}\rangle \\ &= \sum_{ijkl} {c_{ij}}^* c_{kl}\, {}_A\langle i|_B \langle j|O_A\otimes \mathbb{1}_B\,|k\rangle_A\,|l\rangle_B \\ &= \sum_{ijkl} {c_{ij}}^* c_{kl}\, {}_A\langle i|O_A\,|k\rangle_A \underbrace{{}_B\langle j|\,l\rangle_B}_{\delta_{jl}} \\ &= \sum_{ik} \underbrace{\sum_j {c_{ij}}^* c_{kj}}_{=:\rho_{ki}} \underbrace{{}_A\langle i|O_A\,|k\rangle_A}_{=(O_A)_{ik}} \\ &= \sum_{ik} \rho_{ki} {O_A}_{ik} = \sum_k (\rho O_A)_{kk} = \operatorname{tr}(\rho O_A).\end{aligned}$$

ρ_{ki} sind die Komponenten der sogenannten *Dichtematrix* (auch *Dichteoperator* genannt) auf $\mathcal{H}_A$ und $(O_A)_{ik}$ ist eine Matrixdarstellung des Operators O_A in $\mathcal{H}_A$.
Durch das obige Vorgehen haben wir einen reinen Zustand $|\psi_{AB}\rangle = \sum_{ij} c_{ij}\,|i\rangle_A\,|j\rangle_B$ in $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ in einen gemischten Zustand auf $\mathcal{H}_A$ mit der Dichtematrix

$$\rho = \sum_{ki} \rho_{ki}\,|k\rangle_A\,{}_A\langle i| = \sum_{kij} {c_{ij}}^* c_{kj}\,|k\rangle_A\,{}_A\langle i|$$

überführt. Die Umkehrung, d.h. den Übergang von einem gemischten Zustand in einen reinen Zustand in einem größeren Hilbertraum nennt man *Reinigung* (engl. *purification*). Dies ist immer möglich. Dazu sei eine Dichtematrix ρ auf einem n-dimensionalen Hilbertraum $\mathcal{H}_A$ gegeben. Dann hat ρ die Darstellung

$$\rho = \sum_{i=1}^{r} p_i\,|\psi_i\rangle\,\langle\psi_i|$$

mit $p_1 > 0$ und $r \le n$. Dabei wurde verwendet, dass ρ hermitesch ist ($\rho = \rho^\dagger$) und positiv definit $\rho \ge 0$, d.h. alle Eigenwerte sind größer gleich 0, was wir später zeigen werden. Die obige Gleichung ist dann die Spektraldarstellung. Wähle nun $\dim \mathcal{H}_B \ge r$ und $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ als

$$|\psi\rangle = \sum_{i=1}^{r} \sqrt{p_i}\,|\psi_i\rangle\,|\chi_i\rangle\,.$$

Hierbei ist $\{|\psi_i\rangle\}$ die Orthonormalbasis der Eigenzustände von ρ in $\mathcal{H}_A$ und $\{|\chi_i\rangle\}$ eine beliebig gewählte Orthonormalbasis in $\mathcal{H}_B$. Dann ist $|\psi\rangle$ die *purification* von ρ, denn es gilt $c_{ij} = \delta_{ij}\sqrt{p_i}$ und damit $\rho_{ij} = \delta_{ij} p_i$.

Eigenschaften der Dichtematrix

(i) ρ ist hermitesch, denn aus $\rho_{ki} = \sum_j {c_{ij}}^* c_{kj}$ folgt $\rho^\dagger_{ki} = (\rho^T_{ki})^* = \sum_j c_{kj} {c_{ij}}^* = \rho^T_{ik} = \rho_{ki}$. Damit besitzt ρ die Spektraldarstellung

$$\rho = \sum_i \rho_i\,|\psi_i\rangle\,\langle\psi_i|,$$

wobei $|\psi_i\rangle$ die Eigenzustände und $\rho_i \in \mathbb{R}$ die reellen Eigenwerte von ρ sind.

(ii) $\rho \geq 0$, d.h. $\langle\psi|\rho|\psi\rangle \geq 0$ für alle $|\psi\rangle \in \mathcal{H}$, denn für $|\psi\rangle := \sum_i \alpha_i |i\rangle$ gilt

$$\langle\psi|\rho|\psi\rangle = \sum_{ij} \alpha_i{}^* \alpha_j \underbrace{\langle i|\rho|j\rangle}_{\rho_{ij}} = \sum_{ijk} \alpha_i{}^* \alpha_j c_{jk}{}^* c_{ik}$$
$$= \sum_k \left(\sum_i \alpha_i{}^* c_{ik}\right)\left(\sum_j \alpha_j c_{jk}{}^*\right) = \sum_k \left\|\sum_i \alpha_i{}^* c_{ik}\right\|^2 \geq 0.$$

(iii) $\rho_{kk} = \sum_j |c_{kj}|^2$. Damit gilt für die Spur von ρ: $\operatorname{tr}\rho = \sum_k \rho_{kk} = \sum_{jk} |c_{kj}|^2 = 1$, falls der ursprüngliche Vektor normiert war. Umgekehrt kann man ρ immer so wählen, dass $\operatorname{tr}\rho = 1$ gilt. Damit $\operatorname{tr}\rho = \sum_i p_i = 1$ mit $0 \leq p_i \leq 1$.

Interpretation von ρ

1) Zustände werden in allgemeinen (offenen) Systemen durch Dichtematrizen beschrieben.

2) Reine Zustände erhält man durch perfekte *Präparation*. Die gemischten Zustände (Dichtematrix) ergeben sich durch sogenannte *Ensemble-Präparation* , d.h. man erhält das Ergebnis $|\psi_i\rangle$ mit der Wahrscheinlichkeit p_i, falls die Dichtematrix die Gestalt $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ hat.

   Bemerkung: Die Ensemble-Interpretation (bzw. Präparation) ist nicht eindeutig! Dazu betrachten wir die Präparation von Qubits. Quelle 1 erzeugt jeweils mit Wahrscheinlichkeit $1/2$ die Zustände $|0\rangle$ und $|1\rangle$. Daraus ergibt sich die Dichtematrix

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

   Quelle 2 erzeugt ebenfalls jeweils mit Wahrscheinlichkeit $1/2$ die Zustände $(|0\rangle+|1\rangle)/\sqrt{2}$ und $(|0\rangle - |1\rangle)/\sqrt{2}$. Für die Dichtematrix ergibt sich ebenfalls

$$\rho = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Spezialfall

Wir betrachten den Gesamtraum $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ und den Produktzustand $|\psi_{AB}\rangle = |\psi_A\rangle \otimes |\psi_B\rangle$ (keine Korrelation zwischen A und B). Dann gilt für den Operator $O = O_A \otimes \mathbb{1}_B$

$$\langle O\rangle = \langle\psi_A|O_A|\psi_A\rangle \underbrace{\langle\psi_B|\psi_B\rangle}_{=1} = \operatorname{tr}\big(O_A \underbrace{|\psi_A\rangle\langle\psi_A|}_{\rho_A}\big).$$

Damit entspricht die Dichtematrix $\rho = |\psi_A\rangle\langle\psi_A|$ genau einem reinen Zustand $|\psi_A\rangle$.
Ein leicht nachprüfbares Kriterium ist folgendes: Sei ρ eine Dichtematrix mit $\operatorname{tr}\rho = 1$. ρ ist ein reiner Zustand, d.h. $\rho = |\psi\rangle\langle\psi|$ für ein $|\psi\rangle \in \mathcal{H}_A$ genau dann wenn $\operatorname{tr}(\rho^2) = 1$.

Beweis. „⇒“ Es gilt $\rho^2 = |\psi\rangle\langle\psi|\psi\rangle\langle\psi| = |\psi\rangle\langle\psi| = \rho$ und damit $\operatorname{tr}\rho^2 = \operatorname{tr}\rho = 1$.
„⇐“ Es ist $\operatorname{tr}\rho^2 = \sum_i p_i^2 = 1$ mit $0 \leq p_i \leq 1$ und $\sum_i p_i = 1$. Dann muss $p_k = 1$ gelten für ein k und $p_i = 0$ für $i \neq k$, d.h. $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| = |\psi_k\rangle\langle\psi_k|$. □

Für allgemeine gemischte Zustände ist $\operatorname{tr}(\rho^2) = \sum_i p_i^2 < 1$.

Beispiel: Qubit oder Spin 1/2-Teilchen Die Dichtematrix ist selbstadjungiert, d.h. $\rho^\dagger = \rho$, d.h. ρ lässt sich durch $c_0\mathbb{1} + \vec{c}\cdot\vec{\sigma}$ darstellen. Wegen $1 \stackrel{!}{=} \operatorname{tr}\rho = 2c_0$ ist $c_0 = 1/2$ und für die Erwartungswerte der Pauli-Matrizen gilt

$$\langle\sigma_x\rangle = \operatorname{tr}(\sigma_x\rho) = c_0 \operatorname{tr}\sigma_x + c_1 \operatorname{tr}\sigma_x^2 + c_2 \operatorname{tr}\sigma_x\sigma_y + c_3 \operatorname{tr}\sigma_x\sigma_z$$

mit $\sigma_x^2 = \mathbb{1}$, $\sigma_x\sigma_y = i\sigma_z$, $\sigma_x\sigma_z = -i\sigma_y$ und $\operatorname{tr}\sigma_i = 0$ folgt

$$\langle\sigma_x\rangle = 2c_1.$$

Analog erhält man

$$\langle\sigma_y\rangle = 2c_2 \quad \text{und} \quad \langle\sigma_z\rangle = 2c_3.$$

Führt man den *Bloch-Vektor*

$$\vec{p} = \begin{pmatrix} \langle\sigma_x\rangle \\ \langle\sigma_y\rangle \\ \langle\sigma_z\rangle \end{pmatrix} = \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} \in [-1,1]^3$$

ein, so lässt sich die Dichtematrix schreiben als

$$\rho = \frac{1}{2}\left(\mathbb{1} + \vec{p}\cdot\vec{\sigma}\right)$$

(mit $\vec{\sigma} := (\sigma_x, \sigma_y, \sigma_z)^T$). Explizit lautet ρ

$$\rho = \frac{1}{2}\begin{pmatrix} 1+p_z & p_x - ip_y \\ p_y + ip_y & 1-p_z \end{pmatrix}.$$

Damit dieses ρ eine Dichtematrix ist, muss außerdem $\operatorname{tr}\rho^2 \leq 1$ gelten, wobei Gleichheit für reine Zustände gilt. Mit der obigen Darstellung folgt

$$\begin{aligned} \operatorname{tr}\rho^2 &= \frac{1}{4}\left((1+p_z)^2 + p_x^2 + p_y^2 + (1-p_z)^2 + p_x^2 + p_y^2\right) \\ &= \frac{1}{2}(1 + p_x^2 + p_y^2 + p_z^2) = \frac{1}{2}(1 + |\vec{p}|^2) \leq 1. \end{aligned}$$

Damit muss $|\vec{p}|^2 \leq 1$ gelten. Die Zustände lassen sich durch die *Bloch-Kugel* visualisieren (vgl. Abbildung 2.1).

Allgemeine reine Zustände

Wir betrachten $\rho = (\mathbb{1} + \vec{p}\cdot\sigma)/2$ für reine Zustände, d.h. $|\vec{p}| = 1$. Eine mögliche Parametrisierung von $\vec{p}$ sind Kugelkoordinaten

$$\vec{p} = \begin{pmatrix} \sin\theta\cos\varphi \\ \sin\theta\sin\varphi \\ \cos\theta \end{pmatrix}.$$

Daraus ergeben sich die Eigenwerte von ρ zu

$$\frac{1}{2}\left(1 \pm |\vec{p}|\right) = \begin{cases} 0 \\ 1 \end{cases} \quad \text{für reine Zustände.}$$

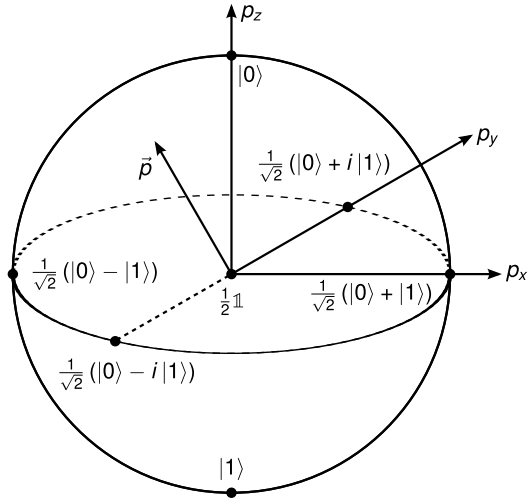


Abbildung 2.1: Visualisierung eines Zustandes $\vec{p}$ in der Blochkugel. Ist $|\vec{p}| = 1$, so ist der Zustand rein, für $|\vec{p}| < 1$ ist er gemischt.

Der Eigenvektor zum Eigenwert 1 ist gegeben durch

$$|\psi(\theta,\varphi)\rangle = \begin{pmatrix} \exp(-i\varphi/2)\cos(\theta/2) \\ \exp(i\varphi/2)\sin(\theta/2) \end{pmatrix}.$$

Damit kann man ρ als Projektor $\rho = |\psi(\theta,\varphi)\rangle\,\langle\psi(\theta,\varphi)|$ schreiben. $|\psi(\theta,\varphi)\rangle$ beschreibt einen „spin-up"-Zustand in $\vec{p}$-Richtung, denn es gilt

$$\begin{aligned}\left\langle \hat{\vec{n}}\cdot\vec{\sigma}\right\rangle &= \operatorname{tr}(\hat{\vec{n}}\cdot\sigma\rho) = \operatorname{tr}(n_x\sigma_x\rho) + \operatorname{tr}(n_y\sigma_y\rho) + \operatorname{tr}(n_z\sigma_z\rho)\\ &= n_x\operatorname{tr}(\sigma_x\rho) + n_y\operatorname{tr}(\sigma_y\rho) + n_z\operatorname{tr}(\sigma_z\rho)\\ &= n_x\langle\sigma_x\rangle + n_y\langle\sigma_y\rangle + n_z\langle\sigma_z\rangle = \hat{\vec{n}}\cdot\vec{p}.\end{aligned}$$

Insbesondere gilt $\langle\vec{p}\cdot\vec{\sigma}\rangle \equiv \left\langle\hat{\vec{p}}\cdot\vec{\sigma}\right\rangle = +1$. Der Eigenzustand zum Eigenwert 0 wird als *antipodaler Zustand* oder *spin-down* bezeichnet.

Reduzierte Dichtematrix

Wir betrachten eine Dichtematrix ρ auf $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ (z.B. erzeugt durch einen reinen Zustand auf $\mathcal{H}\otimes\mathcal{H}_C$) und einen Operator $O = O_A \otimes \mathbb{1}_B$. Für die Bestimmung des Erwartungswertes $\langle O\rangle = \operatorname{tr}(O\rho)$ wählen wir die Orthonormalbasis $|k\rangle_A$ in $\mathcal{H}_A$ und $|l\rangle_B$ in $\mathcal{H}_B$ und bilden die orthonormale Produktbasis $|i\rangle = |k\rangle_A\,|l\rangle_B$ in $\mathcal{H}$. i ist in diesem Fall ein Doppelindex, d.h. $\sum_i$ bedeutet $\sum_{k,l}$. Damit gilt für die Dichtematrix

$$\rho = \sum_{ij}\rho_{ij}\,|i\rangle\,\langle j| = \sum_{kl,mn}\rho_{kl,mn}\,|k\rangle_A\,|l\rangle_B\,{}_A\langle m|_B\langle n|$$

und es ergibt sich für den Erwartungswert

$$\begin{aligned}\langle O\rangle = \mathrm{tr}(O\rho) &= \sum_{kl,mn,rs} \rho_{kl,mn}\, {}_A\langle r|_B\langle s|O_A\otimes \mathbb{1}_B\,|k\rangle_A\,|l\rangle_B \underbrace{{}_A\langle m|_B\langle n|\,r\rangle_A\,|s\rangle_B}_{={}_A\langle m|r\rangle_{AB}\langle n|s\rangle_B=\delta_{mr}\delta ns} \\ &= \sum_{kl,mn} \rho_{kl,mn}\,{}_A\langle m\,|O_A|\,k\rangle_{AB}\langle n|\,l\rangle_B \\ &= \sum_{km}\underbrace{\sum_l \rho_{kl,ml}}_{(\rho_A)_{km}}\,{}_A\langle m|O_A\,|k\rangle_A = \mathrm{tr}(\rho_A O_A),\end{aligned}$$

wobei die letzte Spur über A auszuführen ist. Zusammengefasst gilt $\langle O\rangle \equiv \langle O_A\rangle = \mathrm{tr}(O_A\rho_A)$. Man definiert daher die *reduzierte Dichtematrix* $\rho_A = \mathrm{tr}_B\,\rho$ durch die partielle Spur tr_B, wobei die partielle Spur nur über die Basis von $\mathcal{H}_B$ summiert, d.h.

$$(\rho_A)_{km} := \sum_l \rho_{kl,ml}.$$

Analog ergibt sich

$$\rho_B = \mathrm{tr}_A\,\rho \text{ mit } (\rho_B)_{km} = \sum_l \rho_{lk,lm}.$$

Schmidt-Zerlegung (nach Erhard Schmidt, 1906)
Wir betrachten einen reinen Zustand $|\psi_{AB}\rangle \in \mathcal{H}_A\otimes\mathcal{H}_B$. Ist $\{|\alpha\rangle_A\}$ eine Orthonormalbasis von $\mathcal{H}_A$ und $\{|\beta\rangle_B\}$ eine Orthonormalbasis von $\mathcal{H}_B$ so lässt sich $|\psi_{AB}\rangle$ entwickeln als

$$|\psi_{AB}\rangle = \sum_{\alpha,\beta} c_{\alpha\beta}\,|\alpha\rangle_A\,|\beta\rangle_B = \sum_\alpha \underbrace{|\alpha_A\rangle}_{=:|i\rangle_A=|i=\alpha\rangle_A} \underbrace{\sum_\beta c_{\alpha\beta}\,|\beta\rangle_B}_{=:|j\rangle_B=|j=\alpha\rangle_B}.$$

1) $|\psi_{AB}\rangle = \sum_i |i\rangle_A\,|i\rangle_B$
   Bemerkung: Eine der Basen (in A oder B) kann frei (z.B. als Orthonormalbasis) gewählt werden, die andere im Allgemeinen nicht (diese ist i. A. keine Orthonormalbasis).

2) Zusammenhang mit Dichtematrix: Wähle $|i\rangle_A$ als Orthonormalbasis in $\mathcal{H}_A$ und berechne die partielle Spur

   $$\rho_A = \mathrm{tr}_B\,|\psi_{AB}\rangle\,\langle\psi_{AB}| = \sum_{ij} |i\rangle_A\,{}_A\langle j|\,\mathrm{tr}_B\,|i\rangle_B\,{}_B\langle j|.$$

   Es gilt $\mathrm{tr}_B\,|i\rangle_B\,{}_B\langle j| = \sum_k {}_B\langle k|\,i\rangle_{BB}\langle j|\,k\rangle$. Wählt man $|k\rangle_B$ so, dass ${}_B\langle k|\,i\rangle B = \delta_{ik}$, so folgt

   $$\rho_A = \sum_{ij} |i\rangle_A\,{}_A\langle j|_B\langle j|\,i\rangle_B.$$

   Nun wähle eine Eigenbasis von ρ_A als Orthonormalbasis $\{|i\rangle_A\}$ in $\mathcal{H}_A$. Damit gilt

   $$\rho_A = \sum_i p_i\,|i\rangle_A\,{}_A\langle i|,$$

   wobei p_i der Eigenwert zum Eigenvektor $|i\rangle_A$ ist. Daraus folgt ${}_B\langle j|\,i\rangle_B = p_i\delta_{ij}$. Für diese Wahl der Basis in A sind die Basiszustände $\{|i\rangle_B\}$ in $\mathcal{H}_B$ orthogonal, aber nicht normiert!

3) Normiere $|i'\rangle_B := \frac{1}{\sqrt{p_i}} |i\rangle_B$ und erhalte eine Orthonormalbasis.

Zusammengefasst erhalten wir mit

$$|\psi_{AB}\rangle = \sum_i \sqrt{p_i} \, |i\rangle_A \, |i'\rangle_B$$

die *Schmidt-Zerlegung* von $|\psi_{AB}\rangle$ bezüglich einer bestimmten (von $|\psi_{AB}\rangle$ abhängigen) Orthonormalbasis $|i\rangle_A$ und $|i'\rangle_B$. Zukünftig benutzen wir die Orthonormalbasis in B ebenfalls das Symbol $|i\rangle_B$.

Korollar (reduzierte Dichtematrix)**.** *Es gilt*

$$\begin{aligned}
\rho_A = \mathrm{tr}_B \, |\psi_{AB}\rangle \, \langle\psi_{AB}| &= \mathrm{tr}_B \sum_{ij} \sqrt{p_i p_j} \, \underbrace{|i\rangle_A \, |i\rangle_{B \; A}\langle j|_B \langle j|}_{|i\rangle_{A\,A}\langle j|} \otimes |i\rangle_{B \; B}\langle j| \\
&= \sum_{ij} \sqrt{p_i p_j} \, |i\rangle_{A \; A}\langle j| \underbrace{\mathrm{tr}_B \, |i\rangle_{B \; B}\langle i|}_{\delta_{ij}} \\
&= \sum_i p_i \, |i\rangle_{A \; A}\langle i|.
\end{aligned}$$

Analog erhält man $\rho_B = \mathrm{tr}_A \, |\psi_{AB}\rangle \, \langle\psi_{AB}| = \sum_i p_i \, |i\rangle_{B \; B}\langle i|$.
Die reduzierten Dichtematrizen ρ_A *und* ρ_B *haben die selben Eigenwerte* p_i*. Falls* $\dim \mathcal{H}_A \neq \dim \mathcal{H}_B$*, dann müssen die überzähligen Eigenwerte im größeren Raum gleich Null sein. Die Basis* $|i\rangle_B$ *in* H_B *ist eine Eigenbasis von* ρ_B *genauso wie* $|i\rangle_A$ *eine Eigenbasis von* ρ_A *(in* $\mathcal{H}_A$*) ist.*

Eindeutigkeit der Schmidt-Zerlegung

p_i ist eindeutig für gegebenes $|\psi_{AB}\rangle$. Im Fall ohne Entartung (d.h. alle p_i verschieden) sind $|i\rangle_A$ und $|i\rangle_B$ eindeutig bis auf eine Transformation

$$|i_A\rangle \to e^{i\varphi} \, |i\rangle_A \qquad\qquad |i\rangle_B \to e^{-i\varphi} \, |i\rangle_B \, .$$

Falls Entartung auftritt gibt es weitere Freiheiten, die wir hier aber nicht diskutieren.

Definition (Schmidt-Zahl): Sei $|\psi_{AB}\rangle$ gegeben. Daraus ergeben sich die Dichtematrizen $\rho_A = \mathrm{tr}_B \, |\psi_{AB}\rangle \, \langle\psi_{AB}|$ und $\rho_B = \mathrm{tr}_A \, |\psi_{AB}\rangle \, \langle\psi_{AB}|$. Dann ist die *Schmidt-Zahl* definiert als

$$\begin{aligned}
\mathrm{SZ} &:= \text{Anzahl der Eigenwerte ungleich Null von } \rho_A \\
&= \text{Anzahl der Eigenwerte ungleich Null von } \rho_B.
\end{aligned}$$

Definition (Verschränkung (engl. entanglement)): Ein reiner Zustand $|\psi_{AB}\rangle$ ist *verschränkt* falls $\mathrm{SZ} > 1$ oder äquivalent, falls $|\psi_{AB}\rangle$ nicht als Produkt $|\psi_{AB}\rangle = |\varphi\rangle_A \, |\chi\rangle_B$ dargestellt werden kann (denn sonst wäre $\mathrm{SZ} = 1$).

GHJW-Theorem (Gisin, Hughston, Josza, Wootters)

Die Ensemble-Interpretation der Dichtematrix ρ ist nicht eindeutig. Zum Beispiel ist für ein Qubit

$$\begin{aligned}
\rho &= \frac{1}{2}\mathbb{1}, \\
\rho &= \frac{1}{2} \, |0\rangle \, \langle 0| + \frac{1}{2} \, |1\rangle \, \langle 1| = \frac{1}{2} \, |+\rangle \, \langle +| + \frac{1}{2} \, |-\rangle \, \langle -|
\end{aligned}$$

mit $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. Dies lässt sich auf beliebigdimensionale Zustände verallgemeinern:

$$\begin{aligned}\rho_A &= \sum_{i=1}^{n_1} p_i^{(1)} \left|\psi_i^{(1)}\right\rangle \left\langle\psi_i^{(1)}\right| \\ &= \sum_{i=1}^{n_2} p_i^{(2)} \left|\psi_i^{(2)}\right\rangle \left\langle\psi_i^{(2)}\right|\end{aligned}$$

für zwei Ensembles $\{p_i^{(1)}, |\psi_i\rangle^{(1)}\}$ und $p_i^{(2)}, \left|\psi_i^{(2)}\right\rangle$.

Satz (GHJW)**.** *Es gibt eine Erweiterung von $\mathcal{H}_A$ auf $\mathcal{H}_A \otimes \mathcal{H}_B$ mit $\dim \mathcal{H}_B \equiv n = \max_i n_i$ und $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ mit $\rho_A = \mathrm{tr}_B |\psi\rangle\langle\psi|$ (Purification/Reinigung von ρ_A) und eine Observable $B^{(k)}$ auf $\mathcal{H}_B$ so, dass das k-te Ensemble $\{p_i^{(k)}, \left|\psi_i^{(k)}\right\rangle\}$ durch die Messung von $B^{(k)}$ realisiert wird:*

$$|\psi\rangle = \sum_i \sqrt{p_i^{(k)}} \left|\psi_i^{(k)}\right\rangle \left|\chi_i^{(k)}\right\rangle$$

mit $B^{(k)} = \sum_i \lambda_i^{(k)} \left|\chi_i^{(k)}\right\rangle \left\langle\chi_i^{(k)}\right|$, wobei $\lambda_i^{(k)} \neq \lambda_j^{(k)}$ für $i \neq j$ (keine Entartung).

Beweis. Schreibe zwei beliebige Ensemble-Zerlegungen auf:

$$\rho_A = \sum_i p_i |\psi_i\rangle\langle\psi_i| = \sum_i p_i' |\psi_i'\rangle\langle\psi_i'|$$

und $|\psi\rangle = \sum_i \sqrt{p_i}\,|\psi_i\rangle\,|\chi_i\rangle$ und $|\psi'\rangle = \sum_i \sqrt{p_i'}\,|\psi_i'\rangle\,|\chi_i'\rangle$, wobei $\{|\chi_i\rangle\}$ und $\{|\chi_i'\rangle\}$ beliebige Orthonormalbasen in $\mathcal{H}_B$ sind. Dann gilt

$$\rho_A = \mathrm{tr}_B |\psi\rangle\langle\psi| = \mathrm{tr}_B \left|\psi'\right\rangle\left\langle\psi'\right|.$$

Für die Schmidt-Zerlegung wähle eine Orthonormalbasis $\{|k_A\rangle\}$ auf $\mathcal{H}_A$ als Eigenbasis von ρ_A so, dass $|k_A\rangle$ Eigenvektor zum Eigenwert λ_k ist. Damit ergibt sich

$$\begin{aligned}|\psi\rangle &= \sum_k \sqrt{\lambda_k}\,|k_A\rangle\,|k_B\rangle\,, \\ \left|\psi'\right\rangle &= \sum_k \sqrt{\lambda_k}\,|k_A\rangle\,\left|k_B'\right\rangle\,,\end{aligned}$$

wobei $\{|k_B\rangle\}$ und $\{|k_B'\rangle\}$ Orthonormalbasen auf $\mathcal{H}_B$ sind, die aus der Schmidt-Zerlegung von $|\psi\rangle$ bzw. $|\psi'\rangle$ entstehen. Im Allgemeinen gilt $\{|k_B\rangle\} \neq \{|k_B\rangle'\}$. Da beides Orthonormalbasen sind, gibt es eine Basiswechselmatrix U_B mit

$$|k_B\rangle = U_B \left|k_B'\right\rangle.$$

Dadurch können wir $|\psi\rangle$ ausdrücken durch

$$|\psi\rangle = (\mathbb{1}_A \otimes U_B)\left|\psi'\right\rangle = \sum_i \sqrt{p_i'}\,\left|\psi_i'\right\rangle U_B \left|\chi_i'\right\rangle.$$

Definiert man $|\varphi_i\rangle := U_B |\chi_i'\rangle$ so ist $\{|\varphi_i\rangle\}$ eine Orthonormalbasis auf $\mathcal{H}_B$. Um das Ensemble $\{p_i, |\psi_i\rangle\}$ zu erhalten, misst man bezüglich $\{|\chi_i\rangle\}$ bzw. die Observable $B = \sum_i \lambda_i |\chi_i\rangle\langle\chi_i|$. Für das Ensemble $\{p_i', |\psi_i'\rangle\}$ misst man $\{|\varphi_i\rangle\}$ bzw. $B' = \sum_i \mu_i |\varphi_i\rangle\langle\varphi_i|$. □

2.3 Messung

Eine Observable A ist selbstadjungiert, d.h. $A^\dagger = A$ und lässt sich in die Spektraldarstellung $A = \sum_i a_i P_i$ überführen, wobei die $a_i \in \mathbb{R}$ die Eigenwerte bzw Messergebnisse von A sind und P_i die zugehörenden orthogonalen Projektoren. Diese haben die Eigenschaften

$$\begin{aligned} P_i P_j &= \delta_{ij} P_i & &\Rightarrow [P_i, P_j] = 0 \\ P_i^\dagger &= P_i & &\Rightarrow P_i \geq 0 \\ \sum_i P_i &= \mathbb{1}. \end{aligned}$$

Eine Messung von A entspricht einer Messung von allen P_i.

Modell für eine Messung (von Neumann)
Betrachte den Hilbertraum $\mathcal{H}_{\text{ges}} = \mathcal{H} \otimes \mathcal{H}_P$, wobei A und P_i Operatoren auf $\mathcal{H}$ sind und $\mathcal{H}_P$ den *Messapparat* (engl. *pointer*) darstellt. Ein allgemeiner Zustand $|\psi\rangle \in \mathcal{H}$ lässt sich in der Eigenbasis $\{|\psi_i\rangle\}$ von A entwickeln:

$$|\psi\rangle = \sum_i c_i |\psi_i\rangle$$

mit $c_i \in \mathbb{C}$. Für die Eigenzustände $|\psi_i\rangle$ gilt $P_j |\psi_i\rangle = \delta_{ij} |\psi_i\rangle$ und für eine vollständige Messung benötigen wir $\dim \mathcal{H}_P \equiv n \geq \dim \mathcal{H}$. Sei außerdem $|0\rangle, \ldots, |n\rangle$ eine Orthonormalbasis in $\mathcal{H}_P$. Der *Messprozess* läuft folgendermaßen ab:

1) Verschränkung von System und Messapparat

$$|\psi\rangle |0\rangle = \sum_i c_i |\psi_i\rangle |0\rangle \xmapsto{U = e^{itH}} \sum_i c_i |\psi_i\rangle |i\rangle$$

2) Messung in $\mathcal{H}_P$ in der Orthonormalbasis $\{|i\rangle\}$. Das Ergebnis i ergibt sich mit Wahrscheinlichkeit $|c_i|^2$ und man erhält den Zustand $|\psi_i\rangle |i\rangle$ (im Fall ohne Entartung).

Wie muss der Hamiltonian H aussehen, um das gewünschte Verhalten zu erzielen?

$$H = \lambda \sum_i P_i \otimes \Big(\underbrace{(|0\rangle \langle i| + |i\rangle \langle 0|)}_{\sigma_x^{(0i)}} + \sum_{j \neq 0,i} |j\rangle \langle j| \Big)$$

mit einer Kopplungskonstanten λ zwischen System und Messapparat. Daraus ergibt sich

$$\begin{aligned} H^2 &= \lambda^2 \sum_{ij} \underbrace{P_i P_j}_{\delta_{ij} P_i} \otimes \left(\sigma_x^{(0i)} + \sum_{k \neq 0,1} |k\rangle \langle k| \right) \left(\sigma_x^{(0j)} + \sum_{k \neq 0,1} |k\rangle \langle k| \right) \\ &= \lambda^2 \sum_i P_i \otimes \left(\sigma_x^{(0i)} \sigma_x^{(0i)} + \underbrace{\sigma_x^{(0i)} \sum_{k \neq i} |k\rangle \langle k|}_{=0} + \underbrace{\sum_{k \neq i} |k\rangle \langle k| \sigma_x^{(0i)}}_{=0} + \sum_{k \neq 0,i} \sum_{l \neq 0,i} |k\rangle \underbrace{\langle k| l\rangle}_{\delta_{kl}} |l\rangle \right) \\ &= \lambda^2 \sum_i P_i \otimes \left(\underbrace{\mathbb{1}^{(0i)}}_{|0\rangle\langle 0| + |i\rangle\langle i|} + \sum_{k \neq 0,i} |k\rangle \langle k| \right) = \lambda^2 \sum_i P_i \otimes \mathbb{1}. \end{aligned}$$

Analog erhält man $H^3 = HH^2 = \lambda^3 \sum_i P_i \sigma_x^{(0i)}$ und höhere Potenzen und somit für den Zeitentwicklungsoperator

$$U(t) = e^{-itH} = \sum_{n=0}^{\infty} \frac{(-it)^n}{n!} H^n = \ldots = \cos(\lambda t) \sum_i P_i \otimes \mathbb{1} + i \sin(\lambda t) \sum_i P_i \otimes \sigma_x^{(0i)}.$$

Für $\lambda t = \pi/2$ erhalten wir

$$U\left(\frac{\pi}{2\lambda}\right) = i \sum_i P_i \otimes \sigma_x^{(0i)}$$

und daraus

$$U\left(\frac{\pi}{2\lambda}\right) |\psi\rangle |0\rangle = \sum_i c_i \sum_j P_j \otimes \sigma_x^{(0j)} |\psi_i\rangle |0\rangle = \sum_i c_i \underbrace{\sigma_x^{(0i)} |0\rangle}_{=|i\rangle}.$$

Schwache Messung ($0 < \lambda t < \pi/2$)
Wähle den Hilbertraum $\mathcal{H}_P$ für die Messung mit $\dim \mathcal{H}_B \geq \dim \mathcal{H}_A + 1 \equiv n + 1$. Eine Orthonormalbasis in $\mathcal{H}_P$ sei gegeben durch $|0\rangle, |1\rangle, \ldots, |n\rangle$. $|0\rangle$ steht dabei für den Anfangszustand der Messapparatur und die Zustände $|1\rangle, \ldots, |n\rangle$ für die der Orthonormalbasis von $\mathcal{H}_A$ zugeordneten Zustände.
Wir betrachten eine Messung in $\mathcal{H}_B$ für $\varphi := \lambda t$

$$U(t) |\psi\rangle |0\rangle = \cos(\varphi) |\psi\rangle |0\rangle + i \sin\varphi \sum_{i=1}^{n} |\psi_i\rangle |i\rangle$$

Mit Wahrscheinlichkeit $p = \sin^2\varphi$ erhalten wir eine projektive Messung auf $\mathcal{H}_A$ mit einem Ergebnis $i \in \{1, 2, \ldots, n\}$ und zugehörigem Zustand $|\psi_i\rangle \in \mathcal{H}_A$.
Mit Wahrscheinlichkeit $1 - p = \cos^2\varphi$ passiert keine Messung und der Zustand $|\psi\rangle$ bleibt erhalten.

Beispiel: Wir betrachten ein Qubit, die Observable $A := \sigma_z$ mit den Projektoren $P_0 = |0\rangle\langle 0|$ und $P_1 = |1\rangle\langle 1|$. Es ergibt sich

mit Wahrscheinlichkeit p

$$|\psi\rangle \xmapsto{\langle\psi|P_0|\psi\rangle = |\langle\psi|0\rangle|^2} \frac{P_0 |\psi\rangle}{\langle\psi| P_0 |\psi\rangle} = |0\rangle,$$
$$|\psi\rangle \xmapsto{\langle\psi|P_1|\psi\rangle = |\langle\psi|1\rangle|^2} \frac{P_1 |\psi\rangle}{\langle\psi| P_1 |\psi\rangle} = |1\rangle,$$

mit Wahrscheinlichkeit $p - 1$

$$|\psi\rangle \xmapsto{1} \frac{\mathbb{1} |\psi\rangle}{\langle\psi| \mathbb{1} |\psi\rangle}.$$

Definiere die Operatoren

$$F_0 = p |0\rangle\langle 0|$$
$$F_1 = p |1\rangle\langle 1|$$
$$F_2 = (1-p)\mathbb{1}.$$

Messung auf gemischten Zuständen
Sei ρ eine Dichtematrix. Dann ergibt sich

mit Wahrscheinlichkeit p

$$\rho \xmapsto{\langle 0|\rho|0\rangle = \mathrm{tr}(P_0\rho)} \frac{P_0\rho P_0}{\mathrm{tr}(P_0\rho)} = \frac{|0\rangle\langle 0|\rho|0\rangle\langle 0|}{\langle 0|\rho\rangle 0} = |0\rangle\langle 0|,$$
$$\rho \xmapsto{\langle 1|\rho|1\rangle = \mathrm{tr}(P_1\rho)} \frac{P_1\rho P_1}{\mathrm{tr}(P_1\rho)},$$

mit Wahrscheinlichkeit $1-p$

$$\rho \xmapsto{\mathrm{tr}(\mathbb{1}\rho)=1} \frac{\mathbb{1}\rho\mathbb{1}}{\mathrm{tr}(\mathbb{1}\rho)} = \rho.$$

Mit den Operatoren $M_0 := \sqrt{F_0} = \sqrt{p}\,|0\rangle\langle 0|$, $M_1 := \sqrt{F_1} = \sqrt{p}\,|1\rangle\langle 1|$ und $M_2 = \sqrt{1-p}\mathbb{1}$ lässt sich dies schreiben als

$$\rho \mapsto \begin{cases} |0\rangle\langle 0| = \frac{M_0\rho M_0}{\mathrm{tr}(M_0^\dagger M\rho)} & \text{mit Wahrscheinlichkeit } \mathrm{tr}(M_0^\dagger M_0\rho) = \mathrm{tr}(F_0\rho) = p\langle 0|\rho|0\rangle, \\ |1\rangle\langle 1| = \frac{M_2\rho M_1}{\mathrm{tr}(M_1^\dagger M_2\rho)} & \text{mit Wahrscheinlichkeit } \mathrm{tr}(M_1^\dagger M_1\rho) = \mathrm{tr}(F_1\rho) = p\langle 1|\rho|1\rangle, \\ \rho = \frac{M_2\rho M_2}{\mathrm{tr}(M_2^\dagger M_2\rho)} & \text{mit Wahrscheinlichkeit } \mathrm{tr}(M_2^\dagger M_2\rho) = \mathrm{tr}(F_2\rho) = 1-p. \end{cases}$$

Bemerkung: Die F_a $(a = 0, 1, 2)$ haben fast die selben Eigenschaften wie die P_i bei projektiven Messungen:

$$F_a^\dagger = F_a, \qquad \sum_a F_a, = \mathbb{1} \qquad F_a \geq 0,$$

aber sie sind *nicht orthogonal*, d.h. im Allgemeinen ist $F_aF_b \neq \delta_{ab}F_a$.

Allgemeine Beschreibung von schwachen Messungen
Eine schwache Messung kann durch ein *positive operator valued measure* ($POVM$) beschrieben werden.

gemischter Zustand ρ_A auf $\mathcal{H}_A$	$\leftrightarrow$	reiner Zustand auf $\mathcal{H}_A \otimes \mathcal{H}_B$
verallgemeinerte Messung auf $\mathcal{H}_A$ (POVM)	$\leftrightarrow$	orthogonale (projektive) Messung auf $\mathcal{H}$ mit $\mathcal{H}_A \subseteq \mathcal{H}$, z.B. $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$.

Betrachte $\rho_A = \mathrm{tr}_B\,\rho_{AB}$ und wähle $\rho_{AB} = \rho_A \otimes \rho_B$ bezüglich einer festen Dichtematrix ρ_B. Orthogonale Messungen auf $\mathcal{H}_A \otimes \mathcal{H}_B$ sind gegeben durch

$$\rho_A \otimes \rho_B \mapsto \frac{P_a\rho_A \otimes \rho_B P_a}{\mathrm{tr}_{AB}(P_a\rho_A \otimes \rho_B)}$$

mit Wahrscheinlichkeit

$$\mathrm{tr}_{AB}(P_a\rho_A \otimes \rho_B) = \mathrm{tr}_A\,\mathrm{tr}_B(P_a\rho_A \otimes \rho_B).$$

Wir versuchen nun, $\mathrm{tr}_B(P_a\rho_A \otimes \rho_B) =: F_a\rho_A$ durch eine Matrixmultiplikation von F_a mit ρ_A darzustellen. Mit den Orthonormalbasen $|i\rangle_A$ und $|\mu\rangle_B$ haben wir komponentenweise

$$\sum_{ij} F_{a,ij}\rho_{A,ij} = \sum_{ij\mu\nu} P_{a,j\nu i\mu}\rho_{A,ij}\rho_{B,\mu\nu} \text{ für alle } \rho_A.$$

Hier wie im Folgenden bezeichnen lateinische Indizes Komponenten aus $\mathcal{H}_A$ und griechische aus $\mathcal{H}_B$. Durch Koeffizientenvergleich erhalten wir

$$F_{a,ij} = \sum_{\mu\nu} P_{a,j\nu i\mu}\rho_{B,\mu\nu}.$$

Die Anzahl der F_a ist beschränkt durch die Anzahl der P_a (Projektoren auf den Eigenwert a), welche gegeben ist durch $\dim \mathcal{H}_A \otimes \mathcal{H}_B = \dim \mathcal{H}_A \cdot \dim \mathcal{H}_B$. Diese kann viel größer sein als $\dim \mathcal{H}_A$.

Eigenschaften der F_a

(i) Hermitizität $F_a^\dagger = F_a$

(ii) Positivität $F_a \geq 0$

(iii) Vollständigkeit $\sum_a F_a = \mathbb{1}_A$

aber keine Orthogonalität. Ein Satz von Operatoren $\{F_a\}$ mit den Eigenschaften (i)-(iii) heißt *POVM*.

Unterschied zu orthogonalen (projektiven) Messungen

$F_a = M_a^\dagger M_a$ ist ein POVM-Element und $M_a := \sqrt{F_a}$ der zugehörige *Messoperator*.
Im Spezialfall orthogonaler Messungen ist $F_a = P_a, P_a^2 = P_a$, d.h. $M_a = \sqrt{P_a} = P_a$. Damit ist das POVM-Element gleichzeitig der Messoperator.
Der Zustand nach der Messung im Raum $\mathcal{H}_A$ ist gegeben durch

$$\rho'_A = \mathrm{tr}_B \frac{P_a\rho_A \otimes \rho_B P_a}{\mathrm{tr}(P_a\rho_A \otimes \rho_B)} = \rho'_A(\rho_a, \{F_a\}).$$

Die letzte Gleichheit soll andeuten, dass es keinen einfachen Zusammenhang für diese allgemeine Konstruktion der POVM gibt. Allerdings gibt es weniger optimale, aber dafür einfachere Konstruktionen (ähnlich der von-Neumann-Messung).

Konstruktion von POVM analog zu von-Neumann-Messungen

(i) Erweitere ρ_A in $\mathcal{H}_A$ auf $\rho_A \otimes \rho_B$ in $\mathcal{H}_A \otimes \mathcal{H}_B$, wobei $\rho_B = |0\rangle\langle 0|$ als reiner Zustand in B gewählt werden kann.

(ii) Kopple A und B durch eine unitäre Evolution in $\mathcal{H}_A \otimes \mathcal{H}_B$:

$$\rho = \rho_A \otimes |0\rangle\langle 0| \mapsto U\rho_A \otimes |0\rangle\langle 0|U^\dagger$$

(iii) Führe orthogonale Messung in $\mathcal{H}_B$ aus mit $P_a \equiv \mathbb{1} \otimes P_{B,a} = \mathbb{1} \otimes P_a$:

$$\rho \mapsto \frac{P_a U\rho_A \otimes |0\rangle\langle 0|U^\dagger P_a}{\mathrm{tr}_A \mathrm{tr}_B(P_a U\rho_A |0\rangle\langle 0|U^\dagger)}.$$

Die Wahrscheinlichkeit für den Messwert a ist gegeben durch

$$\operatorname{tr}_A \operatorname{tr}_B(P_a U \rho_A \otimes |0\rangle \langle 0|U^\dagger) = \operatorname{tr}_A F_a \rho_A.$$

Damit gilt

$$\sum_{ij} F_{a,ij}\rho_{A,ij} = \sum_{kij\mu\nu} P_{a,\nu\mu} U_{(k\mu)(i0)} {U^*}_{(k\nu)(j0)} \rho_{A,ij}.$$

Durch Koeffizientenvergleich erhält man

$$F_{a,ji} = \sum_{k\mu\nu} P_{a,\nu\mu} U_{(k\mu)(i0)} {U^*}_{(k\nu)(j0)}.$$

Wähle die Projektoren (bzw. Basis in $\mathcal{H}_B$) nun so, dass $P_{a,\nu\mu} = \mathbb{1}_A \otimes |a\rangle_B \, {}_B\langle a|$, d.h. $P_{a,\nu\mu} = \delta_{\nu\mu}\delta_{a\nu}$ gilt. Damit folgt

$$F_{a,ji} = \sum_k \underbrace{U_{(ka)(i0)}}_{=:M_{a,ki}} \underbrace{{U^*}_{(ka)(j0)}}_{=:(M_a^\dagger)_{jk}} = \sum_k M_{a,ki}(M_a^\dagger)_{jk},$$

d.h.

$$F_a = M_a^\dagger M_a.$$

Für den Zustand nach der Messung ergibt sich

$$\rho'_A = \frac{\operatorname{tr}_B(P_a U \rho_A \otimes |0\rangle \langle 0|U^\dagger P_a)}{\operatorname{tr}_A \operatorname{tr}_B(P_a U \rho_A \otimes |0\rangle \langle 0|U^\dagger)} = \frac{M_a \rho_A M_a^\dagger}{\operatorname{tr}(M_a^\dagger M_a \rho_A)},$$

wobei die Spur im letzten Term die Spur im Raum $\mathcal{H}_A$ ist.

Zusammenfassung POVM

$\{F_a\}$, $a = 1, \ldots, k$ wobei k größer als $\dim \mathcal{H}_A$ sein kann. Es gilt $F_a \geq 0, F_a^\dagger = F_a$ damit $F_a = M_a^\dagger M_a$ und $\sum_a F_a = \mathbb{1}$. Nach einer Messung erhält man den Zustand

$$\rho'_A = \frac{M_a \rho_A M_a^\dagger}{\operatorname{tr}(M_a^\dagger M_a \rho_A)}$$

mit der Wahrscheinlichkeit $\operatorname{tr}(M_a^\dagger M_a \rho_A)$.
Wozu sind POVM gut?

Beispiel: (siehe Nielsen & Chuang, [NC10]) Eine Quelle produziert entweder den Zustand $|0\rangle$ oder $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. Wegen $\langle 0| + \rangle \neq 0$ kann man die beiden Symbole nicht mit Sicherheit unterscheiden. Orthogonale Messungen in der Basis $\{|0\rangle, |1\rangle\}$ (oder $\{|+\rangle, |-\rangle\}$)ergeben bei Ausgang „0“ (oder „+“) kein schlüssiges Ergebnis.
Betrachte folgende POVM:

$$F_+ = \frac{\sqrt{2}}{1+\sqrt{2}} |1\rangle \langle 1|,$$
$$F_0 = \frac{\sqrt{2}}{1+\sqrt{2}} |-\rangle \langle -|,$$
$$F_x = \mathbb{1} - F_0 - F_+.$$

Beim präparierten Zustand $|0\rangle$ kommt nie das Ergebnis „+“ heraus, bei $|+\rangle$ kommt nie „0“ heraus. Der Preis dafür ist, dass es fälle gibt, in denen „x“ herauskommt, d.h. keine Entscheidung zwischen „0“ und „+“ fällt.
Falls $|0\rangle, |+\rangle$ jeweils mit Wahrscheinlichkeit 1/2 präpariert werden, dann liefert die orthogonale Messung in 1/4 der Fälle ein eindeutiges Ergebnis.
Für die POVM ist die Dichtematrix

$$\rho = \frac{1}{4}\begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix},$$

sodass sich mit Wahrscheinlichkeit

$$\operatorname{tr}(F_+\rho) = \frac{\sqrt{2}}{1+\sqrt{2}}\frac{1}{4}\operatorname{tr}\left(\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}\right) = \frac{\sqrt{2}}{4(1+\sqrt{2})}$$

das Ergebnis „+“ und mit Wahrscheinlichkeit

$$\operatorname{tr}(F_0\rho) = \frac{\sqrt{2}}{1+\sqrt{2}}\frac{1}{4}\operatorname{tr}\left(\frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}\right) = \frac{\sqrt{2}}{4(1+\sqrt{2})}$$

das Ergebnis „0“ eintrifft. Damit können wir mit Wahrscheinlichkeit

$$\operatorname{tr}(F_+\rho) + \operatorname{tr}(F_0\rho) = \frac{1}{2+\sqrt{2}} = 1 - \frac{1}{\sqrt{2}} \approx 0.29$$

die beiden Zustände sicher unterscheiden.

Neumark-Theorem (oder Naimark-Theorem)

Benannt nach Mark A. Naimark (1909-1978), russischer Mathematiker.
Wir wissen bereits, dass orthogonale Messungen auf erweitertem Hilbertraum auf POVM in $\mathcal{H}_A$ führen. Umkekehrt stellt sich die Frage, ob es für jedes POVM mit $\mathcal{H}_A$ eine Erweiterung von $\mathcal{H}_A$ und orthogonale Messungen gibt, welche dieses POVM erzeugen. Die Antwort ist „Ja“!

Theorem (Neumark)**.** *Jedes POVM $\{F_a\}$, $a = 1, \ldots, n$ auf $\mathcal{H}_A$ mit $N = \dim \mathcal{H}_A \leq n$ (wobei $\operatorname{rang} F_a = 1$ gelten soll) lässt sich als orthogonale Messung auf $\mathcal{H}$ mit $\mathcal{H}_A \subseteq \mathcal{H}_A = \mathcal{H}_A \oplus \mathcal{H}_A^\perp$ realisieren, wobei $\dim \mathcal{H} \geq n$.*

Beweis. Sei $F_a = \left|\tilde{\psi}_a\right\rangle\left\langle\tilde{\psi}_a\right|$ mit $\left|\tilde{\psi}_a\right\rangle$ nicht normiert. Aufgrund der Vollständigkeit gilt $\sum_{a=1}^n F_a = \mathbb{1}$. In Komponenten einer Orthonormalbasis $\{|i\rangle\}$ in $\mathcal{H}_A$, $i = 1, \ldots, N$ lautet dies

$$\sum_a F_{a,ij} = \sum_a \left\langle i\middle|\tilde{\psi}_a\right\rangle\left\langle\tilde{\psi}_a\middle| j\right\rangle = \delta_{ij}.$$

Dies ist eine $N \times n$-Matrix in deren Zeilen N orthogonale Vektoren in $\mathcal{H}$ stehen und in deren

Spalten n vollständige Vektoren in $\mathcal{H}_A$

$$\begin{array}{cccc}
\left\langle 1\middle|\tilde{\psi}_1\right\rangle & \left\langle 1\middle|\tilde{\psi}_2\right\rangle & \dots & \left\langle 1\middle|\tilde{\psi}_n\right\rangle \\
\left\langle 2\middle|\tilde{\psi}_1\right\rangle & \left\langle 2\middle|\tilde{\psi}_2\right\rangle & \dots & \left\langle 2\middle|\tilde{\psi}_n\right\rangle \\
\vdots & \vdots & \ddots & \vdots \\
\left\langle N\middle|\tilde{\psi}_1\right\rangle & \left\langle N\middle|\tilde{\psi}_2\right\rangle & \dots & \left\langle N\middle|\tilde{\psi}_n\right\rangle \\
\hline
\left\langle N+1\middle|\tilde{\psi}_1\right\rangle & \left\langle N+1\middle|\tilde{\psi}_2\right\rangle & \dots & \left\langle N+1\middle|\tilde{\psi}_n\right\rangle \\
\vdots & \vdots & \ddots & \vdots \\
\left\langle n\middle|\tilde{\psi}_1\right\rangle & \left\langle n\middle|\tilde{\psi}_2\right\rangle & \dots & \left\langle n\middle|\tilde{\psi}_n\right\rangle
\end{array}$$

Unterhalb der Linie steht die Vervollständigung zu einer $n \times n$-Matrix bzw. einer Orthonormalbasis von $\mathcal{H}_A$ durch $|u_a\rangle = \left|\tilde{\psi}_a\right\rangle + \left|\tilde{\psi}_a^{\perp}\right\rangle$ mit $U_{ij} = \langle i|\, u_j\rangle$ unitär.
Orthogonale Messungen auf $\mathcal{H}$ sind definiert durch

$$P_a = |u_a\rangle\langle u_a|$$

und nach der Messung erhält man den Zustand

$$\rho' = \frac{P_a \rho P_a}{\operatorname{tr}(P_a\rho)} \in \mathcal{H}.$$

Die Projektion auf $\mathcal{H}_A$ ist gegeben durch P_A. Damit

$$\tilde{\rho}'_A = P_A \rho' P_A = \frac{P_a p_a \rho P_a P_A}{\operatorname{tr}(P_a\rho)}$$

mit $\rho_A = P_a\rho P_A = P_A\rho_A P_A$ folgt

$$\begin{aligned}
\tilde{\rho}'_A &= \frac{P_A P_a P_A \rho_A P_A P_a P_A}{\operatorname{tr}(P_A\rho)} = \frac{F_a\rho_A F_a}{\operatorname{tr}(P_A\rho)} \\
&= \frac{\sqrt{F_a}\rho_A\sqrt{F_a}}{\operatorname{tr}(F_a\rho)}
\end{aligned}$$

mit

$$\operatorname{tr}(F_a\rho) = \operatorname{tr}(P_A P_a\rho) = \left\langle \tilde{\psi}_a\middle|\tilde{\psi}_a\right\rangle \operatorname{tr}(P_a\rho) \tag{2.1}$$

und

$$\sqrt{F_a} = \frac{\left|\tilde{\psi}_a\right\rangle\left\langle\tilde{\psi}_a\right|}{\sqrt{\left\langle\tilde{\psi}_a\middle|\tilde{\psi}_a\right\rangle}} = \frac{F_a}{\sqrt{\left\langle\tilde{\psi}_a\middle|\tilde{\psi}_a\right\rangle}}. \tag{2.2}$$

Damit ist

$$\rho'_A = \frac{\sqrt{F_a}\rho_A\sqrt{F_a}}{\operatorname{tr}(F_a\rho)} \overset{(2.2)}{=} \frac{1}{\left\langle\tilde{\psi}_a\middle|\tilde{\psi}_a\right\rangle} \frac{F_a\rho_A F_a}{\operatorname{tr}(F_a\rho)} \overset{(2.1)}{=} \frac{1}{\left\langle\tilde{\psi}_a\middle|\tilde{\psi}\right\rangle^2} \frac{F_a\rho_A F_a}{\operatorname{tr}(P_a\rho)}$$

der korrekte normierte Zustand in $\mathcal{H}_A$ nach der Messung. □

2.4 Zeitentwicklung

Wir betrachten den Produktraum $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. Der Anfangszustand auf $\mathcal{H}_A$ sei gegeben durch die Dichtematrix ρ_A und auf $\mathcal{H}_B$ durch einen beliebigen reinen Zustand, den Vektor $|0\rangle \in \mathcal{H}_B$. Der gesamte Anfangszustand in $\mathcal{H}$ ist dann gegeben durch $\rho := \rho_A \otimes |0\rangle\langle 0|$. Eine unitäre Zeitentwicklung in $\mathcal{H}$ führt zu

$$\rho' = U_{AB}\rho_A \otimes |0\rangle\langle 0|U_{AB}^\dagger.$$

Nach der Zeitentwicklung ist die reduzierte Dichtematrix in $\mathcal{H}_A$ gegeben durch

$$\rho'_A = \mathrm{tr}_B(U_{AB}\rho_A \otimes |0\rangle\langle 0|U_{AB}^\dagger) = \sum_\mu {}_B\langle\mu|U_{AB}|0\rangle_B \rho_A \, {}_B\Big\langle 0\Big|U_{AB}^\dagger\Big|\mu\Big\rangle_B,$$

wobei $\{|\mu\rangle\}$ eine Orthonormalbasis von $\mathcal{H}_B$ bezeichnet und ${}_B\langle\mu|U_{AB}|\mu\rangle_B$ den Operator auf $\mathcal{H}_A$ mit den Matrixelementen

$${}_A\langle i|_B\langle\mu|U_{AB}|0\rangle_B|j\rangle_A = {}_A\langle i|_B\langle\mu|U_{AB}|j\rangle_A|0\rangle_B.$$

Definiere $M_\mu := {}_B\langle\mu|U_{AB}|0\rangle_B$. Die M_μ sind Operatoren auf $\mathcal{H}_A$ und es gilt

$$\rho'_A = \sum_\mu M_\mu \rho_A M_\mu^\dagger =: \Lambda\rho_A.$$

Wegen $U_{AB}^\dagger U_{AB} = U_{AB}U^\dagger U_{AB} = \mathbb{1}_{AB}$ folgt $\sum_\mu M_\mu^\dagger M_\mu = \mathbb{1}_A$.
Die so definierte Abbildung Λ ist *linear* und bildet Operatoren in $\mathcal{H}_A$ wieder auf Operatoren in $\mathcal{H}_A$ ab. Man nennt solche Abbildungen Λ *Superoperatoren* . Die Darstellung durch die M_μ heißt *Operatorsummendarstellung* oder *Kraus-Darstellung* .
Weitere Eigenschaften von Λ folgen aus der Eigenschaft „Dichteoperatoren auf Dichteopratoren“ abzubilden, d.h.

(i) $(\rho'_A)^\dagger = \sum_\mu M_\mu \rho_A^\dagger M_\mu^\dagger = \rho'_A$, da $\rho_A^\dagger = \rho_A$.

(ii) $\mathrm{tr}\,\rho'_A = \sum_\mu \mathrm{tr}(\rho_A M_\mu^\dagger M_\mu) = 1$

(iii) $\rho'_A \geq 0$, denn $\langle\psi|\rho'_A|\psi\rangle = \sum_\mu \langle\psi|M_\mu\rho_A M_\mu^\dagger|\psi\rangle = \sum_\mu |\varphi_\mu\rangle\rho_A\langle\varphi_\mu| \geq 0$ mit $|\varphi_\mu\rangle := M_\mu^\dagger|\psi\rangle \in \mathcal{H}_A$ (bzw. $\langle\varphi_\mu| = \langle\psi|M_\mu$) für alle $|\psi\rangle \in \mathcal{H}_A$, da ρ_A positiv semidefinit ist ($\langle\varphi|\rho_A|\varphi\rangle \geq 0$ für alle $|\varphi\rangle \in \mathcal{H}_A$).

Umgekehrt können wir zu gegebenem $\{M_\mu\}$ immer ein $\mathcal{H}_B$ mit $\dim\mathcal{H}_B \geq \dim\mathcal{H}_A$ und U_{AB} unitär auf $\mathcal{H}_A \otimes \mathcal{H}_B$ finden, sodass damit der durch $\{M_\mu\}$ definierte Superoperator realisiert wird (analog zum Naimark-Theorem für POVMs).

Beweis. Definiere $U_{AB}(|\varphi\rangle_A \otimes |0\rangle_B) := \sum_\mu M_\mu|\varphi\rangle_A \otimes |\mu\rangle_B$. Man kann zeigen, dass U_{AB} Skalarprodukterhaltend ist und deshalb auf eine unitäre Operation erweitert werden kann. □

Nicht-Eindeutigkeit der Kraus-Darstellung

Bilde die partielle Spur in der Basis $\{|\nu\rangle\}$ statt $\{|\mu\rangle\}$. Dann gilt

$${}_B\langle\nu| = \sum_\mu U_{\nu\mu}\,{}_B\langle\mu|$$

und deshalb

$$\Lambda\rho_A = \sum_\nu N_\nu \rho_A N_\nu^\dagger$$

mit $N_\nu = U_{\nu\mu} M_\mu$, wobei $N_\nu = {}_B\langle\nu|U_{AB}\,|0\rangle_B$ verwendet wurde.
Wir behaupten, dass alle Darstellungen eines Superoperators auf diese Art zusammenhängen.

POVM als Superoperator

1) Nicht selektive Messung (keine bzw. vor Messung in $\mathcal{H}_B$):

$$\rho \mapsto \sum_\mu \sqrt{F_\mu}\rho\sqrt{F_\mu} = \sum_\mu M_\mu \rho M_\mu^\dagger = \Lambda\rho$$

mit $\sum_\mu M_\mu M_\mu^\dagger = \mathbb{1}$ und Λ stellt einen Superoperator dar.

2) Selektive Messung (nach Messung in $\mathcal{H}_B$)

$$\rho \mapsto \frac{\sqrt{F_\mu}\rho\sqrt{F_\mu}}{\mathrm{tr}(F_\mu\rho)}.$$

Diese Messung kann nicht durch einen Superoperator ausgedrückt werden, da der Zusammenhang nicht linear in ρ ist.

Forderungen an einen Superorperator Λ, welcher Dichtematrizen auf Dichtematrizen abbildet:

(0) Linearität

(1) Hermizitätserhaltung: $(\Lambda\rho)^\dagger = \Lambda\rho$, falls $\rho = \rho^\dagger$

(2) Spurerhaltung $\mathrm{tr}\,\Lambda\rho = 1$, falls $\mathrm{tr}\,\rho = 1$

(3) Positivität $\Lambda\rho \geq 0$, falls $\rho \geq 0$

Reichen diese Bedingungen, um eine Kraus-Darstellung (d.h. eine Darstellung aus unitären Operatoren im größeren Hilbertraum) zu garantieren? Nein, man muss noch etwas mehr fordern (stärkere Version von (3))

(3') *vollständige Positivität* (*complete positivity*): Für alle $\mathcal{H}_B$ und ρ auf $\mathcal{H}_A \otimes \mathcal{H}_B$ gilt $(\Lambda \otimes \mathbb{1}_B)\rho \geq 0$, falls $\rho \geq 0$

Beispiel (positive, aber nicht vollständig positive Abbildungen): Wir betrachten die *partielle Transposition*. Für $\rho \geq 0$ gilt $T\rho := \rho^T \geq 0$, da ρ^T die selben Eigenwerte wie ρ besitzt. Dies ist eine komplette Transposition.
Aber für den Bell-Zustand $\rho = |\varphi_+\rangle\langle\varphi_+|$ mit $|\varphi_+\rangle := (|00\rangle + |11\rangle)$ ($|00\rangle := |0\rangle_B\,|0\rangle_B$) gilt

$$\begin{aligned}
\rho &= \frac{1}{2}\left(|00\rangle\langle 00| + |11\rangle\langle 11| + |00\rangle\langle 11| + |11\rangle\langle 00|\right) \\
&= \frac{1}{2}\left(|0\rangle_B\,{}_B\langle 0| \otimes |0\rangle_A\,{}_A\langle 0| + |1\rangle_B\,{}_B\langle 1| \otimes |1\rangle_A\,{}_A\langle 1|\right. \\
&\qquad \left. + |0\rangle_B\,{}_B\langle 1| \otimes |0\rangle_A\,{}_A\langle 1| + |1\rangle_B\,{}_B\langle 0| \otimes |1\rangle_A\,{}_A\langle 0|\right) \\
&= \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.
\end{aligned}$$

Nach partieller Transformation im Untersystem $\mathcal{H}_B$ gilt

$$(T_B \otimes \mathbb{1}_A)\rho = \frac{1}{2}\left(|0\rangle_B\,{}_B\langle 0| \otimes |0\rangle_A\,{}_A\langle 0| + |1\rangle_B\,{}_B\langle 1| \otimes |1\rangle_A\,{}_A\langle 1| \right.$$
$$\left. + |\mathbf{1}\rangle_B\,{}_B\langle \mathbf{0}| \otimes |0\rangle_A\,{}_A\langle 1| + |\mathbf{0}\rangle_B\,{}_B\langle \mathbf{1}| \otimes |1\rangle_A\,{}_A\langle 0|\right)$$

Diese Matrix hat den dreifachen Eigenwert $1/2$ und den Eigenwert $-1/2$, ist also keine Dichtematrix!

Theorem (Kraus-Darstellung)
Eine Abbildung $\Lambda : \rho \mapsto \Lambda\rho$ besitzt eine Kraus-Darstellung

$$\Lambda\rho = \sum_\mu M_\mu \rho M_\mu^\dagger$$

mit $\sum_\mu M_\mu^\dagger M_\mu = \mathbb{1}$, falls (0), (1), (2) und (3') erfüllt sind.
Bemerkung: Damit existiert auch U_{AB} unitär, welches Λ unter partieller Spurbildung erzeugt.
Führe zunächst die Methode des relativen Zustandes ein:

Methode des relativen Zustands
Sei $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ ($\dim \mathcal{H}_B = \dim \mathcal{H}_A \equiv N$) mit der Orthonormalbasis $\{|i\rangle_A\}$ in $\mathcal{H}_A$ und $\{|i'\rangle_B\}$ in $\mathcal{H}_B$. Definiere den Zustand $\left|\tilde{\psi}\right\rangle = \sum_{i=1}^N |i\rangle_A\,|i'\rangle_B$. Dieser ist nicht normiert auf 1, denn $\left\langle \tilde{\psi}\middle|\,\tilde{\psi}\right\rangle = N$.
Sei $|\varphi\rangle_A := \sum_i a_i\,|i\rangle \in \mathcal{H}_A$ gegeben. Wir möchten $M_A\,|\varphi\rangle_A$ für einen Operator M_A auf $\mathcal{H}_A$ angeben:

1) Es gilt $|\varphi\rangle_A = {}_B\left\langle \varphi^*\middle|\,\tilde{\psi}\right\rangle$ (*partielles Skalarprodukt*) mit dem *relativen Zustand* $|\varphi^*\rangle_B := \sum_i a_i^*\,|i'\rangle_B$
2) $M_A\,|\varphi\rangle_A = {}_B\langle\varphi^*\,|M_A \otimes \mathbb{1}_B|\,\tilde{\varphi}\rangle$

Methode des relativen Zustands für Superoperatoren
Betrachte $\left|\tilde{\psi}\right\rangle := \sum_i |i\rangle_A\,|i'\rangle_B$. Die zugehörige Dichtematrix ist gegeben durch

$$\tilde{\rho} = \left|\tilde{\psi}\right\rangle\left\langle\tilde{\psi}\right| = \sum_{ij} |i\rangle_A\,{}_A\langle j| \otimes \left|i'\right\rangle_B\,{}_B\left\langle j'\right|.$$

Der Spuroperator Λ wirkt nur auf $\mathcal{H}_A$ mit

$$(\Lambda \otimes \mathbb{1})\tilde{\rho} = \sum_{ij} \Lambda(|i\rangle_A\,{}_A\langle j|) \otimes \left|i'\right\rangle_B\,{}_B\left\langle i'\right|.$$

Sei $|\varphi\rangle_A = \sum_i a_i\,|i\rangle_A$ und $|\varphi^*\rangle_B = \sum_i a_i^*\,|i'\rangle_B$. Damit

$${}_B\langle\varphi^*|(\Lambda \otimes \mathbb{1})\tilde{\rho}\,|\varphi^*\rangle_B = \sum_{ij} a_i a_j^* \Lambda(|i\rangle_A\,{}_A\langle j|) = \Lambda(|\varphi\rangle_A\,{}_A\langle\varphi|).$$

Beweis des Kraus-Darstellungs-Theorems. Aus (3') folgt: Λ ist vollständig positiv und damit $\Lambda \otimes \mathbb{1}$ positiv, d.h.

$$(\Lambda \otimes \mathbb{1}) \left|\tilde{\psi}\right\rangle \left\langle\tilde{\psi}\right| = \sum_\mu q_\mu \left|\phi_\mu\right\rangle \left\langle\phi_\mu\right| \geq 0$$

mit $q_\mu \geq 0$.
Die Methode des relativen Zustands liefert

$$\Lambda \left|\varphi\right\rangle_A {}_A\langle\varphi| = {}_B\left\langle\varphi^* \right| \Lambda \otimes \mathbb{1} \left|\tilde{\psi}\right\rangle \left\langle\tilde{\psi}\right| \left.\varphi^*\right\rangle_B = \sum_\mu q_\mu {}_B\langle\varphi^*|\,\phi_\mu\rangle\langle\phi_\mu|\,\varphi^*\rangle_B.$$

Definiere $M_\mu : \mathcal{H}_A \to \mathcal{H}_A$, $|\varphi\rangle_A \mapsto \sqrt{q_\mu}{}_B\langle\varphi^*|\,\phi_\mu\rangle$.
Rechne die Eigenschaften von M_μ nach und erhalte

(i) M_μ ist linear, denn $|\varphi_A\rangle_B \mapsto |\varphi^*\rangle_B$ ist antilinear

(ii) $\Lambda \left|\varphi\right\rangle_A {}_A\langle\varphi| = \sum_\mu M_\mu \left|\varphi\right\rangle_A {}_A\langle\varphi| M_\mu^\dagger$ für alle reinen Zustände in $\mathcal{H}_A$

(iii) Ist ρ_A ein gemischter Zustand in $\mathcal{H}_A$, dann ist $\rho_A = \sum_i p_i \left|\varphi_i\right\rangle \left\langle\varphi_i\right|$. Wegen der Linearität von Λ und M_μ folgt dann

$$\Lambda\rho_A = \sum_\mu M_\mu \rho_A M_\mu^\dagger.$$

(iv) $\mathrm{tr}(\Lambda\rho_A) = \mathrm{tr}(\rho_A)$, $\sum_\mu M_\mu^\dagger M_\mu = \mathbb{1}$

□

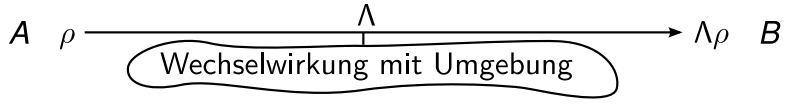
Weitere Folgrungen

(i) Die Anzahl der M_μ ist gegeben durch die Anzahl der $|\phi_\mu\rangle$, welche gegeben ist durch den Rang des Operators $(\Lambda \otimes \mathbb{1}) \left|\tilde{\psi}\right\rangle \left\langle\tilde{\psi}\right|$, der kleiner gleich N^2 ist, wobei $N = \dim \mathcal{H}_A$.

(ii) Die Kraus-Operatoren in den verschiedenen Darstellungen sind unitär verbunden, d.h.

$$N_a = \sum_\mu U_{a\mu} M_\mu.$$

Beispiele von Superoperatoren
Superoperatoren beschreiben die Zeitentwicklung von einem Qubit. Man spricht dabei auch von *Quantenkanälen* .



Beispiel: Depolarisationskanal (depolarization channel)
Bei der Übertragung passiert ein Fehler mit Wahrscheinlichkeit p. Für die folgenden Fehler sei die Wahrscheinlichkeit gleich groß:

Bit-Fehler:	$\|0\rangle \mapsto \|1\rangle$ $\|1\rangle \mapsto \|0\rangle$	$\|\varphi\rangle \mapsto \sigma_x \|\varphi\rangle =: \sigma_1 \|\varphi\rangle$
Phasenfehler:	$\|0\rangle \mapsto \quad \|0\rangle$ $\|1\rangle \mapsto -\|1\rangle$	$\|\varphi\rangle \mapsto \sigma_z \|\varphi\rangle =: \sigma_3 \|\varphi\rangle$
Bit- und Phasenfehler:	$\|0\rangle \mapsto \quad i\|1\rangle$ $\|1\rangle \mapsto -i\|0\rangle$	$\|\varphi\rangle \mapsto \sigma_y \|\varphi\rangle =: \sigma_2 \|\varphi\rangle$

Darstellung

(i) unitäre Darstellung
Wir erweitern den Hilbertraum $\mathcal{H}_A$, der das Qubit beschreibt zu $\mathcal{H}_A \otimes \mathcal{H}_E$, wobei $\mathcal{H}_E$ der ein vierdimensionaler Hilbertraum ist, der die Umgebung (environment) beschreibt. Eine Orthonormalbasis ist gegeben durch $|0\rangle_E, |1\rangle_E, |2\rangle_E$ und $|3\rangle_E$. Wir definieren den Operator

$$U_{AE} : |\varphi\rangle_A |0\rangle_E \mapsto |\psi\rangle_{AE},$$

$$|\psi\rangle_{AE} := \sqrt{1-p}\,|\varphi\rangle_A |0\rangle_E + \sqrt{\frac{p}{3}}\left(\sigma_1 |\varphi\rangle_A |1\rangle_E + \sigma_2 |\varphi\rangle_A |2\rangle_E + \sigma_3 |\varphi\rangle_A |3\rangle_E\right),$$

der unitär auf ganz $\mathcal{H}_A \otimes \mathcal{H}_E$ fortgesetzt werden kann. Die Dichtematrix ist dann gegeben durch

$$\rho'_A = \mathrm{tr}_E\, |\psi\rangle_{AE} \langle\psi|.$$

(ii) Kraus-Darstellung

$$\rho'_A = (1-p)\,|\varphi_A\rangle\langle\varphi_A| + \frac{p}{3}\left(\sigma_1\rho_A\sigma_1 + \sigma_2\rho_A\sigma_2 + \sigma_3\rho_A\sigma_3\right) = \sum_{\mu=0}^{3} M_\mu \rho_A M_\mu^\dagger$$

mit $M_\mu = {}_E\langle\mu\,|U_{AE}|\,0\rangle_E$. Durch Vergleich ergibt sich

$$M_0 = \sqrt{1-p}\mathbb{1}_A =: \sqrt{1-p}\sigma_0 \qquad M_i = \sqrt{\frac{p}{3}}\sigma_i \quad (i = 1,2,3).$$

Man rechnet leicht nach, dass

$$\sum_\mu M_\mu^\dagger M_\mu = \mathbb{1}_A$$

gilt, da $\sigma_i^2 = \mathbb{1}_A$.

(iii) Darstellung durch relativen Zustand
Sei $\left|\tilde{\psi}\right\rangle := |\Phi_+\rangle := (|0\rangle|0\rangle + |1\rangle|1\rangle)/\sqrt{2} \in \mathcal{H}_A \otimes \mathcal{H}_E$ mit $\dim \mathcal{H}_E = 2$. Definiere

$$|\psi_i\rangle := \sigma_i \otimes \mathbb{1}\,|\Phi_+\rangle = \begin{cases} |\Phi_+\rangle, & i = 0, \\ \frac{1}{\sqrt{2}}(|1\rangle|0\rangle + |0\rangle|1\rangle) =: |\Psi_+\rangle, & i = 1, \\ \frac{i}{\sqrt{2}}(|1\rangle|0\rangle - |0\rangle|1\rangle) =: |\Psi_-\rangle, & i = 2, \\ \frac{1}{\sqrt{2}}(|0\rangle|0\rangle - |1\rangle|1\rangle) =: |\Phi_-\rangle, & i = 3. \end{cases}$$

Dann gilt

$$\sigma_i |\varphi\rangle_A = {}_B\langle\varphi^*|\,\psi_i\rangle,$$

wobei $|\varphi\rangle_A = \sum_i a_i |i\rangle_A$ und $|\varphi^*\rangle_B = \sum_i a_i^* |i\rangle_B$.

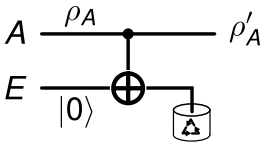
(iv) Darstellung auf der Bloch-Kugel (vgl. Übung)

$$\rho_A = \frac{1}{2}(\mathbb{1} + \vec{p}\cdot\vec{\sigma}) \mapsto \rho'_A = \frac{1}{2}(\mathbb{1} + \vec{p}'\cdot\vec{\sigma})$$

Beispiel: Phasendämpfungskanal (Dephasierungskanal)
(Die Herleitung ist anders als in [Pre]). Für die Umgebung reicht ein Hilbertraum $\mathcal{H}_E$ mit $\dim \mathcal{H}_E = 2$. Definiere

$$\begin{aligned} U_{AE} : |0\rangle_A |0\rangle_E &\mapsto |0\rangle_A |0\rangle_E \\ |1\rangle_A |0\rangle_E &\mapsto \sqrt{1-p}\,|1\rangle_A |0\rangle_E + \sqrt{p}\,|1\rangle_A |1\rangle_E \,. \end{aligned}$$

Schematisch gilt (für $p = 1$):



$\rho'_A = \mathrm{tr}_E(U_{AE}\rho_A \otimes |0\rangle_E\,{}_E\langle 0|U_{AE}^\dagger) = \sum_{\mu=0}^{1} M_\mu^\dagger \rho_A M_\mu$
Die Kraus-Operatoren sind gegeben durch

$$M_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix} \qquad\qquad M_1 = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{p} \end{pmatrix}.$$

Man sieht sofort, dass $M_0^2 + M_1^2 = \mathbb{1}$ gilt. Daraus ergibt sich

$$\rho'_A = \Lambda \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} = M_0^\dagger \rho_A M_0 + M_1^\dagger \rho_A M_1 = \begin{pmatrix} \rho_0 0 & \sqrt{1-p}\rho_{01} \\ \sqrt{1-p}\rho_{10} & \rho_{11} \end{pmatrix}.$$

(In [Pre] findet man $1-p$ als Faktor vor den Nichtdiagonaleinträgen, was in seiner Herleitung ebenfalls korrekt ist.)
Wir betrachten nun eine infinitesimale Zeitentwicklung. Dafür nehmen wir $p = 2\Gamma\Delta t \ll 1$ an mit der Übergangsrate $\Gamma = p/2\Delta t$. (Der Faktor 2 unterscheidet sich von [Pre]). Weiterhin sei $\Delta t = t/n$ mit $n \gg 1$, d.h. wir teilen das Zeitintervall von 0 bis t in n Stücke, wenden jeweils die oben hergeleitete Entwicklung an und erhalten

$$\rho'_A = \Lambda^n \rho_A = \begin{pmatrix} \rho_{00} & (1-p)^{n/2} \\ (1-p)^{n/2}\rho_{10} & \rho_{11} \end{pmatrix}.$$

Wegen

$$(1-p)^{n/2} = (1 - 2\Gamma t/n)^{n/2} \xrightarrow[n\to\infty]{\Delta t\to 0} e^{-\Gamma t}$$

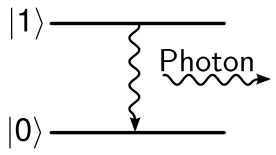
folgt

$$\rho_A(t) = \begin{pmatrix} \rho_{00} & e^{-\Gamma t}\rho_{01} \\ e^{-\Gamma t}\rho_{10} & \rho_1 1 \end{pmatrix} \xrightarrow{t\to\infty} \begin{pmatrix} \rho_{00} & 0 \\ 0 & \rho_{11} \end{pmatrix}.$$

Diese zeitabhängige Dichtematrix beschreibt den kontinuierlichen Übergang von $\alpha|0\rangle_A + \beta|1\rangle_A$ zu $|\alpha|^2 |0\rangle_A\,{}_A\langle 0| + |\beta|^2 |1\rangle_A\,{}_A\langle 1|$ ($|\alpha|^2 = \rho_{00}$, $|\beta|^2 = \rho_{11}$), d.h. die Phasenbeziehung zwischen α und β und somit die Interferenzfähigkeit zerfällt exponentiell mit der Zeit mit der Zeitkonstanten $\tau = 1/\Gamma$.

Bemerkung: Für den exponentiellen Zerfall wurde oben eine wichtige Annahme getroffen: Nach jedem Zeitschritt Δt sei die Umgebung auf den Zustand $|0\rangle_E$ zurückgesetzt. Die Umgebung hat also kein „Gedächtnis“. Dies bezeichnet man als *Markov-Näherung* (vgl. Fermis Goldene Regel).

Beispiel: Amplitudendämpfungskanal
Wir betrachten den Zerfall eines Zustandes unter Emission eines Photons



Ist die Zerfallswahrscheinlichkeit durch p gegeben, so gelten folgende Übergangsregeln

$$\begin{aligned}|0\rangle\,|0\rangle &\mapsto |0\rangle\,|0\rangle\\ |1\rangle\,|0\rangle &\mapsto \sqrt{1-p}\,|1\rangle\,|0\rangle + \sqrt{p}\,|0\rangle\,|1\rangle\,.\end{aligned}$$

Die Kraus-Operatoren sind Übung.
Daraus ergibt sich der Superorperator durch die Abbildungsvorschrift

$$\rho_A \mapsto \Lambda\rho_A = \begin{pmatrix}\rho_{00}+p\rho_{11} & \sqrt{1-p}\rho_{01}\\ \sqrt{1-p}\rho_{10} & (1-p)\rho_{11}\end{pmatrix}.$$

Wie oben erhalten wir unter Verwendung der Markov-Näherung

$$\rho_{11}(t) = e^{-\Gamma t}\rho_{11}(0)$$

und damit

$$\rho_A(t) \xrightarrow{t\to\infty} \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} = |0\rangle\,\langle 0|.$$

Detektion der Photons
Messe in der Basis $|0\rangle_E$, $|1\rangle_E$ (kein/ein Photon in der Umgebung). Dies definiert ein POVM auf $\mathcal{H}_A$ (orthogonale Messung in $\mathcal{H}_E$ bzw. $\mathcal{H}_A \otimes \mathcal{H}_E$) und es gilt

$$|\varphi\rangle_A\,|0\rangle_E \mapsto \sum_\mu M_\mu\,|\varphi\rangle_A\,|\mu\rangle_E$$

mit den Kraus-Operatoren M_μ. Das POVM ist gegeben durch $F_\mu = M_\mu^\dagger M_\mu$, konkret

$$F_0 = \begin{pmatrix}1 & 0\\ 0 & 1-p\end{pmatrix}, \qquad F_1 = \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}.$$

Auf der Zeitskala $t \gg 1/\Gamma$, d.h. $t \to \infty$ bzw. $p \to 1$ erhalten wir wieder eine projektive/orthogonale Messung, denn

$$F_0 \to \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} \qquad F_1 \to \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}.$$

Infinitesimale Zeitentwicklung

Bisher haben wir die Zeitentwicklung der Dichtematrizen $\rho(t) = \Lambda\rho(0)$ durch den Superoperator Λ analog zur Zeitentwicklung der reinen Zustände $|\psi(t)\rangle = U(t)\,|\psi(0)\rangle$ durch eine unitären Operator U betrachtet. Für die reinen Zustände haben wir aber auch die Zeitentwicklung durch die Schrödingergleichung $i\partial_t\,|\psi(t)\rangle = H\,|\psi(t)\rangle$, in der die physikalischen Zusammenhänge durch den Hamiltonoperator H gegeben sind. Wir versuchen nun eine ähnliche Formulierung für die Dichtematrizen zu finden. Das übliche Vorgehen ist, die Dichtematrix

$$\rho(t) = \mathrm{tr}_E\,|\psi(t)\rangle\,\langle\psi(t)|$$

als Spur eines reinen Zustandes $|\psi(t)\rangle \in \mathcal{H}_S \otimes \mathcal{H}_E$ im Produktraum von System und Umgebung zu betrachten. Für $|\psi(t)\rangle$ gilt im (abgeschlossenen) Gesamtsystem die Schrödingergleichung

$$i\partial_t\,|\psi(t)\rangle = H\,|\psi(t)\rangle\,.$$

Der Hamiltonoperator des Gesamtsystems setzt sich zusammen aus

$$H = H_S + H_E + H_{SE},$$

wobei $H_S = H_S \otimes \mathbb{1}$ der Hamiltonoperator des Systems, $H_E = \mathbb{1} \otimes H_E$ der Hamiltonoperator der Umgebung und H_{SE} der Hamiltonoperator der Wechselwirkung zwischen System und Umgebung ist. H_{SE} hat im Allgemeinen die Darstellung $H_{SE} = \sum_i A_S^i \otimes B_E^i$, oft auch nur $H_{SE} = A_S \otimes B_E$.

Bemerkung: Wählt man $\rho_E(0)$ fest, dann gilt $\mathrm{tr}_E(B_E\rho_E(0)) = 0$.

Trivialer Fall: $H_{SE} = 0$
In diesem Fall kann man annehmen, dass $|\psi(0)\rangle = |\psi_S(0)\rangle \otimes |\psi_E(0)\rangle$ gilt. Mit dem Separationsansatz $|\psi(t)\rangle = |\psi_S(t)\rangle \otimes |\psi_E(t)\rangle$ folgt

$$i\partial_t\,|\psi_S(t)\rangle = H_S\,|\psi_S(t)\rangle\,.$$

Daraus ergibt sich für $\rho(t) = |\psi_S(t)\rangle\,\langle\psi_S(t)|$

$$\dot\rho(t) = -iH_S\,|\psi_S(t)\rangle\,\langle\psi_S(t)| + i\,|\Psi_S(t)\rangle\,\langle\psi_S(t)|H_S,$$

d.h.

$$\dot\rho = -i\,[H_S, \rho]\,.$$

Diese Gleichung nennt man *Liouville-Gleichung*. Sie entspricht der Schrödingergleichung für Dichtematrizen.
Nicht-trivialer Fall: $H_{SE} \neq 0$
Hier gilt

$$\begin{aligned}\dot\rho &= \partial_t\rho = \partial_t\,\mathrm{tr}_E\,|\psi\rangle\,\langle\psi| = \mathrm{tr}_E\,\partial_t\,|\psi\rangle\,\langle\psi| \\ &= -i\,\mathrm{tr}_E\,[H, |\psi\rangle\,\langle\psi|] = -i\,\mathrm{tr}_E\,[H_S, |\psi\rangle\,\langle\psi|] - i\,\mathrm{tr}_E\,[H_E, |\psi\rangle\,\langle\psi|] - i\,[H_{SE}, |\psi\rangle\,\langle\psi|]\,.\end{aligned}$$

Im ersten Summand kann die partielle Spur in den Kommutator gezogen werden, da $H_S = H_S \otimes \mathbb{1}_E$ nur in S wirkt. Da die Spurbildung zyklisch ist, folgt

$$\mathrm{tr}_E\,[\mathbb{1}_S \otimes A_E, B] = \mathrm{tr}_E(\mathbb{1}_S \otimes A_E B) - \mathrm{tr}_E(B\mathbb{1}_S \otimes A_E) = 0$$

für beliebige Operatoren A_E auf $\mathcal{H}_E$ und B auf $\mathcal{H}_S \otimes \mathcal{H}_E$. Somit gilt insgesamt

$$\dot{\rho} = -i\,[H_S, \rho] - i\,\mathrm{tr}_E\,[H_{SE}, |\psi\rangle\,\langle\psi|]\,.$$

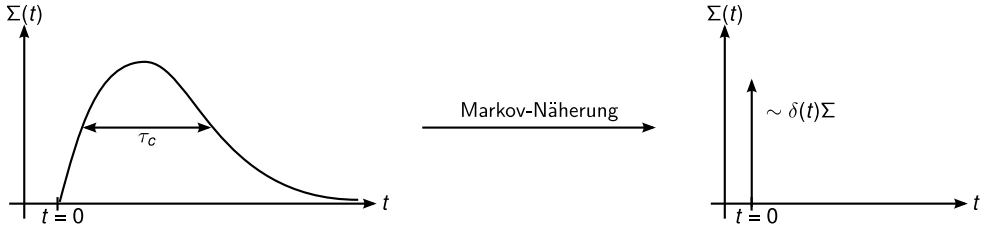
Das Problem an dieser Darstellung ist, dass dies keine geschlossene Differentialgleichung für $\rho(t)$ ist!
Man kann aber zeigen, dass ρ die sogenannte *Nakajima-Zwanzig-Gleichung* erfüllt

$$\dot{\rho}(t) = -i\,[H_S, \rho(t)] + \int_0^t \mathrm{d}t' \Sigma(t-t')\rho(t')$$

mit dem Selbstenergie-Superoperator $\Sigma(t)$. Diese Gleichung ist zwar geschlossen für ρ, ist aber trotzdem problematisch, denn diese Gleichung ist eine komplizierte Integro-Differentialgleichung für $\rho(t)$. Die rechte Seite hängt von der „Vergangenheit“ $\rho(t')$ für $t' < t$ ab.
Als Näherung nimmt man an, dass die Umgebung (das System E) kein Gedächtnis hat bzw. die Erinnerungszeit τ_c sehr viel kürzer als die relevante Zeitskala des Systems S ist. Diese Näherung nennt man *Markov-Näherung*.
In unserem Fall bedeutet dies konkret, dass der Selbstenergie-Superoperator ersetzt wird durch eine Delta-Distribution, d.h.

$$\Sigma(t) \approx \Sigma_{\text{Markov}} = \delta(t) \int_0^\infty \mathrm{d}t' \Sigma(t') =: \delta(t)\Sigma,$$

wobei die Stärke (gegeben durch die Integration) die selbe bleiben soll.



Damit erhalten wir folgende Differentialgleichung für $\rho(t)$

$$\dot{\rho}(t) = -i\,[H_S, \rho(t)] + \Sigma\rho(t) =: L\rho(t),$$

wobei L ein linearer Superoperator ist, der sogenannte *Lindblad-Operator* (engl. *Lindbladian*). Die formale Lösung der Differentialgleichung ist

$$\rho(t) = e^{Lt}\rho(0) =: \Lambda(t)\rho(0).$$

Der Superoperator $\Lambda(t) := e^{Lt}$ entspricht dem Zeitentwicklungsoperator der Schrödingergleichung $U(t) = e^{itH}$. Wir nehmen an, dass Λ ein Superoperator ist, d.h. insbesondere vollständig positiv. Dies ist oft nicht trivial zu zeigen, wenn man von $H = H_S + H_E + H_{SE}$ ausgeht und die Markov-Näherung verwendet.
Unter dieser Annahme können wir das Kraus-Theorem verwenden und erhalten

$$\rho(t) = e^{Lt}\rho(0) = \Lambda(t)\rho(0) = \sum_\mu M_\mu(t)\rho(0)M_\mu^\dagger(t)$$

mit den Kraus-Operatoren M_μ. Die infinitesimale Zeitentwicklung von ρ ist gegeben durch

$$\rho(\mathrm{d}t) = \rho(0) + \dot{\rho}(0)\mathrm{d}t = (\mathbb{1} + L\mathrm{d}t)\rho(0). \tag{2.3}$$

Die infinitesimale Zeitenwicklung der Kraus-Operatoren ist

$$M_0(\mathrm{d}t) = \mathbb{1} + J\mathrm{d}t \qquad\qquad M_\mu(\mathrm{d}t) = L_\mu\sqrt{\mathrm{d}t} \quad (\mu > 0).$$

Setzt man dies ein, so ergibt sich

$$\begin{aligned}\rho(\mathrm{d}t) &= (\mathbb{1} + J\mathrm{d}t)\rho(0)(\mathbb{1} + J^\dagger\mathrm{d}t) + \sum_{\mu>0} L_\mu\rho(0)L_\mu^\dagger\mathrm{d}t \stackrel{!}{=} (\mathbb{1} + L\mathrm{d}t)\rho(0) \\ &= \rho(0) - i\left[H_S, \rho(0)\right]\mathrm{d}t + \Sigma\rho(0)\mathrm{d}t.\end{aligned}$$

Wähle nun $J = -iH_S + K$. Die Normierung der Kraus-Operatoren erfordert

$$\mathbb{1} = \sum_\mu M_\mu^\dagger M_\mu = \mathbb{1} + \mathrm{d}t(-iH_S + iH_S + 2K + \sum_{\mu>0} L_\mu^\dagger L_\mu^\dagger) + \mathcal{O}(\mathrm{d}t^2),$$

sodass

$$K = -\frac{1}{2}\sum_{\mu>0} L_\mu^\dagger L_\mu$$

gelten muss. Einsetzen aller dieser Bedingungen in (2.3) ergibt

$$\dot{\rho} \equiv L\rho = -i\left[H_S, \rho\right] + \sum_{\mu>0}\left(L_\mu\rho L_\mu^\dagger - \frac{1}{2}L_\mu^\dagger L_\mu\rho - \frac{1}{2}\rho L_\mu^\dagger L_\mu\right).$$

Der Kommutator beschreibt die Zeitentwicklung nach der Schrödingergleichung ohne Wechselwirkung mit der Umgebung. Diese Gleichung nennt man (verallgemeinerte bzw. Quanten“)*Mastergleichung* in der *Lindblad-Form*. Die L_μ bezeichnet man ebenfalls als *Lindblad-Operatoren* oder *Quantensprungoperatoren*. Im Folgenden ersetzen wir H_S wieder durch H, da die Wechselwirkung mit der Umgebung in den Lindblad-Operatoren steckt.

Bemerkungen:

(i) Die Mastergleichung gilt nur, wenn die Markov-Näherung gemacht wurde. Sie ist keine exakte Gleichung!

(ii) Die Lindblad-Operatoren L_μ enthalten Raten Γ für Zerfälle ($\sim e^{-\Gamma t} \sim 1 - \Gamma t$). Vergleiche dagegen die unitäre Zeitentwicklung $\cos(\omega t) \sim 1 - \omega^2 t^2$. Die Markov-Näherung benötigt genügend viele Zustände in der Umgebung (Kontinuum, großes System) ($\rightarrow$ vgl. Fermis Goldene Regel).

(iii) Die $U(t)$ bei der Schrödingergleichung bilden eine Gruppe (parametrisch), während die $\Lambda(t) = e^{Lt}$ bei der Mastergleichung nur eine Halbgruppe (ohne Inverses) für $t \geq 0$ bilden.

Beispiel: Dephasierung
Betrachte ein Qubit und die Operatoren

$$H = \frac{\omega}{2}\sigma_z \qquad\qquad L_1 = \sqrt{\frac{\Gamma}{2}}\sigma_z.$$

Dann lautet die Mastergleichung

$$\begin{aligned}\dot{\rho} &= -i\frac{\omega}{2}\left[\sigma_z, \rho\right] + \frac{\Gamma}{2}\left(\sigma_z \rho \sigma_z - \rho\right) \\ &= \begin{pmatrix} 0 & (-i\omega - \Gamma)\rho_{01} \\ (i\omega - \Gamma)\rho_{10} & 0 \end{pmatrix},\end{aligned}$$

wenn $\rho = \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix}$. Damit gilt

$$\rho_{01}(t) = e^{-i\omega-\Gamma}t\rho_{01}(0) \qquad\qquad \rho_{10}(t) = e^{-i\omega-\Gamma}t\rho_{10}(0).$$

Wählt man $\rho_{10}(0) = \rho 01(0)$, so folgt

$$p_x = \langle\sigma_x\rangle = 2\,\mathrm{Re}\,\rho_{01} = 2(\mathrm{Re}\,\rho_{01}(0))\cos(\omega t)e^{-\Gamma t}$$

für die x-Komponente des Bloch-Vektors $\vec{p}$.

Index

Literaturverzeichnis

[ADR82] Alain Aspect, Jean Dalibard, and Gérard Roger. Experimental test of bell's inequalities using time- varying analyzers. *Physical Review Letters*, 49(25):1804–1807, December 1982.

[BB$^+$84] Charles H Bennett, Gilles Brassard, et al. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, volume 175. Bangalore, India, 1984.

[Eke91] Artur K. Ekert. Quantum cryptography based on bell's theorem. *Physical Review Letters*, 67(6):661–663, August 1991.

[HW97] Scott Hill and William K. Wootters. Entanglement of a pair of quantum bits. *Physical Review Letters*, 78(26):5022–5025, June 1997.

[Mer93] N. David Mermin. Hidden variables and the two theorems of john bell. *Reviews of Modern Physics*, 65(3):803–815, July 1993.

[NC10] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, December 2010.

[Pre] John Preskill. Lectures notes for physics 229: Quantum information and computation. http://www.theory.caltech.edu/people/preskill/ph229/.

[Sho94] P.W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *, 35th Annual Symposium on Foundations of Computer Science, 1994 Proceedings*, pages 124–134, 1994.

[Woo98] William K. Wootters. Entanglement of formation of an arbitrary state of two qubits. *Physical Review Letters*, 80(10):2245–2248, March 1998.