

## **Inhalt**

1. Juni 2023

### **Physik des Klimas**

Solarkonstante und Paleoklima

Klimamodelle

### **Quantentheorie**

Das QuBit

Interferometer

BB84 - Quantenkryptographie

Quantenradierer

### **Klassische Systeme**

SI-Einheiten

Kalendersysteme

Die Zeitgleichung

Zeitsysteme

Zeitmessung

Die Gezeiten - Ebbe und Flut

Gezeiten und Tageslänge

Der Nachthimmel

Die Kosmische Entfernungsleiter

Landkarten und der metrische Tensor

### **Spezielle Relativitätstheorie - SRT**

Grundlagen der SRT

Philosophischer Hintergrund der SRT

SRT - Effekte

Das Zwillingsparadoxon

Beschleunigte Systeme und das Rindler-Universum

# Solarkonstante und Paleoklima

Thomas Filk, Universität Freiburg

Unser Klima wird in erster Linie durch die Sonne bestimmt. Von ihr stammt die Energie, die nahezu sämtliche dynamischen Vorgänge auf der Erde antreibt. Eine zweite Energiequelle besteht in Zerfallsprozessen radioaktiver Elemente im Erdinneren. Diese Energieform können wir jedoch für das Verständnis des Klimas vernachlässigen.

Die Energieform, die in der Sonne durch Kernfusionsprozesse entsteht - streng genommen sollte man natürlich immer von „Umwandlung“ sprechen, d.h., bei Kernfusionsprozessen wird Kernenergie in thermische (Bewegungs-)Energie, Strahlungsenergie sowie in Neutrinos umgewandelt -, erreicht uns in Form von elektromagnetischer Strahlung, hauptsächlich im sichtbaren Bereich. Die Oberfläche der Sonne hat eine Temperatur von rund 5800 Kelvin und die zugehörige thermische Strahlung hat ihr Maximum bei rund 500 nm, das entspricht Licht im grün-blauen Bereich.

Der erste Abschnitt wird auf die Solarkonstante eingehen, d.h., die Menge an Energie, die pro Zeiteinheit (Sekunde) und pro Flächeneinheit (Quadratmeter) bei der Erde oberhalb der Atmosphäre ankommt. In diesem Zusammenhang gehen wir auch auf Phänomene wie die Albedo der Erde ein. Außerdem betrachten wir verschiedene Faktoren, die in der Vergangenheit einen Einfluss auf die Solarkonstante bzw. die Einstrahlung der Sonnenstrahlung auf die Erde gehabt haben und damit unser Klima beeinflusst haben könnten. Schließlich betrachten wir auch kurz das Gebiet der Paläoklimatologie, das sich mit dem Klima im Verlauf der Geschichte der Erde beschäftigt.

## 1 Die Solarkonstante

Die Solarkonstante ist definiert als das langjährige Mittel der Intensität pro Flächeneinheit der Sonneneinstrahlung oberhalb der Erdatmosphäre. Die Intensität ist dabei die Energie, die pro Sekunde auf eine bestimmte Fläche - in diesem Fall ein Quadratmeter senkrecht zur Strahlungsrichtung - trifft (siehe Abb. 1). Die IAU (International Astronomical Union) hat 2015 die Solarkonstante aufgrund neuerer Messungen auf den Wert

$$S = 1361 \text{ J} \cdot \text{s}^{-1} \cdot \text{m}^{-2} \quad (1)$$

festgelegt. In älteren Büchern findet man oft den Wert  $S = 1367 \text{ J} \cdot \text{s}^{-1} \cdot \text{m}^{-2}$ , der von 1982 bis 2015 Gültigkeit hatte. Heute misst man die Solarkonstante mit Satelliten, wobei die gemessenen Intensitäten auf den mittleren Abstand Erde-Sonne - die Astronomische Einheit - umgerechnet wird.

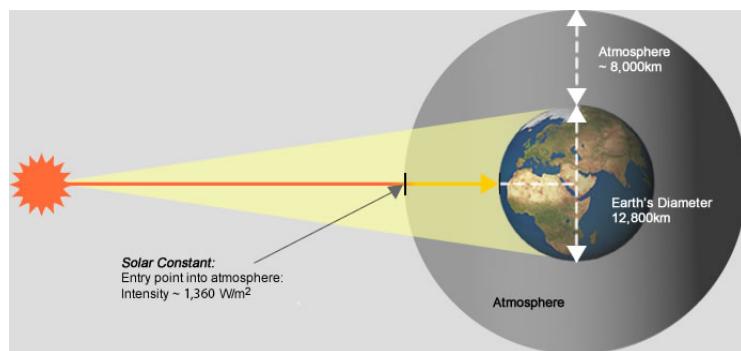


Abbildung 1: Definition der Solarkonstanten. Die gemessene Intensität der Sonnenstrahlung oberhalb der Erdatmosphäre wird auf den mittleren Abstand Sonne-Erde umgerechnet. (aus [1])

## 1.1 Sonnenaktivität

Streng genommen handelt es sich bei der Solarkonstanten nicht um eine Naturkonstante. Sie unterliegt kleinen Schwankungen. Eine winzige Schwankung entsteht durch den 11-Jahres-Zyklus der Sonnenaktivität. Im sichtbaren Bereich machen diese Schwankungen aber nur rund 0,1% aus. Lediglich im UV- bzw. im Röntgen-Bereich können diese Schwankungen wesentlich größer sein, allerdings wird diese Strahlung in höheren Schichten unserer Atmosphäre reflektiert bzw. absorbiert und erreicht den Erdboden größtenteils nicht. Allerdings lassen sich Schwankungen in der mittleren Jahrestemperatur von der Größenordnung von  $0,1^{\circ}\text{C}$  mit einer Periode von 11 Jahren über längere Zeiträume nachweisen. Außerdem gab es in der Vergangenheit häufiger Perioden, in denen die Sonne insgesamt weniger aktiv war und die möglicherweise zu kleinen Eiszeiten geführt haben. Bekannt sind solche Perioden in der Zeit zwischen dem 14. und 18. Jahrhundert (das sogenannte Spörer-Minimum und das Maunder-Minimum; wobei ein direkter Bezug zum Klima in dieser Periode immer noch umstritten ist), beispielsweise durch Wintergemälde von Pieter Bruegel dem Älteren und seinen Söhnen.

Seit Beginn des 17. Jahrhunderts (seit der Erfindung des Teleskops) wurden die Sonnenflecken direkt beobachtet, sodass es gute Aufzeichnungen gibt. Für die Perioden davor eignen sich manche Isotopmessungen (z.B.  $^{14}\text{C}$  und  $^{10}\text{Be}$ ). Diese Isotope entstehen hauptsächlich in der Atmosphäre durch den Einfluss der kosmischen Strahlung, die wiederum durch eine starke Sonnenaktivität und die damit verbundenen Sonnenwinde abgeschwächt wird. Dies führt zu einer Korrelation zwischen der Häufigkeit dieser Isotope in Bohrproben, die auf bestimmte Zeiten datiert werden können, und der Sonnenaktivität: Höhere Isotopenanteile lassen auf geringere Sonnenaktivität schließen, da zu diesen Zeiten die kosmische Strahlung ungehinderter in die Atmosphäre dringen konnte.

## 1.2 Milanković-Zyklen

Für unser Klima (vermutlich) relevante Schwankungen sind die sogenannten Milanković-Zyklen, benannt nach dem serbischen Mathematiker Milutin Milanković (1879–1958). Hierbei handelt es sich um regelmäßige Oszillationen in den Parametern der Erdumlaufbahn um die Sonne. Bei diesen Parametern handelt es sich insbesondere um die Exzentrizität der Erdbahn, die Neigung der Erdachse und die Präzession der Erdachse.

### 1.2.1 Die Exzentrizität der Erdumlaufbahn

In einem reinen Zwei-Körper-Problem mit einer  $1/r^2$ -Kraft (manchmal als nicht-relativistisches Kepler-Problem bezeichnet) bewegt sich ein leichter Körper (Erde) um einen schweren Körper (Sonne) auf einer elliptischen Bahn, wobei sich der schwere Körper in einem der Brennpunkte der Ellipse befindet. Dies gilt ganz allgemein für die Relativkoordinate zwischen den beiden Himmelskörpern, auch wenn die Masse des leichteren Himmelskörpers im Vergleich zu dem schwereren Himmelskörper nicht vernachlässigt werden kann.

Durch den Einfluss der anderen Planeten, insbesondere Jupiter und Saturn, verändert sich die Bahnkurve der Erde jedoch im Verlauf der Zeit. Insbesondere kann auch die Exzentrizität der elliptischen Bahn zwischen einer fast kreisförmigen Erdumlaufbahn ( $\epsilon = 0,0006$ ) und einer schwach elliptischen Bahn ( $\epsilon = 0,058$ ) variieren [5]. Diese Werte schwanken periodisch mit einer Periode von rund 405 000 Jahren, wobei auch Unterzyklen von der Größenordnung von 100 000 Jahren existieren.

Derzeit beträgt der Wert rund  $\epsilon = 0,0167$ , was einer Schwankung in der Entfernung zwischen Erde und Sonne im Bereich zwischen 147,09 Millionen Kilometern und 152,10 Millionen Kilometern entspricht. Obwohl die Differenz in diesen Werten nur rund 3.4% ausmacht, bedeutet dies für die Intensität der Sonnenstrahlung eine Schwankung von rund 6,8% im Verlauf eines Jahres (1). Bei

einer entsprechend größeren Exzentrizität sind auch diese Schwankungen größer und können bis zu 24% ausmachen.

### 1.2.2 Neigung der Erdachse

Im Vergleich zur Ekliptik, also der Ebene der Erdumlaufbahn um die Sonne, ist die Drehachse der Erde um rund  $23,5^\circ$  geneigt. Dieser Neigungswinkel ändert sich aufgrund der Einflüsse anderer Planeten mit einer Periode von rund 41.000 Jahren und schwankt zwischen  $22,1^\circ$  und  $24,5^\circ$ .

Auch diese Schwankung hat zunächst einen jahreszeitlichen Einfluss auf unser Klima: Ist der Neigungswinkel größer, ist der Unterschied im Einfallswinkel der Sonne zwischen Sommer und Winter entsprechend größer, d.h., die jahreszeitlichen Schwankungen fallen stärker aus. Das wiederum kann einen Einfluss darauf haben, wie stark Schnee- und Eisflächen im Sommer abtauen und sich somit zurückbilden. Außerdem haben diese Schwankungen einen Einfluss auf verdunstende Wassermengen in höheren Breitengraden und somit auf den dortigen Niederschlag, was sich beispielsweise im Winter auf erhöhten Schneezuwachs bei Gletschern auswirken kann.

### 1.2.3 Präzession der Erdachse

Da die Erde keine ideale Kugelform hat sondern entlang der Erdachse etwas abgeplattet ist, also entlang des Äquators etwas „dicker“ als entlang von Längengraden (der Abstand vom Erdzentrum zum Nord- bzw. Südpol ist um rund 21 Kilometer kleiner als der Abstand vom Erdzentrum zum Äquator, wobei hier für die Erde vereinfachend die Form eines Rotationsellipsoids angenommen wird). Der gravitative Einfluss von Sonne und Mond (in geringerem Maß auch der von anderen Planeten, insbesondere Jupiter und Saturn) bewirkt ein Drehmoment, das die Erdachse aufrichten würde, falls sich die Erde nicht drehte. Wegen der Drehimpulserhaltung wird die Drehachse zur Seite gedreht und rotiert langsam um eine Senkrechte zur Erdbahn ((Bild!)). Diese Drehung bezeichnet man als Präzession. Sie hat eine Periode von rund 25 800 Jahren.

Der Haupteffekt der Neigung der Erdachse sind die Jahreszeiten, die auf der Nord- und Südhalbkugel der Erde um ein halbes Jahr relativ zueinander verschoben sind. Die Präzession bewirkt, zusammen mit den anderen Orbitalparametern, dass die Unterschiede zwischen den Jahreszeiten (insbesondere zwischen Sommer und Winter) hinsichtlich ihrer Intensität verschieden stark ausfallen können. Wenn beispielsweise die Elliptizität der Erbahn (d.h. die Exzentrizität) sehr groß ist, kann die Richtung der Erdachse relativ zu den Hauptachsen die Strahlungsunterschiede zwischen Sommer und Winter entweder verstärken (wenn der Sommer mit dem Perihel zusammenfällt) oder abschwächen (wenn Sommer mit dem Aphel zusammenfällt).

Ein weiterer wesentlicher Faktor für das Klima ist, dass die Nordhalbkugel der Erde größere Landmassen hat als die Südhalbkugel, die eine größere Wasserfläche hat. Insofern spielt es eine Rolle, ob die oben erwähnte Verstärkung der Unterschiede zwischen Sommer und Winter für die Nord- oder für die Südhalbkugel zutrifft.

Während man in der physikalischen und astrophysikalischen Literatur für die Präzession der Erde einen Wert von 25 800 (oder aufgerundet 26 000) Jahren findet, findet man in der Literatur zur Klimaphysik bzw. zu den Milanković-Zyklen oftmals einen Wert von 23 000 Jahren. Für die Physik (z.B. die Bestimmung des Frühlingspunkts und den damit zusammenhängenden Jahreszeiten) ist die Richtung der Erdachse relativ zur Sonne wichtig. Für die Klimaforschung ist man eher an der Richtung der Erdachse relativ zum Perihel bzw. Aphel interessiert. Wegen der Periheldrehung der Erde, verschieben sich diese Punkte aber langsam. Der kombinierte Effekt von Präzession und Periheldrehung führt zu der verkürzten Periode von 23 000 Jahren.

### 1.3 Das Sonnenalter

Auf sehr langen Zeitskalen nimmt die Solarkonstante zu: In 100 Millionen Jahren um rund 1% [4]. Zu Beginn der Erdgeschichte betrug die Sonnenintensität nur rund 70% ihres heutigen Werts. Das Klima auf der Erde hätte somit wesentlich kälter sein müssen und alles Wasser auf der Erde hätte gefroren sein müssen. Es gibt aber deutliche Hinweise darauf, dass es insgesamt meist wärmer auf der Erde gewesen ist. Dies bezeichnet man als das *Faint young Sun paradox* [4]. Eine mögliche Lösung ist, dass der Kohlendioxideinhalt der Atmosphäre in früheren Zeiten (z.B. aufgrund von Vulkanismus) wesentlich höher war als heute. Diese Frage ist aber noch nicht endgültig geklärt.

## 2 Die Albedo

Die Albedo ist ein Maß dafür, wie stark ein Gegenstand eine Strahlung reflektiert. Es ist so etwas wie der totale elastische Wirkungsquerschnitt eines Gegenstands für elektromagnetische Strahlung. Allerdings handelt es sich nicht um eine Fläche, sondern um ein Verhältnis: das Verhältnis von reflektierter Intensität zu eingestrahlter Intensität. Man kann die Albedo als Funktion der Wellenlänge bzw. der Frequenz betrachten (da es um die reflektierte Strahlung geht, soll die Wellenlänge bzw. Frequenz erhalten bleiben), meist interessiert man sich aber für die Summe über das gesamte Spektrum.

Wenn man ein Foto von der Erde betrachtet, aufgenommen von einem Satelliten oder, besser noch, von einer Raumsonde oder Rakete auf dem Weg zum Mond oder einem anderen Ort im Sonnensystem, ist alles, was man von der Erde sieht, reflektierte Strahlung (siehe Abb. 2). Auf einem solchen Bild sieht man sofort, welche Teile der Erde eine hohe und welche eine niedrige Albedo haben: Schnee- und Wolkenfelder haben eine hohe Albedo, ebenso Eisfelder; Wasser und Wälder haben eine sehr niedrige Albedo. Sand bzw. Wüste oder Steppen haben eine mittlere Albedo.



Abbildung 2: Die Erde, aufgenommen von der Crew der Apollo-17 Mission im Jahre 1972. Deutlich erkennbar sind die Antarktis, der afrikanische Kontinent, die Insel Madagaskar und die saudi-arabische Halbinsel. Die sehr stark reflektierenden Gebiete sind weiß, das sind Schnee- und Wolkenflächen. Die Wüsten sind deutlich heller als Waldgebiete oder Grasflächen. Sehr dunkel sind die Meere. Diese Helligkeiten entsprechen der Albedo der jeweiligen Flächen. (aus [2])

Die Albedo hat einen sehr großen Einfluss auf unser Klima. Je größer die Albedo eines Planeten ist, umso geringer ist (bei gleichbleibenden anderen Faktoren) die Oberflächenerwärmung. Während die Erde insgesamt eine Albedo von 0,3 hat, hat beispielsweise der Planet Venus aufgrund seiner dichten Wolkenschicht eine Albedo von 0,7. Obwohl Venus deutlich näher an der Sonne ist als die Erde und aus diesem Grunde eine doppelt so hohe Solarkonstante hat, wäre ihre Temperatur aufgrund der Albedo kühler als die der Erde. Tatsächlich ist ihre Oberflächentemperatur jedoch wesentlich höher (bei 460°C). Der Grund ist der Treibhauseffekt: Die Atmosphäre von Venus besteht zu 96% aus Kohlendioxid.

### 3 Aufbau der Atmosphäre

Die Atmosphäre der Erde wird in verschiedene Schichten unterteilt, von denen die untersten drei Schichten - die Troposphäre, die Stratosphäre und die Mesosphäre - den größten Einfluss auf unser Klima haben. Sie sind durch ihre Temperaturgradienten definiert. Die beiden darüber liegenden Schichten - die Thermosphäre (100–600 km) und die Exosphäre (600–200 000 km) - haben keinen direkten Einfluss auf unser Wetter bzw. Klima.

#### 3.1 Die Troposphäre

Die Troposphäre ist die unterste Atmosphärenschicht, in der sich nahezu alle Wettervorgänge abspielen. Definiert ist sie durch einen negativen Temperaturgradienten, d.h., in dieser Schicht nimmt die Temperatur mit der Höhe ab. Sie erstreckt sich an den Polen bis in eine Höhe von rund 6–8 km, in den Tropen bis zu einer Höhe von 12–18 km. Im Durchschnitt hat sie eine Höhe von 13 km.

Die Abnahme der Temperatur hängt mit der Druckabnahme zusammen. Der Druck nimmt nahezu exponentiell mit der Höhe ab (dies gilt auch weit über die Troposphäre hinaus). Aus diesem Grund dehnt sich aufsteigende Luft aus (sie passt sich praktisch instantan dem Umgebungsdruck an) und wird dabei kühler. Dieser Vorgang erfolgt nahezu adiabatisch, d.h., es findet kein Wärmeaustausch mit der Umgebung statt. Aus diesem Grund nimmt die Temperatur mit der Höhe ab.

Eine instabile Wetterlage liegt vor, wenn die Abkühlung eines Luftpakets bei seinem Aufstieg aufgrund des verminderten Drucks langsamer erfolgt, als es der Temperatur der Umgebung entspricht. In diesem Fall hat das Luftpaket in einer bestimmten Höhe eine höhere Temperatur als die Umgebung, aber es hat denselben Druck. Höhere Temperatur aber gleicher Druck bedeutet, dass die Dichte des Luftpakets geringer ist als die Dichte der Umgebungsluft und somit ist das Luftpaket leichter und steigt weiter in die Höhe. Es findet somit eine Konvektion statt. Nimmt die Temperatur eines Luftpakets jedoch beim Aufstieg schneller ab, als die Temperatur der Umgebung, bleibt das Luftpaket dichter und steigt nicht weiter bzw. sinkt wieder. In diesem Fall ist die Lage stabil. Insbesondere herrscht eine stabile Wetterlage bei einer Inversionslage, d.h., wenn die Temperatur lokal mit der Höhe zunimmt. In aufsteigende Luft nimmt der Druck immer noch ab, sie kühlte sich somit ab und ihre Temperatur bleibt unter der Temperatur der Umgebung. Somit ist dieses Luftpaket dichter als die Luft der Umgebung und sinkt wieder ab.

Gehen wir an der Erdoberfläche von einer mittleren Temperatur von rund 18°C aus, so kann die Temperatur bis zur Obergrenze der Troposphäre auf rund –50°C bis –60°C abnehmen.

#### 3.2 Die Stratosphäre

Oberhalb der Troposphäre beginnt die Stratosphäre, wobei diese beiden Atmosphärenschichten durch die sogenannte Tropopause getrennt sind. In der Stratosphäre nimmt die Temperatur mit zunehmender Höhe zu und kann in rund 50 km Höhe wieder nahezu bei 0°C liegen. In der Stratosphäre liegt die Ozonschicht. Das Ozon absorbiert die UV-Strahlung, was zu einer Erwärmung führt. Da hier die höher liegenden Luftsichten eine höhere Temperatur haben, kommt es in der Stratosphäre praktisch nicht mehr zur Konvektion. Wolken, z.B. Gewitterwolken, die bis in die Stratosphäre reichen, bilden dort meist einen sogenannten Amboss, d.h. eine flache ausgedehnte Struktur, in der keine Konvektion mehr stattfindet.

#### 3.3 Die Mesosphäre

Oberhalb der Stratosphäre in rund 50–60 km Höhe beginnt die Mesosphäre. Sie reicht bis ungefähr 80–90 km. In dieser Schicht findet man kaum noch Ozon, sodass die Temperatur in der Mesosphäre wieder

abnimmt, teilweise bis deutlich unter  $-140^{\circ}\text{C}$ . Dies ist die kälteste Schicht unserer Atmosphäre.

### 3.4 Thermosphäre und Exosphäre

In rund 85 km Höhe beginnt die Thermosphäre. Hier ist die Luft so dünn, dass die Atome von einzelnen Photonen auf sehr hohe Geschwindigkeiten beschleunigt werden können, und da die mittlere Weglänge der Atome bzw. Moleküle sehr groß ist, kommt es kaum zu einem Austausch. Die Temperatur nimmt daher wieder zu, bis teilweise auf über  $1000^{\circ}\text{C}$ .

Oberhalb der Thermosphäre in rund 600–700 km Höhe beginnt die sogenannte Exosphäre. Die Temperatur ändert sich hier nicht - die Luft wird so dünn, dass man von Temperatur im thermodynamischen Sinne kaum sprechen kann. Der Übergang zwischen beiden Schichten ist fließend. Eine Definition definiert die Grenze zwischen diesen beiden Schichten über die mittlere freie Weglänge der Atome bzw. Teilchen. Als Obergrenze der Exosphäre wird meist die Schicht definiert, in der die Sonnenwinde einen größeren Einfluss auf die Teilchen haben als das Gravitationsfeld der Erde. Diese Schicht liegt bei rund 200 000 km.

### 3.5 Homosphäre und Heterosphäre

Bis zu ungefähr der gleichen Höhe wie die Mesosphäre, d.h. bis rund 85 km, reicht auch die sogenannte Homosphäre. Das ist der Bereich der Atmosphäre, der als „well mixed“ (gut durchmischt) gilt. In diesem Bereich ändert sich die Zusammensetzung der Luft kaum, d.h., in diesem Bereich besteht die Atmosphäre zu rund 78% aus Stickstoff, 21% Sauerstoff, 0,94% Argon sowie Kohlendioxid, Neon, Helium, Methan, Stickoxide und weitere Spurengase. Bis zu dieser Höhe findet ausreichend vertikale Durchmischung der Luft statt, sodass diese Verhältnisse bestehen bleiben. Oberhalb von rund 85 km beginnt die Heterosphäre. Hier ist die Luft so dünn, dass es zu einer Trennung der verschiedenen Gasanteile entsprechend ihrer molekularen Gewichte kommt. Sauerstoff, Stickstoff und Argon bleiben zunächst weg, später in größerer Höhe auch Helium, sodass es in den äußersten Schichten praktisch nur noch Wasserstoff gibt. Wasserstoff und zu einem geringeren Anteil Helium sind auch die einzigen Gase, die dem gravitativen Einfluss der Erde entkommen und in den Weltraum entweichen können.

## 4 Anmerkungen

(1) Der Grund für den Faktor 2 zwischen der Schwankung im Abstand und der Schwankung in der Intensität der Sonneneinstrahlung liegt in dem  $1/r^2$ -Gesetz der Intensität als Funktion des Abstands:

$$\frac{1}{(r \pm \Delta r)^2} \approx \frac{1}{r^2} \mp 2 \frac{\Delta r}{r}. \quad (2)$$

## Literatur

- [1] Solarkonstante; Uni Kassel,  
<https://www.greenrhinoenergy.com/solar/radiation/images/solar-constant.jpg>
- [2] NASA Image and Video Library, <https://images.nasa.gov/details/as17-148-22727>
- [3] NASA Earth Fact Sheet; <https://nssdc.gsfc.nasa.gov/planetary/factsheet/earthfact.html>
- [4] Wikipedia „Faint young Sun paradox“ [https://en.wikipedia.org/wiki/Faint\\_young\\_Sun\\_paradox](https://en.wikipedia.org/wiki/Faint_young_Sun_paradox).

[5] Wikipedia „Milankovitch cycles“. [https://en.wikipedia.org/wiki/Milankovitch\\_cycles](https://en.wikipedia.org/wiki/Milankovitch_cycles).

[6] Wikipedia „Solarkonstante“. <https://de.wikipedia.org/wiki/Solarkonstante>.

# Klimamodelle

Thomas Filk, Universität Freiburg

Für das Verständnis unseres Klimas sowie die Möglichkeiten, Vorhersagen zu treffen, wie sich unser Klima in den nächsten Jahrzehnten entwickeln wird und wie wir die Auswirkungen eines Klimawandels, der mittlerweile kaum noch bestritten wird, möglichst gering halten können, spielen Klimamodelle eine wichtige Rolle. Es sind auch gerade die Aussagen solcher Klimamodelle, die von Leugnern eines anthropogenen (d.h. vom Menschen verursachten) Klimawandels immer wieder in Frage gestellt werden. Daher ist es wichtig, einen ungefähren Eindruck von Klimamodellen zu bekommen.

In diesem Kapitel betrachten wir zunächst sehr einfache Klimamodelle, die sich auf den globalen Energiehaushalt der Erde beziehen. Insgesamt ist dieser Energiehaushalt nahezu ausgeglichen. Außerdem handelt es sich um statische Modelle, d.h., wir untersuchen die Gleichgewichtszustände solcher Modelle und wie sich diese verändern, wenn sich äußere Parameter verändern. Gegen Ende dieses Kapitels beschreiben wir auch kurz die wichtigsten Modelle, die heute für Klimaproxgnosen verwendet werden. Die Modelle lassen sich kaum noch in ihrer Komplexität überschauen, aber man kann zumindest grob beschreiben, welche Effekte von diesen Modellen berücksichtigt werden.

## 1 Grundlagen

Für die einfachen Klimamodelle, die hier besprochen werden, sind die folgenden Grundlagen relevant. Ausführlicher wird auf diese Themen in Kapitel [Solarkonstante und Paleoklima](#) eingegangen.

### 1.1 Die Solarkonstante

Die Solarkonstante ist definiert als das langjährige Mittel der Intensität der Sonneneinstrahlung oberhalb der Erdatmosphäre. Die Intensität ist dabei die Energie, die pro Sekunde auf eine bestimmte Fläche - in diesem Fall ein Quadratmeter senkrecht zur Strahlungsrichtung - trifft (siehe Kapitel [Solarkonstante und Paleoklima](#)). Die IAU (International Astronomical Union) hat 2015 die Solarkonstante aufgrund neuerer Messungen auf den Wert

$$S = 1361 \text{ J} \cdot \text{s}^{-1} \cdot \text{m}^{-2} = 1361 \text{ W} \cdot \text{m}^{-2} \quad (1)$$

festgelegt.

Auf Zeitskalen von einigen hundert Jahren liegt die dominante Schwankung der Solarkonstante in den 11-jährigen Aktivitätszyklen der Sonne. Diese wirken sich aber im sichtbaren Bereich kaum aus. Die beobachteten Schwankungen in der Temperatur im langjährigen Mittel aufgrund dieser Sonnenzyklen liegt in der Größenordnung von  $0,1^\circ\text{C}$ . Zum Einfluss längerfristiger Schwankungen (z.B. der sogenannten Milanković-Zyklen), siehe Kapitel [Solarkonstante und Paleoklima](#).

### 1.2 Die Albedo

Einen wesentlich größeren Einfluss auf das Klima hat die Albedo der Erde (siehe Kapitel [Solarkonstante und Paleoklima](#)). Hierbei handelt es sich um die Reflektivität der Erde für den Wellenlängenbereich der Sonnenstrahlung. Je größer die Albedo, umso größer ist die Menge an Energie, die direkt wieder in den Weltraum abgegeben wird und somit auf den Energiehaushalt der Erde keinen Einfluss hat. Derzeit liegt die Albedo der Erde bei  $a = 0,3$ , d.h. rund 30% der eingestrahlten Energie wird entweder an den Wolken oder der Erdoberfläche zurück in den Weltraum reflektiert. Den wichtigsten

Beitrag zur Albedo der Erde liefern neben den Wolken die Schnee- und Eisflächen der nördlichen und südlichen Breitengrade. Auch landschaftliche Veränderungen, beispielsweise das Abholzen von Nadelwäldern (die eine sehr kleine Albedo haben) und die Ersetzung durch Wiesen oder Städte, haben einen Einfluss. Der weitaus größte Unsicherheitsfaktor bei Klimamodellen sind die Wolken und die damit zusammenhängende Albedo: Es ist bisher kaum verstanden, wie die Wolkendichte auf den Klimawandel reagieren wird.

### 1.3 Aufbau der Erdatmosphäre

Ebenfalls wichtig für Klimamodelle ist der Aufbau der unteren Atmosphäre. Bis zu einer Höhe von rund 12–13 km (in Polnähe und im Winter ist diese Schicht etwas tiefer, in den Tropen kann sie bis zu 18–20 km betragen) reicht die Troposphäre. In dieser Schicht spielt sich unser Wettergeschehen ab. Definiert ist diese Schicht durch einen negativen Temperaturgradienten, d.h., die Temperatur nimmt in dieser Schicht mit zunehmender Höhe ab. Den Haupteinfluss auf dieses Verhalten hat die Tatsache, dass sich aufsteigende Luftmassen, deren Druck daher abnimmt, abkühlen. Nimmt die Temperatur eines solchen Luftpaketes mit der Höhe nicht rasch genug ab, bleibt es wärmer als die Umgebungstemperatur. Dadurch steigt es weiter auf, da ein wärmeres Luftpaket (bei gleichem Druck) weniger dicht und damit leichter ist. Auf diese Weise kommt es zur Konvektion: Warme Luftschichten steigen auf, kühlen sich dabei ab, je nach Luftfeuchtigkeit sinkt die Temperatur unter den sogenannten Taupunkt - hier kommt es zur Kondensation von Wasser und Wolkenbildung. Bei der Kondensation wird Energie frei, die den Auftrieb der Luftmassen nochmals beschleunigt.

Oberhalb der Troposphäre liegt die Stratosphäre, die bis in eine Höhe von rund 50 km reicht. Hier nimmt die Temperatur mit der Höhe wieder zu, von rund  $(-50)^\circ$  –  $(-80)^\circ\text{C}$  (abhängig von der Jahreszeit und dem Breitengrad) bis wieder zum Gefrierpunkt von  $0^\circ\text{C}$ . Der Grund dafür ist, dass sich in dieser Schicht die Ozonschicht der Erde befindet. Diese absorbiert ultraviolette Sonnenstrahlung und erwärmt sich dadurch. Die weiteren Schichten (Mesosphäre, Thermosphäre und Exosphäre) spielen für das Klima eine untergeordnete Rolle.

Die Luft der Atmosphäre besteht im Wesentlichen aus Stickstoff ( $\text{N}_2$ , rund 78%) und Sauerstoff ( $\text{O}_2$ , rund 21%) sowie dem Edelgas Argon ( $\text{Ar}$ , rund 0,9%). Der Rest sind Spurengase, deren Anteile meist in ppm (parts per million), ppb (parts per billion) oder ppt (parts per trillion) angegeben werden. Hierbei handelt es sich um Volumenanteile, also im Wesentlichen Anteile pro Teilchenzahl, da wir bei Gasen (insbesondere Gasen, die sich wie ideale Gase verhalten) das Eigenvolumen der Moleküle vernachlässigen können.  $\text{CO}_2$  (Kohlendioxid) ist mit derzeit rund 420 ppm oder 0,042% das häufigste Gas, es folgen  $\text{Ne}$  (Neon, 18 ppm),  $\text{He}$  (Helium, 5 ppm),  $\text{CH}_4$  (Methan, 1870 ppb). Außerdem enthält die Luft je nach Luftfeuchtigkeit bis zu 4 Volumenprozent  $\text{H}_2\text{O}$  (Wasserdampf).

## 2 Einige globale, statische Klimamodelle

Ein Klimamodell ist ein System mathematischer Gleichungen, in das klimarelevante Parameter (z.B. Luft- und Wassertemperatur in bestimmten Luftschichten bzw. Wassertiefen, Niederschläge, Luftfeuchtigkeit, Eisfläche und Dicke von Eisschichten, ... eventuell ortsabhängig) eingehen und das den zeitlichen Verlauf solcher Parameter beschreibt. Die einfachsten Klimamodelle betrachten lediglich die Durchschnittstemperatur der Erde (also keine Ortsabhängigkeiten). Weitere Vereinfachungen bestehen darin, die Zeitabhängigkeit zu vernachlässigen, d.h., es werden Gleichgewichtszustände untersucht. Selbst auf diesem Niveau unterscheiden sich Klimamodelle noch hinsichtlich der Eigenschaften der Erdoberfläche, die berücksichtigt werden.

Die folgenden Klimamodelle sind von dieser globalen Struktur und berücksichtigen zuneh-

mend mehr Einflüsse. Man erkennt, wie schnell die Modell komplexer werden, wie andererseits aber auch einfache Modelle schon bestimmte nicht-triviale Effekte beschreiben können. Abbildung 1 gibt global die Strahlungsbilanz der Atmosphäre wieder, die näherungsweise in den angeführten Modellen beschrieben wird.

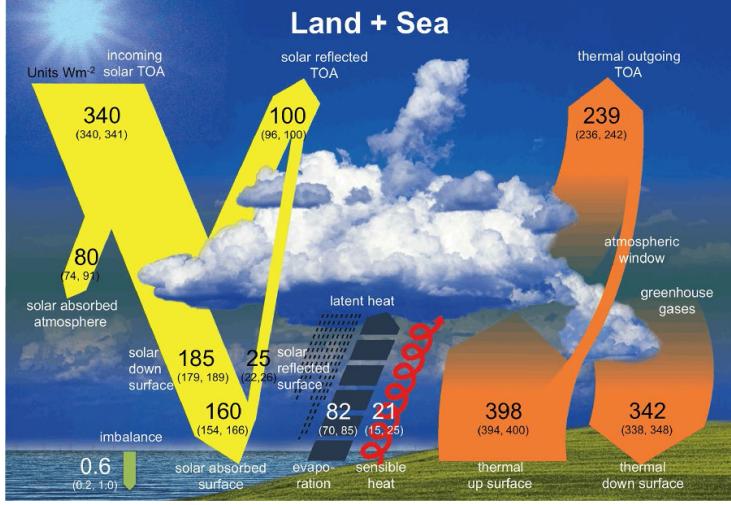


Abbildung 1: Einfaches Diagramm zur Veranschaulichung des radiativen Energiehaushalts der Atmosphäre. TOA = top of atmosphere. Die Zahlen bezeichnen eine Strahlungsintensität (in Watt pro Quadratmeter). In Klammern ist jeweils das Intervall der Unsicherheit angegeben. (Aus [1])

Im Folgenden verwenden wir auch das Kirchhoff'sche Strahlungsgesetz, wonach der Absorptionskoeffizient und der Emissionskoeffizient eines Körpers gleich sind. Wir setzen hier die integrierte diffuse Strahlung voraus, d.h. wir betrachten keine Richtungsabhängigkeiten. Allerdings berücksichtigen einige der Klimamodelle eine Frequenzabhängigkeit dieser Koeffizienten. Die Interpretation sogenannter Emissionskoeffizienten in diesem Kapitel bedeutet meist, dass die Atmosphäre für bestimmte Frequenzen absorbierend und für andere Frequenzen durchlässig ist. Daher ist die Forderung, dass die frequenzabhängige Absorption gleich der entsprechenden Emission ist, eher eine Näherung.

## 2.1 Klimamodell 1 - ohne Atmosphäre

Wir betrachten zunächst ein sehr einfaches Klimamodell, bei dem folgende Faktoren berücksichtigt werden:

1. eine konstante Intensität der Sonnenstrahlung oberhalb der Atmosphäre, dies ist die Solarkonstante  $S = 1\,361\text{ J} \cdot \text{s}^{-1} \cdot \text{m}^{-2}$ ,
2. eine konstante Albedo von  $a = 0,3$ ,
3. eine Erdoberfläche, welche die absorbierte Energie in alle Richtungen als Schwarzkörperstrahlung entsprechend ihrer Oberflächentemperatur  $T$  abgibt. Hier wird das Stefan-Boltzmann-Gesetz

$$L = \sigma T^4 \quad (2)$$

angenommen, d.h., die abgestrahlte Intensität  $L$  (Energie pro Zeit und Fläche) ist proportional zur 4. Potenz der absoluten Temperatur  $T$  der Erdoberfläche. Die Proportionalitätskonstante ist die Stefan-Boltzmann-Konstante  $\sigma = 5,67 \cdot 10^{-8}\text{ J} \cdot \text{s}^{-1} \cdot \text{m}^{-2} \cdot \text{K}^{-4}$ . (2)

Bis auf die Oberflächentemperatur der Erde sind alle Parameter bekannt. Für die gesamte Intensität der absorbierten Strahlung gilt:

$$L_{ab} = (1 - a)S\pi R^2. \quad (3)$$

Der Faktor  $(1 - a)$  beschreibt die Reduktion der eingestrahlten Intensität  $S$  aufgrund der Albedo.  $\pi R^2$  (wobei  $R$  der Radius der Erde ist) beschreibt die Querschnittsfläche der Erde, die von der Sonnenstrahlung getroffen wird. Meist argumentiert man, dass dies die Fläche des Schattens hinter der Erde ist, wo die Sonnenstrahlung fehlt, weil sie auf die Erde getroffen ist.

Diese absorbierte Intensität muss im Gleichgewicht gleich der emittierten Intensität sein (siehe Abb. 2). Bei der emittierten Intensität soll es sich in diesem Modell um die Intensität der Schwarzkörperstrahlung der Erde bei ihrer Oberflächentemperatur  $T_s$  handeln (siehe Abb. 2, (a)):

$$L_{\text{em}} = \sigma T_s^4 4\pi R^2. \quad (4)$$

Hier wird als emittierende Fläche die gesamte Oberfläche der Erde angenommen (daher der Faktor 4). Setzt man  $L_{\text{ab}} = L_{\text{em}}$  folgt

$$(1 - a)S = 4\sigma T_s^4 \quad \Rightarrow \quad T_s = \sqrt[4]{\frac{(1 - a)S}{4\sigma}} = 254,6 \text{ K}. \quad (5)$$

Nach diesem einfachen Modell müsste die durchschnittliche Erdtemperatur knapp 255 K oder  $-18^\circ\text{C}$  betragen. Das ist natürlich viel zu kalt. Bei diesen Temperaturen wären die Ozeane gefroren. Die tatsächliche Jahresdurchschnittstemperatur der Erde liegt (2017) bei  $15^\circ\text{C}$  oder 288 K.

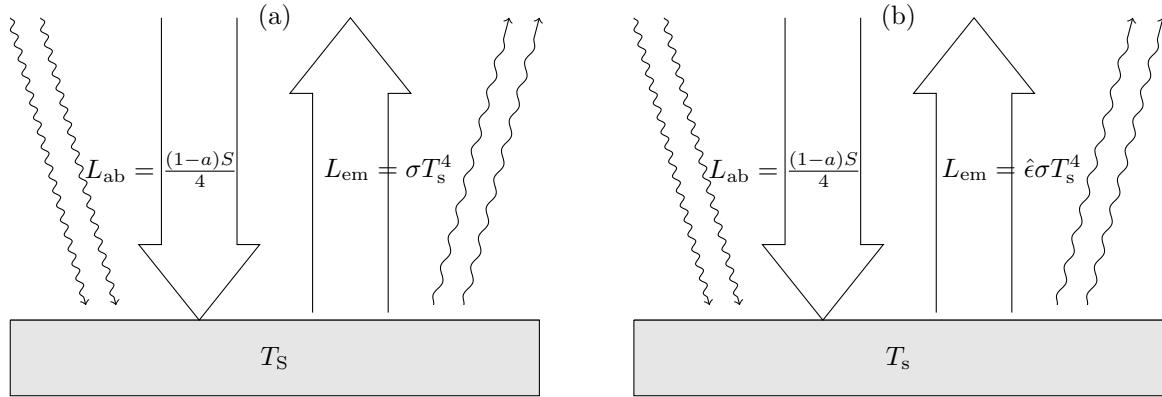


Abbildung 2: Zwei einfache Klimamodelle. (a) In diesem Modell wird sämtliche von der Sonne einfallende Strahlung (abzüglich der Albedo) von der Erdoberfläche absorbiert und von dieser als Schwarzkörperstrahlung wieder in den Weltraum abgegeben. (b) In diesem Modell gibt die Erdoberfläche ebenfalls alle absorbierte Strahlung wieder ab, allerdings nicht als Schwarzkörperstrahlung sondern um einen Faktor  $\hat{\epsilon}$  geringer. Dieser Faktor beschreibt den Einfluss des Treibhauseffekts, gibt aber keine Erklärung.

## 2.2 Klimamodell 2 - ohne Atmosphärensicht, mit Emissionsfaktor

Es gibt mehrere Gründe, weshalb das erste Klimamodell eine deutlich zu tiefe Oberflächentemperatur der Erde liefern könnte. Zum Einen könnte es eine Energiequelle geben, die bisher nicht berücksichtigt wurde. Dafür kommt praktisch nur die Erde selbst in Frage, insbesondere die Kernzerfälle im Inneren der Erde. Hier zeigt eine genauere Überlegung jedoch, dass dieser Effekt zu klein ist und für die Klimabetrachtungen nicht berücksichtigt zu werden braucht.

Weiterhin könnte die Annahme falsch sein, dass die Erde eine Schwarzkörperstrahlung zu einer Temperatur  $T$  abgibt, die ihrer Oberflächentemperatur entspricht. Wir führen für unser zweites Klimamodell einen Emissionsfaktor  $\hat{\epsilon}$  ein, setzen also für die emittierte Intensität

$$L_{\text{em}} = \hat{\epsilon} \sigma T^4 4\pi R^2 \quad (6)$$

an.  $\hat{\epsilon} = 1$  entspricht dem alten Modell. Man spricht in diesem Fall auch schon mal von einem „grauen Strahler“ (siehe Abb. 2(b)). Wir erhalten nun für die Oberflächentemperatur der Erde:

$$(1 - a)S = 4\hat{\epsilon}\sigma T^4 \quad \Rightarrow \quad T = \sqrt[4]{\frac{(1 - a)S}{4\hat{\epsilon}\sigma}} = \sqrt[4]{\frac{1}{\hat{\epsilon}}} 254,6 \text{ K}. \quad (7)$$

Setzen wir für  $T$  die beobachtete Oberflächentemperatur von 288 K ein, erhalten wir  $\hat{\epsilon} = 0,61$ . Wir haben an dieser Stelle unser „Klimamodell“ nicht zur Bestimmung der Oberflächentemperatur  $T$  benutzt, sondern zur Bestimmung von  $\hat{\epsilon}$ . Gelegentlich wird behauptet, der niedrigere Wert von  $\hat{\epsilon}$  beschreibe den Treibhauseffekt. Das ist zwar nicht falsch, bedarf aber einer Erklärung. Inwieweit  $\hat{\epsilon}$  direkt gemessen und somit aus diesem Modell die Oberflächentemperatur der Erde bestimmt werden kann, werden wir noch diskutieren.

### 2.3 Klimamodell 3 - eine total absorbierende Atmosphärensenschicht

Wir betrachten nun ein Modell, bei dem zwischen der einfallenden Strahlung, die durch die Solarkonstante gegeben ist, und der ausfallenden Strahlung, die in den bisherigen beiden Modellen durch die Temperatur der Erdoberfläche gegeben ist, noch eine Schicht liegt, welche den Einfluss der Treibhausgase beschreiben soll. Diese Schicht sei vollkommen durchlässig für die einfallende (kurzwellige) Strahlung der Sonne, sie absorbiere aber die langwellige Strahlung zur Temperatur  $T_s$ , die von der Erde emittiert wird. Außerdem emittiere sie eine langwellige Strahlung zu ihrer Atmosphärentemperatur  $T_a$  (siehe Abb. 3). Diese Emission erfolgt sowohl in Richtung Weltraum (nach oben) als auch zurück in Richtung Erdoberfläche. Wir haben es nun also mit zwei Temperaturen zu tun: einmal die Temperatur der Erdoberfläche  $T_s$  sowie weiterhin die Temperatur  $T_a$  der Atmosphärensenschicht, die wir uns als sehr dünn denken.

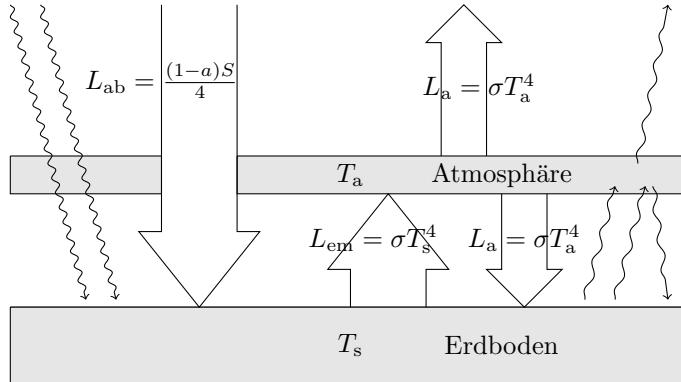


Abbildung 3: Der Treibhauseffekt - Klimamodell mit einer Atmosphärensenschicht. Die von der Erdoberfläche emittierte langwellige Strahlung wird in der Atmosphäre absorbiert und als Schwarzkörperstrahlung zur Atmosphärentemperatur  $T_a$  sowohl in den Weltraum als auch zurück zur Erdoberfläche emittiert.

Für beide Schichten - Erdoberfläche und Atmosphäre - stellen wir nun Bilanzgleichungen auf. Dabei nehmen wir an, dass die Atmosphäre nur Energie von der Erdoberfläche erhält, die entsprechend dem Stefan-Boltzmann-Gesetz durch  $L_{\text{em}} = \sigma T_s^4$  gegeben ist. Außerdem emittiere die Atmosphäre die Energie in zwei Richtungen - nach oben und nach unten - entsprechend ihrer Temperatur  $T_a$ . Wir erhalten somit die folgenden zwei Bilanzgleichungen ((s) für „surface“, d.h. die Bilanzgleichung des Erdbodens, und (a) für die Bilanzgleichung der Atmosphärensenschicht):

$$(s) \quad \frac{(1 - a)}{4}S + \sigma T_a^4 = \sigma T_s^4 \quad (a) \quad \sigma T_s^4 = 2\sigma T_a^4. \quad (8)$$

Bilanzgleichung (s) beschreibt die Strahlungsbilanz der Erdoberfläche: Einfallend (linke Seite der Gleichung) ist die kurzwellige Strahlung der Sonne sowie ein Teil, der von der Atmosphäre nach unten emittiert wird; ausfallend ist die langwellige Strahlung, welche die Erdoberfläche entsprechend ihrer

Temperatur  $T_s$  emittiert. Bilanzgleichung (a) beschreibt die Strahlungsbilanz der Atmosphärenschicht (die als dünn und von konstanter Temperatur angenommen wird): Einfallend ist die emittierte langwellige Strahlung der Erde, emittiert werden zwei Anteile (nach unten und nach oben in den Weltraum) entsprechend der Temperatur  $T_a$ .

Aus der zweiten Gleichung folgt unmittelbar  $T_s = \sqrt[4]{2} T_a$ , die Temperatur der Erdoberfläche ist also um den Faktor  $\sqrt[4]{2} \approx 1,1892$  höher als die Temperatur  $T_a$  der Atmosphäre. Setzt man die zweite Bilanzgleichung in die erste ein, erhält man für die Atmosphärentemperatur den Wert des ersten Modells  $T_a \approx 254,6\text{ K}$  oder  $-18^\circ\text{C}$ . Dies ist auch plausibel, weil für das Gesamtsystem die Strahlungsbilanz ebenfalls ausgeglichen sein soll. Für die Temperatur der Erdoberfläche erhalten wir nun  $T_s \approx 303\text{ K}$  oder  $30^\circ\text{C}$ . Das ist etwas zu warm wenn man bedenkt, dass die mittlere Temperatur der Erdoberfläche eher bei  $15^\circ\text{C}$  liegt.

An dieses Modell wird offensichtlich, dass die Temperatur, die man aus einer großen Entfernung von der Erde beobachtet, nicht die Temperatur der Erdoberfläche ist (wie in dem ersten Modell), sondern die Temperatur der (oberen) Atmosphärenschicht.

## 2.4 Klimamodell 4 - eine teilabsorbierende Atmosphärenschicht

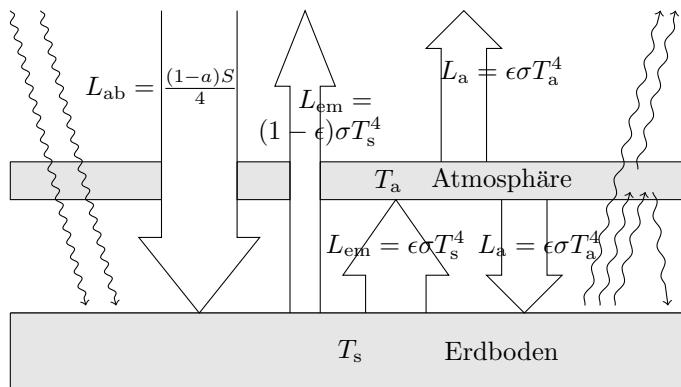


Abbildung 4: Effektiver Treibhauseffekt mit einer teilweise absorbierenden Atmosphärenschicht. Die von der Erdoberfläche emittierte langwellige Strahlung wird teilweise in der Atmosphäre absorbiert und teilweise durchgelassen. Der absorbierte Anteil wird als „graue Strahlung“ mit Emissionskoeffizienten  $\epsilon$  zur Atmosphärentemperatur  $T_a$  sowohl in den Weltraum als auch zurück zur Erdoberfläche emittiert.

In diesem Klimamodell wird angenommen, dass die Atmosphärenschicht nur einen Teil der von der Erde emittierten langwelligen Strahlung absorbiert (parametrisiert durch den Parameter  $\epsilon$ ) und einen Teil direkt durchlässt (siehe Abb. 4). Entsprechend den Kirchhoff'schen Strahlungsgesetzen, wonach Absorptions- und Emissionskoeffizient eines Körpers für eine gegebene Wellenlänge gleich sind (Richtungsabhängigkeiten werden hier vernachlässigt), wird bei diesem Modell angenommen, dass die Atmosphäre für bestimmte Wellenlängen transparent ist - und entsprechend bei diesen Wellenlängen auch nicht emittiert - und andere Wellenlängen absorbiert, die dann auch wieder (diffus, also in alle Richtungen) emittiert werden.

Nun lauten die Bilanzgleichungen:

$$(s) \quad \frac{(1-a)}{4} S + \epsilon \sigma T_a^4 = ((1-\epsilon) + \epsilon) \sigma T_s^4 = \sigma T_s^4 \quad (a) \quad \epsilon \sigma T_s^4 = 2 \epsilon \sigma T_a^4. \quad (9)$$

Aus der zweiten Gleichung (für die Atmosphärenschicht) folgt:

$$T_s^4 = 2 T_a^4 \quad \text{bzw.} \quad T_s = \sqrt[4]{2} T_a. \quad (10)$$

Wenn wir dies in die erste Gleichung einsetzen, erhalten wir für die Atmosphärentemperatur:

$$T_a = \sqrt[4]{\frac{(1-a)S}{4} \frac{1}{2-\epsilon}} = \sqrt[4]{\frac{1}{2-\epsilon}} 254,6\text{ K} \quad (11)$$

und damit für die Bodentemperatur der Erde:

$$T_s = \sqrt[4]{\frac{(1-\epsilon)S}{4}} \frac{2}{2-\epsilon} = \sqrt[4]{\frac{2}{2-\epsilon}} 254,6 \text{ K}. \quad (12)$$

Für  $\epsilon \rightarrow 0$  erhalten wir unser erstes Modell wieder: Die Atmosphäre emittiert keine Strahlung mehr (ihre Temperatur wäre  $T_a = (1/\sqrt[4]{2}) \cdot 254,6 \text{ K} \approx 214,1 \text{ K}$ , bei der aber praktisch keine Strahlung emittiert wird). Sämtliche von der Erdoberfläche absorbierte Energie wird als Schwarzkörperstrahlung bei 254,6 K wieder in den Weltraum abgestrahlt. Für  $\epsilon \rightarrow 1$  erhalten wir das vorherige Modell: Die Atmosphärentemperatur beträgt 254,6 K und die Temperatur der Erdoberfläche ist  $T_s = \sqrt[4]{2} \cdot 254,6 \text{ K} \approx 303 \text{ K}$ .

Wenn wir nun umgekehrt für die durchschnittliche Oberflächentemperatur der Erde den Wert 288 K annehmen, erhalten wir den Wert  $\epsilon \approx 0,7785$ , und für die Atmosphärentemperatur folgt daraus ein Wert von rund 242 K. Über Polareis (siehe Abb. 5) würde sich für eine Erdtemperatur von 268 K eine Atmosphärentemperatur von 226 K ergeben, was recht gut mit den beobachteten Werten übereinstimmt. Der Vorteil dieses Modells ist, dass wir hier einige Effekte studieren und interpretieren können, die bisher nicht so offensichtlich waren.

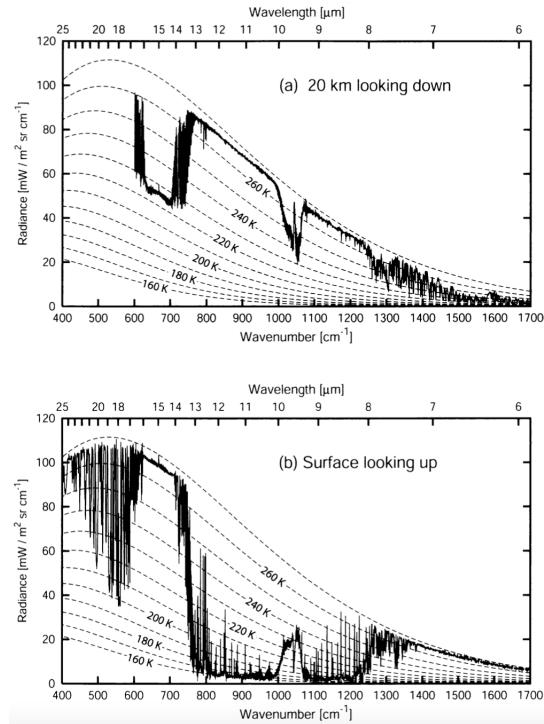


Abbildung 5: (oben) Spektralzerlegung der Erdstrahlung aus großer Höhe betrachtet. (unten) Spektrale Zerlegung der Atmosphärenstrahlung vom Erdboden aus betrachtet. Die Diagramme wurden über Polareis aufgenommen. Die wellenlängenabhängigen Kurven werden verglichen mit Schwarzkörperstrahlungen zu verschiedenen Temperaturen. Aus großer Höhe sieht man Bereiche mit einer Temperatur von rund 268 K - diese entsprechen der Erdoberfläche - und andere Bereiche mit einer Temperatur von rund 226 K - diese entsprechen der oberen Atmosphäre. In Wellenlängenbereichen, in denen die Atmosphäre lichtdurchlässig ist, sieht man im oberen Bild eine Temperatur von rund 268 K, also den Erdboden, und im unteren Bild eine sehr niedrige Temperatur entsprechend dem Welt Raum. Aus [3].

1. Für  $\epsilon \neq 1$  oder 0 entspricht die Strahlung, welche die Erde in den Weltraum abgibt - die also von einer großen Höhe aus beobachtet wird - nicht einer Schwarzkörperstrahlung zu einer festen Temperatur sondern einer Strahlung zu zwei Temperaturen. Damit erhält der Parameter  $\epsilon$  aus dem zweiten Modell eine anschauliche Bedeutung: Bei manchen Wellenlängen ist die Atmosphäre durchlässig und man sieht somit bei diesen Wellenlängen aus großer Höhe die Oberfläche der Erde (und ihre Temperatur). Bei anderen Wellenlängen ist die Atmosphäre absorbierend und undurchsichtig. Bei diesen Wellenlängen sieht man aus einer großen Höhe nur die obere Atmosphärenschicht und dies bei einer deutlich niedrigeren Temperatur (vgl. Abb. 5).

2. Wenn  $\epsilon$  größer wird, also mehr Strahlung von der Atmosphäre absorbiert wird, nimmt die Temperatur der Erdoberfläche zu. Dies ist der eigentliche Treibhauseffekt. Gleichzeitig nimmt in diesem Modell aber auch die Temperatur der oberen Atmosphärenschicht, von der aus die Strahlung in den Weltraum abgegeben wird, zu. Sie ist allerdings insgesamt kälter als in dem Klimamodell 3. Da die einfallende Intensität (die Solarkonstante und die Albedo) mehr oder weniger konstant bleibt, muss auch die abgegebene Intensität in der Summe konstant bleiben. Wenn aber von der Erdoberfläche Strahlung zu einer höheren Temperatur in den Weltraum abgegeben wird, muss die Strahlung von den oberen Atmosphärenschichten zu einer tieferen Temperatur gehören.

## 2.5 Ein einfaches Klimamodell für die Venus

Die Atmosphäre der Venus enthält rund 96% Kohlendioxid. Der Abstand Venus-Sonne beträgt ungefähr 0,72 AU (astronomische Einheiten) und somit ist die Solarkonstante an der Oberfläche der Venusatmosphäre um rund das Inverse von  $(0,72)^2 \approx 0,52$ , also knapp das Doppelte, größer als auf der Erde (dies beruht auf dem  $1/r^2$ -Gesetz der Intensität einer Strahlung als Funktion des Abstands von der Strahlungsquelle). Allerdings ist die Albedo der Venus mit rund 0,75 deutlich größer als die der Erde (von außen erscheint die Venus vollkommen wolkenbedeckt). Insgesamt müsste es daher auf der Venus kälter sein, als auf der Erde. Doch die Oberflächentemperatur der Venus beträgt über 460°C. Um zu sehen, wie eine dickere absorbierende Atmosphärenschicht die Oberflächentemperatur beeinflusst, betrachten wir folgendes Atmosphärenmodell (siehe Abb. 6).

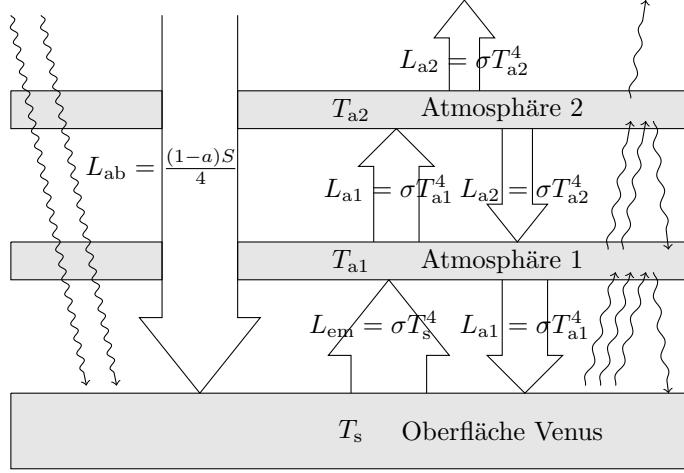


Abbildung 6: Modell für den Treibhauseffekt der Venus - ein Klimamodell mit zwei Atmosphärenschichten. Die von der Venusoberfläche emittierte langwellige Strahlung wird in der ersten Atmosphäre absorbiert und als Schwarzkörperstrahlung zur Atmosphärentemperatur  $T_{a1}$  zur zweiten Schicht wie auch zurück zur Venusoberfläche emittiert. Die zweite Schicht mit der Temperatur  $T_{a2}$  emittiert Strahlung zurück zur ersten Schicht und in den Weltraum.

Die Bilanzgleichungen sind nun (wobei durch die Stefan-Boltzmann-Konstante  $\sigma$  dividiert wurde):

$$(g) \quad \frac{(1-a)S}{4\sigma} + T_{a1}^4 = T_s^4 \quad (a1) \quad T_s^4 + T_{a2}^4 = 2T_{a1}^4 \quad (a2) \quad T_{a1}^4 = 2T_{a2}^4 \quad (13)$$

Aus diesen Gleichungen erhalten wir (die Zahlenwerte gelten zum besseren Vergleich mit den vorherigen Modellen für die Albedo und die Solarkonstante der Erde):

$$T_{a2} = \sqrt[4]{\frac{(1-a)S}{4\sigma}} = 254,6 \text{ K} \quad (14)$$

$$T_{a1} = \sqrt[4]{2} T_{a2} = \sqrt[4]{2} \cdot \sqrt[4]{\frac{(1-a)S}{4\sigma}} = \sqrt[4]{2} \cdot 254,6 \text{ K} \approx 302,8 \text{ K} \quad (15)$$

$$T_s = \sqrt[4]{3} T_{a2} = \sqrt[4]{3} \cdot \sqrt[4]{\frac{(1-a)S}{4\sigma}} = \sqrt[4]{3} \cdot 254,6 \text{ K} \approx 335,1 \text{ K} \quad (16)$$

Die oberste Atmosphärenschicht behält somit ihre Temperatur von 254,6 K, wohingegen die Temperatur am Boden im Vergleich zu dem Einschicht-Modell nochmals gewachsen ist. Man kann sich nun leicht überlegen, dass sich dieses Verhalten bei mehreren Atmosphärenschichten fortsetzt. Letztendlich erhält man, dass immer dieselbe Temperatur von 254,6 K in den Weltraum abgestrahlt wird, diese Temperatur zu den tieferen Schichten hin zunimmt und am Boden theoretisch einen beliebig hohen Wert annehmen kann. Auch wenn sich die Zahlenwerte hier auf die Albedo und die Solarkonstante der Erde beziehen, kann man das Modell leicht auf Venus übertragen. Durch die hohe Konzentration an CO<sub>2</sub> (sowie die Tatsache, dass die Atmosphäre der Venus fast 90 mal dichter ist als die der Erde) wirken dort effektiv viele Atmosphärenschichten total absorzierend.

Modelle für die Atmosphären anderer Planeten oder auch von Monden anderer Planeten (z.B. der Jupiter- und Saturnmonde) sind auch für die Klimaphysik von Interesse, weil man an diesen Modellen testen kann, ob die Zusammenhänge durch die einfachen Modelle schon nahezu korrekt wiedergegeben werden und somit diese einfachen Modelle der Realität entsprechen.

## 2.6 Klimamodell 5 - Erwärmung der Troposphäre und Abkühlung der Stratosphäre

Nun betrachten wir ein Klimamodell mit zwei Atmosphärenschichten, von denen eine Schicht näherungsweise die Verhältnisse in der Troposphäre und eine zweite Schicht die Verhältnisse in der Stratosphäre beschreibt. Es ist das allgemeinste Modell in diesem Kapitel und man erhält alle vorherigen Modelle für spezielle Werte der Parameter.

Die Stratosphäre zeichnet sich dadurch aus, dass sie einen Teil der einfallenden Sonnenstrahlung absorbiert (insbesondere im UV-Bereich durch das in der Stratosphäre vorhandene Ozon) und in Wärme umwandelt. Ansonsten wirken sowohl die Troposphäre als auch die Stratosphäre als „graue Strahler“, die bestimmte Wellenlängen absorbieren und emittieren und andere Wellenlängen transmittieren und in diesem Bereich auch nicht emittieren (siehe Abb. 7). Die zugehörigen Emissionsfaktoren (gleich Absorptionsfaktoren)  $\epsilon_{\text{tro}}$  und  $\epsilon_{\text{str}}$  geben wieder den Anteil einer Strahlung an, die von dieser Schicht absorbiert und emittiert wird.

Die Bilanzgleichungen für die drei Schichten lassen sich anhand von Abb. 6 ablesen und lauten:

$$\text{Erdboden} \quad (1 - \beta_{\text{str}}) \frac{(1-a)S}{4} + \epsilon_{\text{tro}} \sigma T_{\text{tro}}^4 + \epsilon_{\text{str}} (1 - \epsilon_{\text{tro}}) \sigma T_{\text{str}}^4 = \sigma T_{\text{s}}^4 \quad (17)$$

$$\text{Troposphäre} \quad \epsilon_{\text{tro}} \sigma T_{\text{s}}^4 + \epsilon_{\text{str}} \epsilon_{\text{tro}} \sigma T_{\text{str}}^4 = 2 \epsilon_{\text{tro}} \sigma T_{\text{tro}}^4 \quad (18)$$

$$\text{Stratosphäre} \quad \beta_{\text{str}} \frac{(1-a)S}{4} + \epsilon_{\text{tro}} \epsilon_{\text{str}} \sigma T_{\text{tro}}^4 + \epsilon_{\text{str}} (1 - \epsilon_{\text{tro}}) \sigma T_{\text{s}}^4 = 2 \epsilon_{\text{str}} \sigma T_{\text{str}}^4 \quad (19)$$

Wir können diese drei Gleichungen in Matrixform schreiben:

$$\begin{pmatrix} 1 & -\epsilon_{\text{tro}} & -\epsilon_{\text{str}}(1 - \epsilon_{\text{tro}}) \\ -\epsilon_{\text{tro}} & 2\epsilon_{\text{tro}} & -\epsilon_{\text{tro}}\epsilon_{\text{str}} \\ -\epsilon_{\text{str}}(1 - \epsilon_{\text{tro}}) & -\epsilon_{\text{tro}}\epsilon_{\text{str}} & 2\epsilon_{\text{str}} \end{pmatrix} \begin{pmatrix} \sigma T_{\text{s}}^4 \\ \sigma T_{\text{tro}}^4 \\ \sigma T_{\text{str}}^4 \end{pmatrix} = \begin{pmatrix} (1 - \beta_{\text{str}}) \frac{(1-a)S}{4} \\ 0 \\ \beta_{\text{str}} \frac{(1-a)S}{4} \end{pmatrix} \quad (20)$$

Bildet man die Summe dieser drei Gleichungen erhält man die Bilanz des Gesamtsystems:

$$\frac{(1-a)S}{4} = (1 - \epsilon_{\text{tro}})(1 - \epsilon_{\text{str}})\sigma T_{\text{s}}^4 + \epsilon_{\text{tro}}(1 - \epsilon_{\text{str}})\sigma T_{\text{tro}}^4 + \epsilon_{\text{str}}\sigma T_{\text{str}}^4. \quad (21)$$

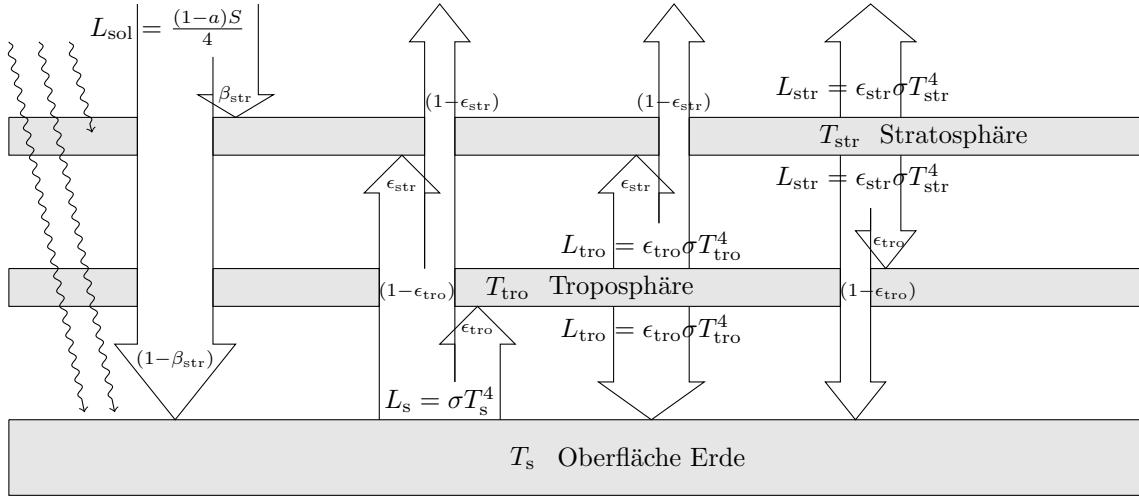


Abbildung 7: Modell für die Strahlungsbilanz mit einer Troposphäre und einer Stratosphäre. Die Stratosphäre absorbiert einen Teil der einfallenden Sonnenstrahlung (Absorptionskoeffizient  $\beta_{str}$ ) und wird dadurch erwärmt. Außerdem absorbiert sie Teile der langwelligen Infrarotstrahlung von der Erdoberfläche und der Troposphäre. Die Erdoberfläche absorbiert und emittiert als schwarzer Körper und hat somit einen Emissionskoeffizienten von 1. Die Troposphäre und Stratosphäre absorbieren und emittieren als „graue Strahler“. Die Troposphäre hat einen Absorptions- und Emissionskoeffizienten von  $\epsilon_{tro}$  und die Stratosphäre einen entsprechenden Koeffizienten  $\epsilon_{str}$ . Der durchgelassene Strahlungsanteil ist jeweils proportional zu  $(1 - \epsilon_{tro/str})$ .

In Anhang 3.1 wird dieses Gleichungssystem gelöst. Aus den Gleichungen 36–38 ergibt sich:

$$T_{str} = \sqrt[4]{\frac{1}{(2 - \epsilon_{str})} \left( 1 + \frac{(1 - \epsilon_{str})}{\epsilon_{str}} \beta_{str} \right)} \cdot 254,6 \text{ K}, \quad (22)$$

$$T_{tro} = \sqrt[4]{\frac{(2 + \epsilon_{str}(1 - \epsilon_{tro}) - (\epsilon_{str} + \epsilon_{tro}(1 - \epsilon_{str}))\beta_{str})}{(2 - \epsilon_{str})(2 - \epsilon_{tro})}} \cdot 254,6 \text{ K}, \quad (23)$$

$$T_s = \sqrt[4]{\frac{((4 - \epsilon_{str}\epsilon_{tro}) - (2 + \epsilon_{tro}(1 - \epsilon_{str}))\beta_{str})}{(2 - \epsilon_{str})(2 - \epsilon_{tro})}} \cdot 254,6 \text{ K}. \quad (24)$$

Für die Parameterwerte  $\beta_{str} = 0,05$ ,  $\epsilon_{str} = 0,11$  und  $\epsilon_{tro} = 0,78$  erhalten wir für die Bodentemperatur  $T_s = 288,06 \text{ K}$ , die Troposphärentemperatur  $T_{tro} = 245,19 \text{ K}$  und die Stratosphärentemperatur  $T_{str} = 236,37 \text{ K}$ . Erhöhen wir nun die beiden Emissionsparameter etwas (um beispielsweise eine erhöhte CO<sub>2</sub>-Konzentration in der Atmosphäre zu modellieren), z.B. auf  $\epsilon_{str} = 0,12$  und  $\epsilon_{tro} = 0,8$ , erhalten wir für die Bodentemperatur  $T_s = 289,43 \text{ K}$ , die Tropospärentemperatur  $T_{tro} = 246,50 \text{ K}$  und die Stratosphärentemperatur  $T_{str} = 235,07 \text{ K}$ . Wir erkennen, dass die Bodentemperatur ebenso wie die Troposphärentemperatur zugenommen hat, wohingegen die Stratosphärentemperatur etwas abgenommen hat. Allerdings zeigt Gl. 36 (bzw. Gl. 22), dass die Stratosphärentemperatur in diesem Modell nicht von dem Emissionsparameter der Troposphäre abhängt. Außerdem ist die Absorption der einfallenden Strahlung für diesen Effekt wichtig: Für  $\beta = 0$  nimmt die Stratosphärentemperatur mit wachsendem  $\epsilon_{str}$  zu. Die Abnahme erfolgt nur, wenn  $\beta > \epsilon_{str}^2/2$  ist (Anmerkung (3))

### 3 Anmerkungen

(1) Der Grund für den Faktor 2 zwischen der Schwankung im Abstand und der Schwankung in der Intensität der Sonneneinstrahlung liegt in dem  $1/r^2$ -Gesetz der Intensität als Funktion des Abstands:

$$\frac{1}{(r \pm \Delta r)^2} \approx \frac{1}{r^2} \mp 2 \frac{\Delta r}{r}. \quad (25)$$

(2) Das Stefan-Boltzmann-Gesetz folgt z.B. aus dem Planck'schen Strahlungsgesetz. Damit lässt sich die Stefan-Boltzmann-Konstante durch andere Naturkonstanten ausdrücken:

$$\sigma = \frac{2\pi^5 k_B^4}{15 h^3 c^2} = 5,670\,374\,42\dots \cdot 10^{-8} \frac{\text{W}}{\text{m}^2 \text{K}^4}. \quad (26)$$

(3) Die genaue Bedingung lautet

$$\beta_{\text{str}} > \frac{\epsilon_{\text{str}}^2}{2 - 2\epsilon_{\text{str}} + \epsilon_{\text{str}}^2}, \quad (27)$$

was aber für kleine Werte von  $\epsilon_{\text{str}}$ , die durchaus realistisch sind, durch  $\beta_{\text{str}} > \epsilon_{\text{str}}^2/2$  genähert werden kann. Man erhält die Bedingung aus der Ableitung von  $T_{\text{str}}$  nach  $\epsilon_{\text{str}}$ , die für eine abnehmende Temperatur als Funktion von  $\epsilon_{\text{str}}$  negativ sein muss.

#### 3.1 Lösung des Gleichungssystems für das Stratosphärenmodell

Aus Gl. 18 folgt:

$$T_{\text{s}}^4 + \epsilon_{\text{str}} T_{\text{str}}^4 = 2T_{\text{tro}}^4. \quad (28)$$

Wir nutzen diese Gleichung, um  $T_{\text{tro}}$  aus dem Gleichungssystem zu eliminieren und erhalten:

$$(1 - \beta_{\text{str}}) \frac{(1-a)S}{4} + \epsilon_{\text{str}} \left(1 - \frac{1}{2}\epsilon_{\text{tro}}\right) \sigma T_{\text{str}}^4 = \left(1 - \frac{1}{2}\epsilon_{\text{tro}}\right) \sigma T_{\text{s}}^4 \quad (29)$$

$$\beta_{\text{str}} \frac{(1-a)S}{4} + \epsilon_{\text{str}} \left(1 - \frac{1}{2}\epsilon_{\text{tro}}\right) \sigma T_{\text{s}}^4 = 2\epsilon_{\text{str}} \left(1 - \frac{1}{4}\epsilon_{\text{tro}}\epsilon_{\text{str}}\right) \sigma T_{\text{str}}^4 \quad (30)$$

oder in Matrixschreibweise:

$$\begin{pmatrix} \left(1 - \frac{1}{2}\epsilon_{\text{tro}}\right) & -\epsilon_{\text{str}} \left(1 - \frac{1}{2}\epsilon_{\text{tro}}\right) \\ -\epsilon_{\text{str}} \left(1 - \frac{1}{2}\epsilon_{\text{tro}}\right) & 2\epsilon_{\text{str}} \left(1 - \frac{1}{4}\epsilon_{\text{tro}}\epsilon_{\text{str}}\right) \end{pmatrix} \begin{pmatrix} \sigma T_{\text{s}}^4 \\ \sigma T_{\text{str}}^4 \end{pmatrix} = \begin{pmatrix} (1 - \beta_{\text{str}}) \frac{(1-a)S}{4} \\ \beta_{\text{str}} \frac{(1-a)S}{4} \end{pmatrix} \quad (31)$$

Nach der allgemeinen Formel für das Inverse einer  $2 \times 2$ -Matrix,

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \implies A^{-1} = \frac{1}{(ad - bc)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \quad (32)$$

folgt mit

$$A = \begin{pmatrix} \left(1 - \frac{1}{2}\epsilon_{\text{tro}}\right) & -\epsilon_{\text{str}} \left(1 - \frac{1}{2}\epsilon_{\text{tro}}\right) \\ -\epsilon_{\text{str}} \left(1 - \frac{1}{2}\epsilon_{\text{tro}}\right) & 2\epsilon_{\text{str}} \left(1 - \frac{1}{4}\epsilon_{\text{tro}}\epsilon_{\text{str}}\right) \end{pmatrix} \quad (33)$$

und

$$\det A = 2\epsilon_{\text{str}} \left(1 - \frac{1}{2}\epsilon_{\text{tro}}\right) \left(1 - \frac{1}{2}\epsilon_{\text{str}}\right) \quad (34)$$

die Gleichung

$$\begin{pmatrix} \sigma T_{\text{s}}^4 \\ \sigma T_{\text{str}}^4 \end{pmatrix} = \frac{1}{\det A} \begin{pmatrix} 2\epsilon_{\text{str}} \left(1 - \frac{1}{4}\epsilon_{\text{tro}}\epsilon_{\text{str}}\right) & \epsilon_{\text{str}} \left(1 - \frac{1}{2}\epsilon_{\text{tro}}\right) \\ \epsilon_{\text{str}} \left(1 - \frac{1}{2}\epsilon_{\text{tro}}\right) & \left(1 - \frac{1}{2}\epsilon_{\text{tro}}\right) \end{pmatrix} \begin{pmatrix} (1 - \beta_{\text{str}}) \\ \beta_{\text{str}} \end{pmatrix} \frac{(1-a)S}{4} \quad (35)$$

Wir erhalten für die Stratosphärentemperatur:

$$T_{\text{str}}^4 = \frac{1}{(2 - \epsilon_{\text{str}})} \left( 1 + \frac{(1 - \epsilon_{\text{str}})}{\epsilon_{\text{str}}} \beta_{\text{str}} \right) \cdot \frac{(1 - a)S}{4\sigma}, \quad (36)$$

und für die Bodentemperatur:

$$T_{\text{s}}^4 = \frac{((4 - \epsilon_{\text{str}}\epsilon_{\text{tro}}) - (2 + \epsilon_{\text{tro}}(1 - \epsilon_{\text{str}}))\beta_{\text{str}})}{(2 - \epsilon_{\text{str}})(2 - \epsilon_{\text{tro}})} \cdot \frac{(1 - a)S}{4\sigma}. \quad (37)$$

Schließlich ergibt sich für die Temperatur der Troposphäre:

$$T_{\text{tro}}^4 = \frac{(2 + \epsilon_{\text{str}}(1 - \epsilon_{\text{tro}}) - (\epsilon_{\text{str}} + \epsilon_{\text{tro}}(1 - \epsilon_{\text{str}}))\beta_{\text{str}})}{(2 - \epsilon_{\text{str}})(2 - \epsilon_{\text{tro}})} \cdot \frac{(1 - a)S}{4\sigma}. \quad (38)$$

## Literatur

- [1] Wild, M., Ohmura, A., Schär, Ch., Müller, G., Folini, D., Schwarz, M., Hakuba, M.Z., Sanchez, A.; *The Global Energy Balance Archive (GEBA) version 2017: a database for worldwide measured surface energy fluxes*; Earth Syst. Sci. Data 9 (2017) 601–613.  
<https://www.klimanavigator.eu/dossier/artikel/011967/index.php>
- [2] NASA Earth Fact Sheet; <https://nssdc.gsfc.nasa.gov/planetary/factsheet/earthfact.html>
- [3] Grant W. Petty; *A First Course in Atmospheric Radiation*, 2nd Ed.; Sundog Publishing, Madison, Wisconsin; 2006.
- [4] Wikipedia „Solarkonstante“. <https://de.wikipedia.org/wiki/Solarkonstante>.

# Das Qubit

Thomas Filk; Universität Freiburg

Das QuBit ist die Einheit der Quanteninformation. Es ist das Quantensystem, das der klassischen Informationseinheit, dem Bit als die Menge bestehend aus zwei Möglichkeiten - {Ja, Nein}, {Wahr, Falsch}, {1,0} - entspricht. Man bezeichnet das QuBit als ein quantenmechanisches Zwei-Zustandssystem, obwohl, wie wir sehen werden, ein solches System wesentlich mehr als zwei mögliche Zustände einnehmen kann - es sind sogar unendliche viele Zustände möglich. Es zeigt sich jedoch, dass eine Messung an solchen Systemen - d.h. das Ablesen oder Auslesen eines solchen Systems - immer nur maximal ein klassisches Bit als Ergebnis, also eine von zwei möglichen Antworten, zulässt. Allerdings gibt es viele Möglichkeiten, wie dieses Ablesen realisiert werden kann bzw. welche Alternativen man in einer solchen Messung auszeichnet.

Es gibt viele Realisierungen solcher Zwei-Zustandssysteme, beispielsweise die Polarisationsfreiheitsgrade eines Photons, die Spinfreiheitsgrade eines Elektrons oder mancher Atome, bestimmte atomare Zweizustandssysteme, manche Ionenfallen, etc. In diesem Abschnitt beschränken wir uns auf die Polarisationsfreiheitsgrade von Photonen, da diese nicht nur sehr anschaulich sind, sondern auch häufig zu Demonstrationszwecken verwendet werden.

## 1 Quantenmechanische Grundbegriffe

Bevor wir konkret auf quantenmechanische Zwei-Zustandssysteme bzw. das QuBit eingehen, sollen einige Grundbegriffe der Quantentheorie, die im Folgenden eine wichtige Rolle spielen, nochmals diskutiert werden. Dabei ist aber zu bedenken, dass schon in Bezug auf diese Grundbegriffen die Meinungen hinsichtlich der Deutung oder Interpretation oftmals weit auseinandergehen.

### 1.1 Observable

Eine Observable ist etwas, das man an einem System messen kann. Die physikalische Realisierung einer Observablen erfolgt durch die Angabe des Messprotokolls, wie eine Messung dieser Observablen an einem System durchzuführen ist. Das Ergebnis einer solchen Messung ist eine Zahl.

Die mathematische Darstellung (Repräsentation) einer Observablen hängt von der Theorie ab, die wir verwenden. In der klassischen Mechanik handelt es sich bei Observablen um Funktionen vom Ort und Impuls (bzw. der Geschwindigkeit), in der Quantenmechanik werden Observable durch sogenannte selbst-adjungierte (oder hermitesche) Operatoren auf einem Hilbert-Raum - einem Vektorraum mit einem Skalarprodukt - repräsentiert. Hier sollte man betonen, dass die mathematische Darstellung einer Observablen nicht den Messprozess repräsentiert (also den dynamische Vorgang der Messung), sondern eher die Informationen, die man durch solche Messungen erlangen kann: die Menge der möglichen Messwerte sowie die Zustände, die mit diesen Messwerten verbunden sind. In diesem Sinne und in Anlehnung an einen Ausdruck von Schrödinger [3] ist eine Observable ein „Katalog von möglichen Ergebnissen“.

### 1.2 Zustand

Auch bei dem Begriff des Zustands sollte man zwischen der physikalischen Realisierung und der mathematischen Darstellung unterscheiden. Die physikalische Realisierung eines Zustands besteht in einer im Prinzip beliebig großen Menge gleichartig präparierter Systeme. Mathematisch handelt es sich

bei einem Zustand um eine Vorschrift, einer Observablen ihren Erwartungswert zuzuordnen. Daher bezeichnet man einen Zustand auch als ein Erwartungswertfunktional. Bei der Interpretation eines Zustands gehen die Meinungen schon auseinander: Ist es sinnvoll, ein einzelnes System durch einen Zustand zu beschreiben oder bezieht sich ein Zustand immer nur auf ein Ensemble von gleichartig präparierten Systemen?

Ist unser Wissen unvollständig, verwenden wir sogenannte gemischte Zustände (in der klassischen Mechanik sind das Wahrscheinlichkeitsverteilungen über einem Zustandsraum, beispielsweise dem Phasenraum; in der Quantentheorie sind das sogenannte Dichtematrizen). Schrödinger bezeichnet einen Zustand, sowohl einen reinen Zustand als auch einen gemischten Zustand, als einen „Katalog von Erwartungen“ [3]. Ein Zustand ist somit eine mathematische Kodierung unseres Wissens über die Art, wie ein System präpariert wurde, sodass wir dieses Wissen für die Vorhersage zukünftiger Messungen an diesem System nutzen können.

In der klassischen Mechanik beschreiben wir einen (reinen Zustand) durch einen Punkt im Phasenraum, also die Angabe eines Ortes und eines Impulses. Diese Angabe ordnet einer Observablen (also einer Funktion über dem Phasenraum) eine Zahl zu: Ihren Wert an der Stelle dieses Phasenraumpunktes. Dies ist der Messwert, den wir bei einer Messung dieser Observablen in diesem Zustand erwarten.

In der Quantenmechanik beschreiben wir einen Zustand mathematisch durch einen Strahl (einen eindimensionalen Unterraum) in einem Hilbert-Raum. Meist wählen wir zur einfacheren Beschreibung einen auf eins normierten Vektor  $|\psi\rangle$  auf diesem Strahl als Repräsentanten, wir können einen Zustand aber auch durch die Angabe eines eindimensionalen Projektionsoperators  $P_\psi$  (der jeden Vektor in dem Hilbert-Raum auf diesen eindimensionalen Unterraum projiziert) darstellen. In der Quantenmechanik von Punktteilchen verwenden wir zur Beschreibung eines Zustands oft eine sogenannte Wellenfunktion  $\psi(x)$ , die man aber als Vektor in einem unendlich dimensionalen Hilbert-Raum (dem Raum der quadratintegrierbaren Funktionen) auffassen kann. Das Erwartungswertfunktional, d.h. die Vorschrift, nach der wir einer Observablen  $A$  eine Zahl  $\langle A \rangle_\psi$  - ihren Erwartungswert in dem Zustand  $\psi$  - zuordnen, ist dann

$$\langle A \rangle_\psi = \langle \psi | A | \psi \rangle = \text{Spur}(P_\psi A) = \int_V \psi(x)^* A \psi(x) dx. \quad (1)$$

Dies sind drei Darstellungen des Erwartungswertfunktionalen, je nachdem, ob man einen Zustand durch einen normierten Vektor, einen Projektionsoperator oder eine normierte Wellenfunktion über einem Volumen  $V$  repräsentiert.

### 1.3 Messung

Der Begriff der Messung ist in der Quantentheorie sehr umstritten und nicht wenige Physiker\*innen, die sich mit den Grundlangen der Quantentheorie beschäftigen, plädieren dafür, diesen Begriff in der Quantentheorie nach Möglichkeit zu vermeiden (ein Beispiel ist [2]). In unserer von der klassischen Physik geprägten Vorstellung wird „Messung“ oft mit „Beobachtung“ identifiziert, d.h., einem Informationsgewinn über ein System, ohne dass dieses System in irgendeiner Form dadurch beeinflusst oder gestört wird. Insbesondere impliziert der Begriff der Messung in der klassischen Physik meist eine Information über den Zustand des Systems unmittelbar *bevor* die Messung vorgenommen wurde. In der Quantentheorie sagt das Ergebnis eines Messprozesses jedoch oft sehr wenig über den Zustand vor der Messung aus, hingegen gibt es uns eine Information über den Zustand des Systems unmittelbar *nach* der Messung. Das „Kollapspostulat“ (korrekter spricht man auch von dem „von Neumann-Lüders-Projektionspostulat“) besagt, dass unmittelbar nach einer Messung ein Quantensystem durch einen

Zustand zu beschreiben ist, der dem beobachteten Messwert entspricht.<sup>1</sup> In diesem Sinne beschreibt eine Messung also eher eine Präparation eines Systems.

Obwohl diese Tatsache in manchen Bereichen (z.B. bei der Messung von Polarisationszuständen von Licht) bekannt und vertraut ist, wird sie leicht vergessen, wenn man beispielsweise an die Ortsmessung oder Impulsmessung von Elektronen oder Atomen denkt: Es besteht die Tendenz, den gemessenen Ort eines Elektrons diesem Teilchen schon vor der Messung zuzuschreiben, obwohl es vielleicht sogar bis zum Augenblick dieser Ortsmessung durch einen wohldefinierten Impuls, also ein über den gesamten Raum ausgedehntes Quantenobjekt, beschrieben wurde. Schon der Begriff „Teilchen“ impliziert, dass man an einen wohldefinierten Ort denkt, was in der Quantentheorie nicht immer zulässig ist. Im didaktischen Bereich hat sich daher der Begriff des Quantenobjekts durchgesetzt, womit noch kein Teilchen- oder Wellencharakter (oder was auch immer für eine spezifische Eigenschaft) festgelegt sein sollte.

## 2 Zwei-Zustandssysteme

In der Quantentheorie werden Zwei-Zustandssysteme durch die Strahlen in einem zweidimensionalen komplexen Vektorraum mit Skalarprodukt beschrieben. Ein Strahl ist dabei ein komplex eindimensionaler Untervektorraum. Etwas anders ausgedrückt: Zwei Vektoren, die sich nur um einen komplexen, multiplikativen Faktor unterscheiden, gehören zum selben Strahl. Wir können also einen Strahl durch einen Vektor auf diesem Strahl (meist wählt man einen auf eins normierten Vektor) repräsentieren.

In diesem Vektorraum zeichnen wir eine Basis aus:

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (2)$$

Jeder Vektor in diesem Vektorraum lässt sich als komplexe Linearkombination dieser beiden Vektoren darstellen. Da wir als Repräsentanten für einen Strahl (eindimensionalen Unterraum) und damit einen Zustand gewöhnlich nur normierte Vektoren zulassen, entsprechen Zustände den Vektoren

$$|x\rangle = \begin{pmatrix} a \\ b \end{pmatrix} = a|0\rangle + b|1\rangle \quad |a|^2 + |b|^2 = 1. \quad (3)$$

Zu bedenken ist allerdings, dass zwei solche Vektoren, die sich nur bezüglich einer Phase (komplexe Zahl vom Betrag 1) unterscheiden, denselben Zustand beschreiben.

### 2.1 Realisation durch die Polarisation von Licht

Um ein konkretes physikalisches Zwei-Zustandssystem vor Augen zu haben, betrachten wir im Folgenden die Polarisationszustände von Licht bzw. von Photonen. Dieses Beispiel eignet sich aus mehreren Gründen besonders: (1) Die Polarisierung von Licht ist auch Schüler\*innen aus dem Alltag (polarisierte Sonnengläser) bzw. der Schule (Brewster-Winkel) bekannt; (2) dieses System wird in nahezu allen Schulversuchen zu Zwei-Zustandssystemen verwendet; (3) es ist immer noch eines der am häufigsten realisierten Zwei-Zustandssysteme in der Quantenoptik (z.B. für Grundlagenexperimente in der Quantenphysik, beispielsweise die berühmten Experimente von Aspect zum Nachweis der Verletzung von Bell'schen Ungleichungen [1]).

---

<sup>1</sup>Falls es mehrere unabhängige Zustände zu demselben Messergebnis gibt, wenn dieser Raum also mehrdimensional ist, ist derjenige Zustand zu wählen, den man durch eine orthogonale Projektion des Zustands vor der Messung auf diesen Raum erhält. Dies ist eine technische Feinheit, die im Folgenden nicht weiter relevant sein wird.

Trifft ein Lichtstrahl auf eine Grenzfläche wie Luft-Wasser oder Luft-Glas, wird ein Teil des Lichts reflektiert und ein Teil in das andere Medium gebucht. Mithilfe eines einfachen Polarisationsfilters (oder einer polarisierenden Sonnenbrille) können wir feststellen, dass das reflektierte Licht eine Richtungsabhängigkeit in Bezug auf die Orientierung dieses Filters aufweist: Unter einer bestimmten Richtung (wir bezeichnen diese Richtung oft als die Polarisationsachse des Filters) sieht man hinter dem Polfilter die größte Helligkeit, unter einer dazu orthogonalen Richtung die geringste Helligkeit. Besteht zwischen dem reflektierten Strahl und dem in das Medium gebuchten Strahl ein Winkel von  $90^\circ$  (in diesem Fall bezeichnet man den Einfallswinkel auch als Brewster-Winkel) ist dieser Effekt am ausgeprägtesten: Nahezu das gesamte reflektierte Licht tritt durch den Polfilter hindurch, wenn die Polarisationsachse parallel zur reflektierenden Grenzfläche steht, und es tritt fast kein Licht hindurch, wenn die Polarisationsachse um  $90^\circ$  gedreht wird.

Wir erkennen hier eine Eigenschaft von Licht, die wir mithilfe eines Polfilters messen können. Wir können diese Eigenschaft mit sogenannten Polarisationsstrahlteilern auch zu nahezu einhundert Prozent präparieren: Ein Polarisationsstrahlteiler besteht im Wesentlichen aus zwei Grenzflächen, die als Diagonalebene in einem Glaswürfel aufeinanderliegen und gelegentlich noch besonders beschichtet sind. Ein Lichtstrahl, der durch eine Seitenfläche senkrecht in den Würfel eindringt, wird gewöhnlich in einen Anteil aufgeteilt, der aus der gegenüberliegenden Fläche in der Verlängerung wieder austritt, und einen Anteil, der an der Diagonalfäche um  $90^\circ$  abgelenkt wird und aus einer Seitenfläche wieder austritt (siehe Abb. 1).

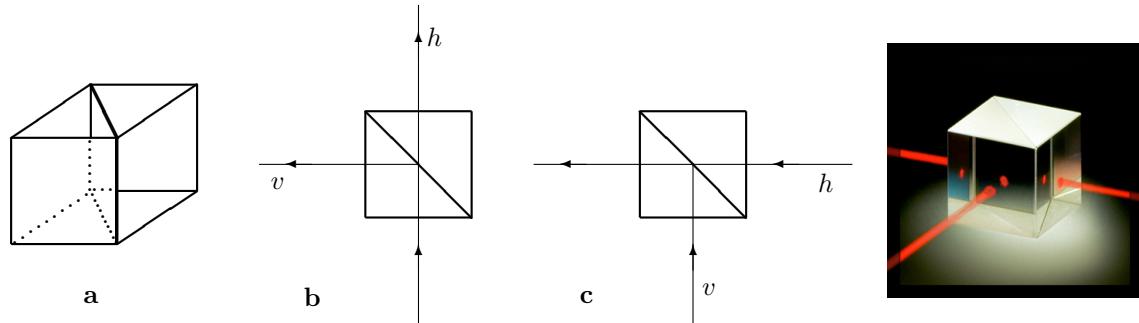


Abbildung 1: **a** Ein Polwürfel oder Polarisationsstrahlteiler besteht aus zwei zusammengesetzten Prismen mit einer besonders präparierten Grenzfläche. **b** Ein einfallender (unpolarisierter) Strahl wird in zwei (orthogonal) polarisierte Strahlen aufgespalten. Der reflektierte Strahl besitzt eine Polarisation parallel zur Grenzfläche ( $v$ ) – in der Abbildung senkrecht zur Bildebene –, der durchgelassene Strahl eine horizontale Polarisation ( $h$ ) in der Bildebene. **c** Umgekehrt kann man auch zwei geeignete polarisierte Strahlen zu einem gemeinsamen Strahl zusammenführen. Die Polarisation des ausfallenden Strahls hängt von der relativen Phase der beiden einfallenden Strahlen ab. Bei Phasengleichheit ist der ausfallende Strahl linear polarisiert, ansonsten elliptisch. Bei umgekehrter Wahl der Polarisationsrichtungen für die einfallenden Strahlen verläuft der ausfallende Strahl in der Abbildung nach oben. (rechts) Abbildung eines Polarisationsstrahlteilers. (aus [1] und [?])

In der Abbildung wurde der Polarisationswürfel so aufgestellt, dass der durchgelassene Strahl eine horizontale Polarisation besitzt, der (um  $90^\circ$ ) reflektierte Strahl eine vertikale Polarisation. Der Polarisationswürfel kann aber auch um die Achse des einfallenden Strahls gedreht werden, sodass eine Aufspaltung in zwei orthogonale Polarisationsrichtungen bezüglich jeder Orientierung des Würfels vorgenommen werden kann.

Eine „Messung“ mit einem solchen Polarisationswürfel kann dadurch erfolgen, dass hinter den beiden Austrittsrichtungen für den Lichtstrahl ein Detektor platziert wird, der das austretende Licht nachweist. Ohne die Detektoren kann man mit dieser Anordnung eine Polarisation präparieren: Man verwendet einfach den Strahl, der an der Seite zu der gewünschten Polarisation austritt, für die weiteren Versuche.

Nun ist bekannt, dass ein Lichtstrahl aus einer sehr großen Anzahl von Photonen bzw. Lichtquanten besteht. Ein Laserpointer strahlt pro Sekunde rund  $10^{18} - 10^{20}$  Photonen ab. Wenn man die Intensität eines Laserstrahls immer weiter verringert, gelangt man schließlich zu einzelnen Photonen. In der Praxis lässt sich das schwer realisieren, da es aufgrund der bosonischen Eigenschaften von Photonen zu sogenannten Bunching-Effekten kommt: Photonen treten bevorzugt als Mehrfachpakte auf. Man erhält also keine Einzelphotonen mit nahezu konstanten Zeitabständen zwischen ihnen. Auf dieses eher technische Detail soll hier aber nicht weiter eingegangen werden. Es ist im Prinzip (mit hohem Kostenaufwand) möglich, gezielt Einzelphotonen herzustellen und für die optischen Experimente, die hier eine Rolle spielen, zu verwenden.

Für Einzelphotonen gelten im Wesentlichen dieselben Regeln, wie für einen Lichtstrahl. Lediglich die Gesetze bezüglich der Intensität von Lichtstrahlen werden durch Gesetze für relative Häufigkeiten - und im Fall von Einzelphotonen durch Wahrscheinlichkeiten - ersetzt. So besagt das Gesetz von Malus, dass die Intensität eines Lichtstrahls, der einen Polarisationsfilter unter einer Orientierung zum Winkel  $\alpha$  durchquert hat und durch einen zweiten Filter tritt, der unter dem Winkel  $\beta$  orientiert ist, um den Faktor  $\cos^2(\alpha - \beta)$  verringert wird. Für einen Strahl mit  $N$  Photonen hinter dem ersten Filter lautet dieses Gesetz: Für die Anzahl  $N_1$  der Photonen, die auch den zweiten Filter passieren, gilt ungefähr:  $N_1/N = \cos^2(\alpha - \beta)$ . Und für ein Einzelphoton lautet das Gesetz: Die Wahrscheinlichkeit, dass ein Photon, welches den ersten Filter passiert hat, auch den zweiten Filter passiert, ist  $\cos^2(\alpha - \beta)$ .

Zur Beschreibung der Polarisation von Einzelphotonen verwenden wir wieder normierte Vektoren in einem zweidimensionalen komplexen Vektorraum. Die horizontale und vertikale Polarisation beschreiben wir meist durch die Einheitsvektoren aus Gl. 2, wobei wir in der Notation die klassischen Bitvariablen durch  $h$  und  $v$  ersetzen:

$$|h\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad |v\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (4)$$

Unter den Winkeln  $+45^\circ$  und  $-45^\circ$  polarisiertes Licht drücken wir entsprechend durch  $+$  und  $-$  aus. In der obigen Basis gilt für die Vektoren zu dieser Polarisation:

$$|+\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad |-\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \quad (5)$$

Linear polarisiertes Licht kann man dadurch beschreiben, dass die allgemeinen Koeffizienten in Gl. 3 reell gewählt werden können. Dies bringt zum Ausdruck, dass es keine relative Phasenverschiebung zwischen den beiden Komponenten (horizontal und vertikal) gibt. Eine Polarisation unter einem allgemeinen Winkel  $\theta$  relativ zur Horizontalen wird durch den Vektor

$$|\theta\rangle = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \quad (6)$$

beschrieben.

Neben den linearen Polarisationen gibt es noch die zirkularen Polarisationen (und ganz all-

gemein die elliptischen Polarisationen):

$$|L\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \quad |R\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}. \quad (7)$$

Hier spricht man von links zirkular polarisiertem Licht bzw. rechts zirkular polarisiertem Licht. Ganz allgemein kann eine Polarisation durch den Vektor

$$|\theta, \delta\rangle = \begin{pmatrix} \cos \theta \\ e^{i\delta} \sin \theta \end{pmatrix} \quad (8)$$

gekennzeichnet werden. Da eine beliebige gemeinsame Phase für die beiden Komponenten denselben Zustand beschreibt, kann die erste Komponente reell gewählt werden. Der Winkel  $\delta$  gibt die Elliptizität der Polarisation an ( $\delta = 0$  entspricht linear polarisiertem Licht und  $\delta = \pm 90^\circ$  entspricht R- bzw. L-zirkular polarisiertem Licht) und der Winkel  $\theta$  kennzeichnet die Lage der großen Halbachse dieser Ellipse.

## 2.2 Die Menge der Observablen

In diesem Abschnitt betrachten wir die Menge der hermiteschen Matrizen in dem zweidimensionalen komplexen Vektorraum, die wir mit der Menge der Observablen identifizieren werden. Als Basis für die hermiteschen Matrizen bieten sich die drei Pauli-Matrizen sowie die Einheitsmatrix an:

$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_3 = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}. \quad (9)$$

Diese vier Matrizen sind hermitesch (d.h. es gilt  $A_{ij} = A_{ji}^*$ ). Eine beliebige reelle Linearkombination ist ebenfalls hermitesch. Umgekehrt lässt sich jede hermitesche  $2 \times 2$ -Matrix als reelle Linearkombination in dieser Form schreiben:

$$A = a_0 \mathbf{1} + \sum_{i=1}^3 a_i \sigma_i = a_0 \mathbf{1} + \mathbf{a} \cdot \boldsymbol{\sigma}_i = \begin{pmatrix} a_0 + a_3 & a_1 + ia_2 \\ a_1 - i a_2 & a_0 - a_3 \end{pmatrix}. \quad (10)$$

Über das charakteristische Polynom erhält man die Eigenwerte  $\lambda_i$ :

$$\lambda_{1/2} = a_0 \pm \sqrt{a_1^2 + a_2^2 + a_3^2} = a_0 \pm |\mathbf{a}|. \quad (11)$$

$\sigma_1$  hat als Eigenvektoren  $|+\rangle$  und  $|-\rangle$ ,  $\sigma_2$  hat als Eigenvektoren  $|R\rangle$  und  $|L\rangle$  und  $\sigma_3$  hat die Eigenvektoren  $|h\rangle$  und  $|v\rangle$ . Da es sich bei  $A$  um eine selbst-adjungierte Matrix handelt, sind die Eigenvektoren von  $A$  immer orthogonal zueinander. Es handelt sich bei den Eigenvektoren einer Observablen also immer um zueinander orthogonale Polarisationsvektoren.

## 2.3 Die Bloch-Kugel

Die Bloch-Kugel veranschaulicht die möglichen Polarisationszustände. Aus Gl. 8 wird schon deutlich, dass man die Polarisationszustände durch zwei Winkel  $\theta$  und  $\delta$  charakterisieren kann, wobei  $\theta \in [0, \pi]$  (nach einer Drehung um  $180^\circ$  erhält man dieselbe lineare Polarisation) und  $\delta \in [-\pi, +\pi]$  (positive Vorzeichen entsprechen rechtszirkulären elliptischen Polarisationen und negative Vorzeichen linkszirkulären). Diese Darstellung legt schon nahe, dass es sich bei der Menge der Polarisationszustände um eine Kugeloberfläche handelt. Trotzdem ist die Identifikation der Winkelvariablen nicht so trivial.

Wie schon erwähnt, lassen sich die (reinen) Zustände auch durch eindimensionale Projektionsmatrizen darstellen, also hermitesche Matrizen mit der Eigenschaft  $P^2 = P$ . Diese Eigenschaft lässt nur die Eigenwerte 1 und 0 zu. Ist ein Eigenwert 1 und der andere 0, projizieren diese Matrizen auf eindimensionale Unterräume. Aus Gl. 11 erkennt man, dass diese beiden Eigenwerte nur in folgendem Fall auftreten:

$$a_0 = \frac{1}{2} \quad \text{und} \quad |\mathbf{a}| = \frac{1}{2}. \quad (12)$$

In zwei Dimensionen können wir also jede Projektionsmatrix auf einen eindimensionalen Unterraum in der Form

$$P = \frac{1}{2}(\mathbf{1} + \mathbf{n} \cdot \boldsymbol{\sigma}) \quad (13)$$

schreiben, wobei  $\mathbf{n}$  ein beliebiger 3-dimensionaler Einheitsvektor ist. Damit wird deutlich, dass sich jeder Zustand durch einen 3-dimensionalen Einheitsvektor darstellen lässt. Diese Darstellung bezeichnet man als Bloch-Kugel (siehe Abb. 2).<sup>2</sup>

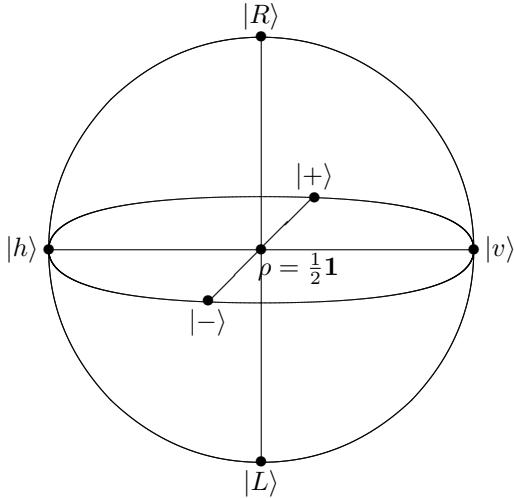


Abbildung 2: Die Bloch-Kugel für Polarisationszustände von Licht. Die Kugeloberfläche repräsentiert die reinen Zustände, das Kugellinnere die gemischten Zustände. Gegenüberliegende Punkte (Antipoden) entsprechen orthogonalen Polarisationszuständen.

Für den Fall  $|\mathbf{n}| < 1$  erhält man sogenannte Dichtematrizen. Sie beschreiben gemischte Zustände aus verschiedenen Polarisationen. Der Spezialfall  $|\mathbf{n}| = 0$  führt auf die Einheitsmatrix (multipliziert mit 1/2). In diesem Fall kann keine Aussage über die Polarisation getroffen werden bzw. alle Polarisationsrichtungen sind gleich wahrscheinlich. Dies ist der maximal gemischte Zustand. Das Innere der Bloch-Kugel kennzeichnet somit gemischte Zustände, die Oberfläche die reinen Zustände.

Man beachte, dass die Bloch-Kugel in einem 3-dimensionalen reellen Vektorraum liegt. Orthogonale Polarisationen ( $h$  und  $v$  oder  $+$  und  $-$  oder auch  $L$  und  $R$ ) liegen auf der Bloch-Kugel antipodisch gegenüber. Es handelt sich bei diesem Vektorraum daher nicht um den Hilbert-Raum der Zustände, in dem orthogonale Polarisationen auch durch orthogonale Vektoren dargestellt werden.

### 3 Verschränkung

Ebenso wie in der klassischen Informationstheorie interessiert man sich in der Quanteninformation nicht nur für einzelne QuBits, sondern für Ketten bzw. Folgen von QuBits. Im Folgenden betrachten wir nur QuBit-Paare, also Systeme zur Darstellung von zwei QuBits. Es zeigt sich, dass man auch in der Quantentheorie (ähnlich wie in der klassischen Informationstheorie) die meisten Operationen auf QuBit-Folgen als Hintereinanderausführung von Operationen auf QuBit-Paaren darstellen kann.

<sup>2</sup>Henri Poincaré hat schon 1892 diese Darstellung für die Polarisationszustände von Licht verwendet, daher wird sie in der Optik auch als Poincaré-Kugel bezeichnet.

Ein System aus zwei QuBits (z.B. die Polarisationszustände von zwei Photonen) wird im Tensorprodukt der beiden einzelnen zweidimensionalen Vektorräume beschrieben. Man erhält auf diese Weise einen vierdimensionalen komplexen Vektorraum. In einem solchen Tensorprodukt beschreibt man die Vektoren gerne mit Doppelindizes. So ist beispielsweise das Tensorprodukt aus einem Vektor  $\mathbf{x}$  und einem Vektor  $\mathbf{y}$  gegeben durch:

$$\mathbf{x} \otimes \mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \otimes \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 y_1 \\ x_1 y_2 \\ x_2 y_1 \\ x_2 y_2 \end{pmatrix}. \quad (14)$$

Das Tensorprodukt der beiden Vektorräume besteht aus allen Linearkombinationen solcher Vektoren, wobei folgendes Distributivgesetz gilt:<sup>3</sup>

$$(\mathbf{x}_1 + \mathbf{x}_2) \otimes (\mathbf{y}_1 + \mathbf{y}_2) = \mathbf{x}_1 \otimes \mathbf{y}_1 + \mathbf{x}_1 \otimes \mathbf{y}_2 + \mathbf{x}_2 \otimes \mathbf{y}_1 + \mathbf{x}_2 \otimes \mathbf{y}_2 \quad (15)$$

sowie

$$\mathbf{0} \otimes \mathbf{y} = \mathbf{x} \otimes \mathbf{0} = \mathbf{0}. \quad (16)$$

Man erkennt sofort, dass in Gl. 14 das Produkt der beiden äußeren Komponenten (die erste und vierte Komponente) gleich dem Produkt der beiden inneren Komponenten ist. Dies kennzeichnet den Vektor (in zwei Dimensionen) eindeutig als separabel, wobei ein Vektor in dem vierdimensionalen Vektorraum als separabel bezeichnet wird, wenn es zwei Vektoren  $\mathbf{x}$  und  $\mathbf{y}$  in den einzelnen Vektorräumen gibt, sodass sich der vierdimensionale Vektor als Tensorprodukt der beiden zweidimensionalen Vektoren schreiben lässt. Falls das nicht möglich ist - allgemein sich also ein Vektor in einem Tensorproduktraum nicht als Tensorprodukt von zwei Vektoren in den einzelnen Vektorräumen schreiben lässt - bezeichnet man den Vektor als verschränkt. In zwei Dimensionen haben wir somit ein einfaches Kriterium für die Verschränktheit von zwei Vektoren: Wenn das Produkt der ersten und vierten Komponente eines Vektors nicht gleich dem Produkt der zweiten und dritten Komponente ist, ist der Vektor verschränkt.

Bei Wellenfunktionen zu zwei Teilsystemen (oder allgemeiner zu zwei unabhängigen Klassen von Freiheitsgraden), die durch Funktionen von zwei Variablen beschrieben werden, bedeutet Separierbarkeit, dass  $\Psi(x,y) = \varphi(x)\psi(y)$ . Die Wellenfunktion lässt sich also als einfaches Produkt von zwei Funktionen darstellen. Für verschränkte Wellenfunktionen gilt das nicht.

Man beachte, dass der Begriff der Verschränktheit nur sinnvoll ist, wenn für einen Vektorraum eine Zerlegung als ein Tensorprodukt definiert ist. Ohne die Angabe einer solchen Produktdarstellung ist der Begriff der Verschränkung nicht definiert. Die Angabe einer solchen Produktdarstellung gibt an, in welcher Form das Gesamtsystem als aus zwei Teilsystemen bestehend gedacht werden kann. Bei zwei Teilchen ist eine solche Zerlegung physikalisch naheliegend (jedes Teilchen definiert sein Teilsystem). Es kann jedoch auch eine Verschränkung beispielsweise zwischen den Polarisationsfreiheitsgraden eines Photons und seinen räumlichen Eigenschaften bestehen. So hat ein einzelnes Photon, das durch einen Doppelspalt tritt, wobei hinter dem rechten Spalt ein +–Polarisationsfilter und hinter dem linken Spalt ein ––Polarisationsfilter steht, eine Verschränkung zwischen seiner Polarisierung und dem Spalt, durch den es getreten ist. Der Zustand

$$|\gamma\rangle = \frac{1}{\sqrt{2}}(|+\rangle \otimes |r\rangle + |-\rangle \otimes |l\rangle) \quad (17)$$

---

<sup>3</sup>Man beachte den Unterschied zur direkten Summe von Vektorräumen bzw. Vektoren:

$$(\mathbf{x}_1 + \mathbf{x}_2) \oplus (\mathbf{y}_1 + \mathbf{y}_2) = \mathbf{x}_1 \oplus \mathbf{y}_1 + \mathbf{x}_2 \oplus \mathbf{y}_2.$$

und  $\mathbf{0} \oplus \mathbf{y}$  sowie  $\mathbf{x} \oplus \mathbf{0}$  sind von null verschiedene Vektoren. Statt  $\mathbf{x} \oplus \mathbf{y}$  schreibt man hier oft einfacher  $(\mathbf{x}, \mathbf{y})$ .

beschreibt ein solches Photon. Die +-Polarisation ist korreliert mit dem rechten Spalt und die --Polarisation mit dem linken. Dieser Zustand ist im Sinne der obigen Definition verschränkt. Die Darstellung des Zustands einzelner Photonen durch solche Verschränkungen zwischen ihrem Polarisationsfreiheitsgrad und ihrer räumlichen Eigenschaft ist ganz sinnvoll, beispielsweise bei der Beschreibung von Experimenten zum Quantenradierer. Meistens hat man es aber mit Verschränkungen zwischen den Polarisationszuständen (bzw. allgemeiner den zwei Zuständen) eines Systems aus zwei Teilchen zu tun.

Die physikalische Bedeutung der Verschränkung liegt unter anderem an folgender Eigenschaft verschränkter Zustände. Ist ein Zustand separierbar, so separiert auch das Absolutquadrat der Komponenten des Vektors, durch den der Zustand beschrieben wird (vgl. Gl. 14). Diese Absolutquadrate entsprechen aber nach der Born'schen Regel bei Einzelsystemen den Wahrscheinlichkeiten, dass zwei Ereignisse gleichzeitig auftreten (bei Funktionenräumen handelt es sich um Wahrscheinlichkeitsdichten, die zur Berechnung einer Wahrscheinlichkeit über ein Volumen zu integrieren sind). Beispielsweise ist in Gl. 14  $w_{ij} = |x_i|^2|y_j|^2$  die Wahrscheinlichkeit für das gleichzeitige Auftreten der beiden Ereignisse  $i$  und  $j$ . Faktorisierende Wahrscheinlichkeiten für zwei Ereignisse bedeutet aber, dass diese Ereignisse unkorreliert sind. Eine Information über das Vorliegen eines der beiden Ereignisse ändert nichts an der Wahrscheinlichkeit für das Auftreten des zweiten Ereignisses.

Bei verschränkten Zuständen ist dies anders: Erhält man eine Information über das Vorliegen eines Ereignisses beispielsweise an einem Teilsystem, kann man neue Vorhersagen über das Eintreffen eines anderen Ereignisses an dem anderen Teilsystem treffen. Im Extremfall liegt das Ereignis an dem anderen Teilsystem sogar eindeutig fest. Dies ist dann besonders erstaunlich, wenn die beiden Teilsysteme sehr weit voneinander entfernt sind. Weshalb sollte eine Messung an einem Teilsystem auf der Erde eine instantane Veränderung des anderen Teilsystems im Andromeda-Nebel zur Folge haben? Diese Eigenschaft verschränkter Zustände hängt eng mit der Problematik der Lokalität der Quantentheorie zusammen.

## 4 Bell-Zustände

In dem Tensorproduktraum von zwei Zwei-Zustandssystemen gibt es eine ausgezeichnete Basis, die man erhält, wenn man die Tensorprodukte der Basisvektoren in den einzelnen Hilbert-Räumen bildet:

$$|h\rangle \otimes |h\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad |h\rangle \otimes |v\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad |v\rangle \otimes |h\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad |v\rangle \otimes |v\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \quad (18)$$

Diese Basisvektoren sind alle separabel. Zur Beschreibung von verschränkten Zuständen bietet es sich oft an, eine Basis zu wählen, bei der alle Basisvektoren maximal verschränkt sind.<sup>4</sup> Die Orthonormalbasis aus den folgenden vier Zuständen hat diese Eigenschaft. Diese Zustände bezeichnet man als

---

<sup>4</sup>Es gibt eine technische Definition von maximal verschränkt: Wenn die Spur über die Projektionsmatrix des Zustands über einen der Teilräume die Identitätsmatrix in dem verbliebenen Teilraum ergibt. Physikalischer ausgedrückt: Ein Zustand heißt maximal verschränkt, wenn der Erwartungswert für beliebige Messungen an einem Teilsystem durch eine Dichtematrix proportional zur Identitätsmatrix - also maximale Unkenntnis - beschrieben werden muss. Dies wird im Folgenden aber nicht benötigt.

Bell-Zustände:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|h\rangle \otimes |h\rangle + |v\rangle \otimes |v\rangle) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad (19)$$

$$|\Phi^-\rangle = \frac{1}{\sqrt{2}}(|h\rangle \otimes |h\rangle - |v\rangle \otimes |v\rangle) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix} \quad (20)$$

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|h\rangle \otimes |v\rangle + |v\rangle \otimes |h\rangle) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \quad (21)$$

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|h\rangle \otimes |v\rangle - |v\rangle \otimes |h\rangle) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}. \quad (22)$$

Den letzten Zustand,  $|\Psi^-\rangle$ , bezeichnet man auch oft als EPR-Zustand (nach Einstein-Podolsky-Rosen). Er hat einige besondere Eigenschaften, insbesondere liegt die Antikorrelation bezüglich jeder Basis vor, d.h. es gilt:

$$|\text{EPR}\rangle = \frac{1}{\sqrt{2}}(|h\rangle \otimes |v\rangle - |v\rangle \otimes |h\rangle) = \frac{1}{\sqrt{2}}(|+\rangle \otimes |-\rangle - |-\rangle \otimes |+\rangle) = \frac{1}{\sqrt{2}}(|R\rangle \otimes |L\rangle - |L\rangle \otimes |R\rangle). \quad (23)$$

Es handelt sich bei diesem Bell-Zustand um den Eigenzustand zum Gesamtdrehimpuls 0 (Photonen haben einen Spin 1, und zwei Photonen können sich zum Gesamtspin 0 oder Gesamtspin 2 verbinden). Dieser Zustand ist invariant unter Drehungen und daher in allen Basissystemen formgleich. Für den Zustand  $|\Phi^+\rangle$  gilt diese Eigenschaft, sofern man sich auf lineare Polarisationen beschränkt.

## Literatur

- [1] Aspect, A., Grangier, Ph., Roger, G.; Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities; Phys. Rev. Lett. 49 (1982) 91–94.  
Aspect, A., Dalibard, J., Roger, G.; Experimental Test of Bell's Inequalities Using Time-Varying Analyzers; Phys. Rev. Lett. 49 (1982) 1804–1807.
- [2] Bell, J.; *Against „Measurement“*; in *62 Years of Uncertainty*, Erice, 5–14 August 1989; auch in *Physics World* 8 (1990) 33–40. Abgedruckt in [3].
- [3] Bell, J.S.; *Speakable and Unspeakable in Quantum Physics*, 2. edition, Cambridge University Press (2004).
- [4] Filk, T; *Quantenmechanik (nicht nur) für Lehramtsstudierende*; Springer-Verlag, 2019.

- [5] Schrödinger, E.; *Die gegenwärtige Situation in der Quantenmechanik*; Die Naturwissenschaften 23 (1935) 807–812, 823–828, 844–849.

# Interferometer

Thomas Filk; Universität Freiburg

Interferometer spielen überall in der Optik eine wichtige Rolle. Sie sind eines der zentralen Instrumente der Experimentalphysik für optische und quantenoptische Versuche. Verwendet man Laserlicht, zeigen Interferometer unter anderem die Wellennatur des Lichts. Mit Einzelphotonen lassen sich jedoch vollkommen neuartige und interessante Effekte erzielen.

Zwei der bekanntesten Interferometer sind das Mach-Zehnder-Interferometer und das Michelson-Interferometer. Beide beruhen auf einem ähnlichen Prinzip: Ein Strahl wird mit einem Strahlteiler geteilt, über ein Spiegelsystem auf einen zweiten Strahlteiler gelenkt und dort wieder zusammengebracht, wobei dann die Interferenzmuster entstehen. Beim Michelson-Interferometer gibt es nur einen Strahlteiler, der sowohl die Aufspaltung als auch die spätere Zusammenführung übernimmt, d.h. das Spiegelsystem reflektiert die aufgeteilten Strahlen wieder zurück auf den ersten Strahlteiler, hinter dem es dann zur Interferenz kommen kann. Beim Mach-Zehnder-Interferometer lenken die Spiegel die beiden Strahlen auf einen zweiten Strahlteiler, hinter dem dann eventuell Interferenzen beobachtet werden können.

Beide Interferometer lassen sich sowohl mit gewöhnlichen Strahlteilern als auch mit polarisationsabhängigen Strahlteilern betreiben. Allerdings ist dies beim Michelson-Interferometer etwas problematischer, da dort die vorgegebene Orientierung die Strahlen in die Lichtquelle lenken würde, wohingegen man einen Ausgang möchte, der senkrecht dazu steht. Das Michelson-Interferometer wird auch beim Nachweis von Gravitationswellen verwendet.

Viele Experimente lassen sich am Mach-Zehnder-Interferometer leichter beschreiben, aber am Michelson-Interferometer leichter durchführen. Daher sollen hier beide Interferometer besprochen werden. In gewissen Hinsicht sind diese Interferometer Erweiterungen des Doppelspalteperiments für kohärentes Licht: Der Strahl bzw. der Zustand eines einzelnen Photons wird aufgeteilt und die beiden Anteile breiten sich entlang unterschiedlicher Wege aus bevor sie wieder zusammengeführt und zur Interferenz gebracht werden. Allerdings kann man an diesen Interferometern die beiden Teile eines Strahls bzw. eines Photonzustands im Prinzip beliebig weit auseinander bringen, was den Effekt erhöht. Außerdem kann man optische Geräte oder Hindernisse in den Weg einer der Teilstrahlen bringen. In der Schule werden viele der (Gedanken-)Experimente mit Mach-Zender- bzw. Michelson-Interferometern unter dem Gesichtspunkt „Ist Welcher-Weg-Information bekannt?“ behandelt, d.h. die Aufgabe der Schüler\*innen besteht darin zu entscheiden, ob man je nach experimentellem Aufbau im Prinzip die Information darüber, welchen Weg ein Photon genommen hat, gewinnen kann oder nicht. In Abhängigkeit davon kann man dann entscheiden, ob ein Interferenzmuster beobachtet werden kann oder nicht.

Beide Interferometer werden für den Fall normaler (d.h., nicht polarisationsabhängiger) Strahlteilern beschrieben. Ein anderer Abschnitt widmet sich dem Mach-Zehnder-Interferometer mit polarisationsabhängigen Strahlteilern. Wir beginnen allerdings mit einer kurzen Beschreibung der mathematischen Darstellung eines Strahlteilers.

## 1 Strahlteiler

Strahlteiler sind die wesentlichen Elemente eines Mach-Zehnder- bzw. Michelson-Interferometers. Es gibt polarisationsabhängige Strahlteiler, bei denen ein einfallender Lichtstrahl in zwei Teilstrahlen mit zueinander orthogonalen Polarisationszuständen aufgespalten wird. Häufiger verwendet man jedoch

nicht polarisationsabhängige Strahlteiler, d.h., der einfallende Strahl wird kohärent in zwei gleichartig polarisierte Teilstrahlen aufgeteilt.

Physikalisch besteht ein Strahlteiler meist aus einer beschichteten Grenzfläche. Die Beschichtung ist gerade so gewählt, dass die Hälfte der Lichtintensität durchgelassen und die andere Hälfte abgelenkt wird. Es gibt auch Strahlteiler mit einer asymmetrischen Aufspaltung, wir beschränken uns hier aber auf die Halbe-Halbe-Strahlteiler.

Wir können die Wirkung eines Strahlteilers durch eine unitäre Matrix beschreiben (unitär, weil zumindest in der Theorie keine Intensität verloren geht, d.h., alle Photonen, die in den Strahlteiler eindringen, kommen auch wieder heraus). Diese Matrix ergibt sich auch direkt aus der Überlegung, wie zwei einfallende Strahlen am Strahlteiler aufgeteilt werden. Diese beiden Vorschriften sind:

$$|1\rangle_i \rightarrow \frac{1}{\sqrt{2}}(|1\rangle_f + i|2\rangle_f) \quad |2\rangle_i \rightarrow \frac{1}{\sqrt{2}}(|2\rangle_f + i|1\rangle_f) \quad (1)$$

Hierbei wurde die in der Streutheorie übliche Konvention verwendet, die einlaufenden ('initial') Zustände durch  $i$  und die auslaufenden ('final') Zustände durch  $f$  zu kennzeichnen. Die (Streu-)Matrix zum Strahlteiler ist dann durch  $S_{ij} = f\langle j|i\rangle_i$  gegeben, also durch die Wahrscheinlichkeitsamplituden zu dem Übergang von einem Anfangszustand  $i$  zu einem Ausgangszustand  $j$  (siehe auch Abb. 1). Die Faktoren  $i$  in Gl. 1 beschreiben eine relative Phasenverschiebung um  $90^\circ$ . Diese Phasenverschiebung tritt immer auf, wenn eine Reflexion an einer Grenzschicht stattfindet, bei der der Winkel zwischen einfallender und ausfallender Welle gerade  $90^\circ$  beträgt. Der Faktor  $1/\sqrt{2}$  ergibt sich daraus, dass das Quadrat dieses Faktors nach der Born'schen Regel die Wahrscheinlichkeit angibt, das Photon in dem jeweiligen Sektor zu messen (hier jeweils  $1/2$ ). Für Laserlicht bedeutet er einfach, dass jeder Teilstrahl die Hälfte der Eingangsintensität besitzt.

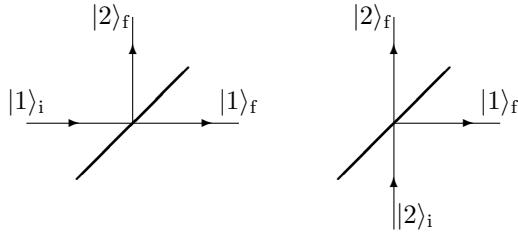


Abbildung 1: Die beiden Strahlengänge in einem Strahlteiler in Abhängigkeit vom Anfangszustand. Ordnet man jeder Reflexion einen Faktor  $i$  zu, erhält man die Matrix in Gl. 2.

Damit erhalten wir die folgende unitäre Matrix, die das Verhalten von Photonen bzw. Licht an einem Halbe-Halbe-Strahlteiler beschreibt:

$$S = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}. \quad (2)$$

Man kann noch relative optische Weglängenänderungen (Phasen) vor- bzw. nachschalten. Diese werden durch eine Diagonalmatrix mit reinen Phasen beschrieben. Auf diese Weise erhalten wir eine allgemeine Darstellung für  $S$  in der Form:

$$S(\alpha, \beta) = \frac{1}{\sqrt{2}} \begin{pmatrix} e^{i\alpha} & 0 \\ 0 & e^{-i\alpha} \end{pmatrix} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} e^{i\beta} & 0 \\ 0 & e^{-i\beta} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} e^{i(\alpha+\beta)} & ie^{i(\beta-\alpha)} \\ ie^{i(\alpha-\beta)} & e^{-i(\alpha+\beta)} \end{pmatrix} \quad (3)$$

Durch die spezielle Wahl  $\alpha = -\beta = \pi/4$  erhalten wir eine andere Darstellung für einen Strahlteiler, die ebenfalls häufig verwendet wird:

$$S' = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}. \quad (4)$$

Das Minuszeichen in der Matrix ist wesentlich, da es sich andernfalls nicht um eine unitäre Matrix handeln würde. Dieses Minuszeichen beschreibt letztendlich die relative Phasenverschiebung von  $180^\circ$  zwischen den beiden reflektierten Strahlen. Anders ausgedrückt: Die Summe der relativen Phasen der reflektierten Strahlen ist  $180^\circ$ , die Summe der relativen Phasen der durchgelassenen Strahlen ist  $0^\circ$ . Dieses Minuszeichen ist aber gerade Schüler\*innen kaum einfach zu vermitteln, wohingegen die obige Erklärung einsichtiger ist und auf dasselbe Ergebnis führt.

## 2 Das Mach-Zehnder-Interferometer

Beim Mach-Zehnder-Interferometer trifft ein Lichtstrahl zunächst auf einen Strahlteiler. Ein Teil des Lichtstrahls wird abgelenkt, der andere Teil durchgelassen. Über zwei Spiegel, an denen die Lichtstrahlen um jeweils  $90^\circ$  reflektiert werden, treffen beide Strahlen von verschiedenen Seiten auf einen zweiten Strahlteiler, wo sie wieder zusammengeführt werden. Hinter dem zweiten Strahlteiler kann man die Intensität des Lichts an zwei möglichen Ausgängen messen.

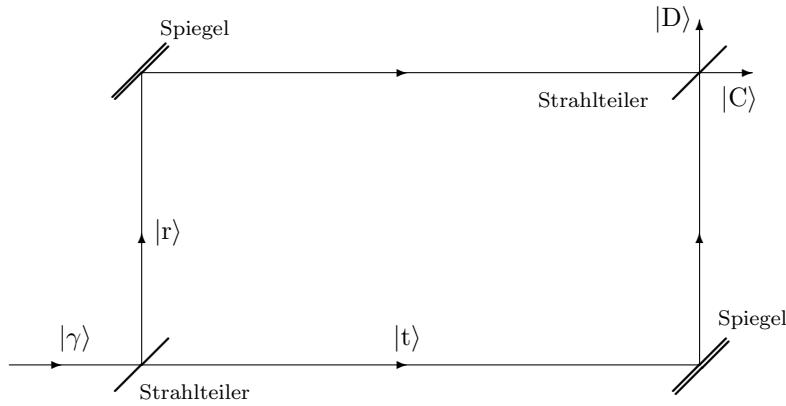


Abbildung 2: Mach-Zehnder-Interferometer. Der Strahl trifft auf einen ersten Strahlteiler. Die beiden Teilstrahlen  $r$  und  $t$  werden von Spiegeln auf einen zweiten Strahlteiler gelenkt, hinter dem es zu konstruktiver bzw. destruktiver Interferenz kommen kann.

Die Interferenz am Ausgang hinter dem zweiten Strahlteiler ergibt sich aus unterschiedlichen optischen Weglängen der zwei Strahlen, die dort zusammentreffen. Meist findet man folgende, sehr vereinfachte Erklärung: Am Ausgang  $D$  treffen zwei Strahlen aufeinander, von denen einer einmal um  $90^\circ$  reflektiert wurde und der andere dreimal. Da jede Reflektion um  $90^\circ$  eine Phasenverschiebung von  $90^\circ$  zur Folge hat ergibt sich zwischen den beiden Strahlen eine relative Phasenverschiebung von  $180^\circ$  und somit destruktive Interferenz. Bei Ausgang  $C$  treffen zwei Strahlen aufeinander, die jeweils einmal reflektiert wurden und somit in Phase sind; es kommt also zu konstruktiver Interferenz.

Im mathematischen Formalismus für Einzelphotonen können wir dies folgendermaßen beschreiben. Ein einfallendes Photon wird durch den Vektor  $|\gamma\rangle$  beschrieben. Am ersten Strahlteiler (ST) wird dieser Zustand in eine Superposition von zwei Zuständen zerlegt: einer dieser Zustände ist der Zustand  $|t\rangle$  (für transmittiert), der andere ist der Zustand des reflektierten Anteils  $|r\rangle$ . Der reflektierte Anteil erhält noch eine Phase von  $90^\circ$ , was durch einen Faktor  $i$  ausgedrückt werden kann. Hinter dem ersten Strahlteiler erhalten wir also folgenden Zustand:

$$|\gamma\rangle \longrightarrow \frac{1}{\sqrt{2}} (|t\rangle + i|r\rangle) . \quad (5)$$

Der Faktor  $1/\sqrt{2}$  ergibt sich daraus, dass das Quadrat dieses Faktors nach der Born'schen Regel die Wahrscheinlichkeit angibt, das Photon in dem jeweiligen Sektor zu messen. Für Laserlicht bedeutet er einfach, dass jeder Teilstrahl die Hälfte der Eingangsintensität besitzt.

Beide Anteile des Strahls werden nun an einem Spiegel reflektiert, erhalten somit einen Faktor  $i$ , der für beide Teile gleich ist. Die beiden Anteile treffen anschließend auf den zweiten Strahlteiler, wobei jeder der beiden Anteile entweder in die Richtung von  $D$  oder von  $C$  gelenkt werden kann. Dies drücken wir durch die Zustände  $|D\rangle$  bzw.  $|C\rangle$  aus, und je nachdem, ob eine Reflektion stattgefunden hat oder nicht, erhalten wir einen weiteren Faktor  $i$ .

$$|\gamma\rangle \xrightarrow{1. ST} \frac{1}{\sqrt{2}}(|t\rangle + i|r\rangle) \xrightarrow{\text{Spiegel}} \frac{i}{\sqrt{2}}(|t\rangle + i|r\rangle) \quad (6)$$

$$\xrightarrow{2. ST} \frac{i}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}(|D\rangle + i|C\rangle) + \frac{i}{\sqrt{2}}(i|D\rangle + |C\rangle)\right) \quad (7)$$

$$= \frac{i}{2}|D\rangle + \frac{i^2}{2}|C\rangle + \frac{i^3}{2}|D\rangle + \frac{i^2}{2}|C\rangle \quad (8)$$

$$= -|C\rangle \quad (9)$$

Das übliche Interferenzmuster im Sinne von „alternierenden Helligkeitsstreifen“ erhält man hier, wenn man die optische Weglänge eines der beiden Teilstrahlen beispielsweise durch eine Verschiebung des reflektierenden Spiegels ändert. Man erhält so alternierend mal den Ausgang  $C$  und mal den Ausgang  $D$  als den Ausgang maximaler Helligkeit, wobei der jeweils andere Ausgang dunkel ist.

In der Praxis, beispielsweise bei der Verwendung von sichtbarem Licht, lässt sich die optische Weglänge kaum so ausmessen, dass man die relativen Phasenunterschiede einzig auf die Anzahl der Reflexionen zurückführen kann. Man wird die optische Weglänge eines der Teilstrahlen durch Verschieben der optischen Elemente so anpassen, dass man beispielsweise hinter einem der Ausgänge kein Licht (destruktive Interferenz) und entsprechend hinter dem anderen Ausgang volle Helligkeit (konstruktive Interferenz) findet.

### 3 Feynman's Summation über alle Wege

Richard Feynman hat eine sehr eingängige Darstellung für die Bestimmung einer Amplitude  $\langle b|a\rangle$  gefunden. Formal kann man schreiben:

$$\langle b|a\rangle = \sum_{\text{Wege } a \rightarrow b} A(\text{Weg}), \quad (10)$$

wobei  $A(\text{Weg})$  eine komplexe Amplitude ist, die sich als Produkt einzelner Beiträge entlang eines Weges schreiben lässt. Summiert wird über alle „Wege“ (Möglichkeiten), unter Einhaltung der physikalischen Gesetze von dem Zustand  $a$  zu dem Zustand  $b$  zu gelangen. Diese Darstellung wollen wir uns am Mach-Zehnder-Interferometer anschauen.

Der Ausgangszustand ist der Zustand  $|\gamma\rangle$  von einem Photon, das am ersten Strahlteiler in das Interferometer trifft. Die möglichen Endzustände sind  $|D\rangle$  und  $|C\rangle$ , bei denen ein Photon in Detektor  $D$  bzw.  $C$  nachgewiesen wird. Für die beiden Amplituden  $\langle D|\gamma\rangle$  und  $\langle C|\gamma\rangle$  gibt es jeweils zwei Wege:

$$\begin{aligned} \langle D|\gamma\rangle &= |\gamma\rangle \rightarrow ST_1 \rightarrow |r\rangle \rightarrow \text{Spiegel} \rightarrow |r\rangle \rightarrow ST_2 \rightarrow |D\rangle \\ &\quad + |\gamma\rangle \rightarrow ST_1 \rightarrow |t\rangle \rightarrow \text{Spiegel} \rightarrow |t\rangle \rightarrow ST_2 \rightarrow |D\rangle \end{aligned} \quad (11)$$

$$\begin{aligned} \langle C|\gamma\rangle &= |\gamma\rangle \rightarrow ST_1 \rightarrow |r\rangle \rightarrow \text{Spiegel} \rightarrow |r\rangle \rightarrow ST_2 \rightarrow |C\rangle \\ &\quad + |\gamma\rangle \rightarrow ST_1 \rightarrow |t\rangle \rightarrow \text{Spiegel} \rightarrow |t\rangle \rightarrow ST_2 \rightarrow |C\rangle. \end{aligned} \quad (12)$$

Zur Berechnung der Phase verwenden wir folgende Regeln:

- Propagation entlang einer geraden Strecke:  $A = 1$ .

Dies ist nicht ganz korrekt, eigentlich müsste man hier als reine Phase die optische Weglänge

nehmen. Da wir nur an relativen Phasenunterschieden interessiert sind können wir diesen Beitrag unberücksichtigt lassen.

- Für jede Reflektion um einen Winkel  $\alpha$  erhalten wir eine Phase  $A = e^{i\alpha}$ . Für Reflektionen um  $90^\circ$  ist dies ein Faktor  $i$ .
- Für jede Verzweigung (z.B. in einem Strahlteiler) erhalten wir einen Faktor  $A = 1/\sqrt{2}$  (das Quadrat dieses Faktors ist gleich der Wahrscheinlichkeit, in eine der beiden Richtungen abgelenkt zu werden).

Damit können wir die Phasen für die obigen Wege ausrechnen:

$$\langle D|\gamma \rangle = \left(\frac{i}{\sqrt{2}}\right)(i)\left(\frac{i}{\sqrt{2}}\right) + \left(\frac{1}{\sqrt{2}}\right)(i)\left(\frac{1}{\sqrt{2}}\right) = -\frac{i}{2} + \frac{i}{2} = 0 \quad (13)$$

$$\langle C|\gamma \rangle = \left(\frac{i}{\sqrt{2}}\right)(i)\left(\frac{1}{\sqrt{2}}\right) + \left(\frac{1}{\sqrt{2}}\right)(i)\left(\frac{i}{\sqrt{2}}\right) = -\frac{1}{2} - \frac{1}{2} = -1. \quad (14)$$

Wir finden diese Faktoren auch in den Gleichungen 7 und 8 wieder.

Feynmans Darstellung einer Amplitude als Summe über Beiträge von Wegen ist oft ganz hilfreich, um sich die möglichen Wege vor Augen zu halten. In einer Kontinuumstheorie wird aus der Summation über Wege ein sogenanntes Funktionalintegral. Die Idee hinter dieser Darstellung beruht auf folgender Überlegung: Ein physikalischer Prozess wird in der Quantentheorie durch den sogenannten Zeitentwicklungsoperator  $\exp(\frac{i}{\hbar} H t)$  (oder, wie beispielsweise beim Mach-Zehnder-Interferometer, durch einen äquivalenten unitären Operator) beschrieben. Diesen unitären Operator kann man als Produkt von Operatoren zu einzelnen Abschnitten der Wege (oder sehr kurzen Zeitabschnitten) schreiben. Die Matrixmultiplikation entspricht aber einer „Summation über alle Wege im Indexraum“. Um das zu verdeutlichen, betrachten wir der Einfachheit halber das Produkt von drei Matrizen  $A$ ,  $B$  und  $C$ , wobei wir die Bra-Ket-Notation von Dirac verwenden:

$$\langle m|(CBA)|n\rangle = \sum_{kl} \langle m|C|l\rangle \langle l|B|k\rangle \langle k|A|n\rangle \quad (15)$$

Für einen festen Wert von  $n$  und  $m$  (Anfangs- und Endzustand) erhalten wir somit eine Darstellung des gesamten Matrixelements als Summe über alle Wege (summiert wird über alle  $k$  und  $l$ )  $n \rightarrow k \rightarrow l \rightarrow m$  und zu jedem Teilprozess eines dieser Wege erhalten wir als „Amplitude“ das entsprechende Element der zugehörigen unitären Matrix. Die Wahrscheinlichkeit für einen Übergang  $n \rightarrow m$  bei dem Prozess  $CBA$  ist dann das Quadrat dieser Amplitude.

Beim Mach-Zehnder-Interferometer multiplizieren wir drei Matrizen: Die erste Matrix beschreibt den ersten Strahlteiler, die zweite Matrix die Reflektion an einem Spiegel (dies ergibt im Wesentlichen nur einen Faktor  $i$ ) und die dritte Matrix den zweiten Strahlteiler.

## 4 Das Michelson-Interferometer

Beim Michelson-Interferometer trifft ein einfallender Strahl bzw. der Zustand eines einfallenden Photons auf einen Strahlteiler und wird dort in zwei Anteile aufgespalten, die sich entlang unterschiedlicher Wege ausbreiten. Beide Teilstrahlen werden an Spiegel totalreflektiert und laufen wieder zurück. Sie treffen schließlich wieder auf den Strahlteiler, wo sie zusammengeführt werden und interferieren können (siehe Abb. 1).

Beim Michelson-Interferometer erhält man zunächst nur einen Ausgang, hinter dem man die Interferenz nachweisen kann. Das ist insofern kein Informationsverlust, da der zweite Ausgang ohnehin nur die am ersten Ausgang fehlende Intensität (zur einlaufenden Gesamtintensität) aufzeigt.

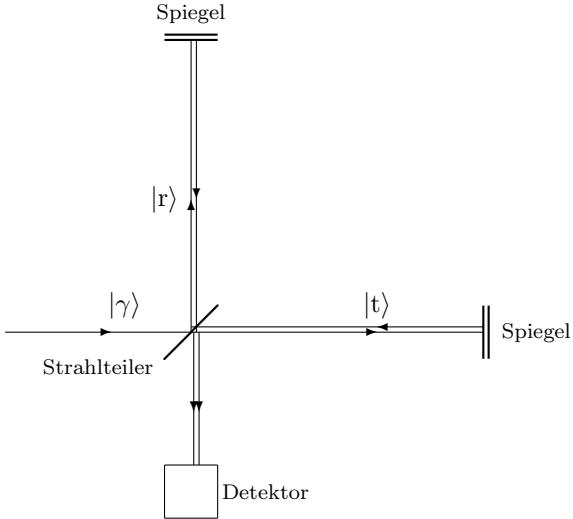


Abbildung 3: Michelson-Interferometer. Nachdem der Strahl von einem Strahlteiler geteilt wurde, treffen die beiden Strahlteile auf Spiegel, die sie auf den Strahlteiler zurücklenken. Hinter dem Strahlteiler kann man mit einem Detektor die Lichtintensität messen. Durch winzige Variationen in den optischen Weglängen kann man Interferenzen beobachten.

Der zweite Ausgang zeigt wieder zurück zum einfallenden Licht und sollte daher nicht mit Detektoren zugebaut werden. Es gibt aber unterschiedliche Varianten dieses Aufbaus, bei denen auch hinter dem zweiten Ausgang Detektoren stehen können.

## 5 Das „Knallerexperiment“

Das sogenannte „Knallerexperiment“ geht auf eine Arbeit von Avshalom Elitzur und Lev Vaidman aus dem Jahr 1993 mit dem Titel „Quantum mechanical interaction-free measurements“ zurück [1]. Sie beschreiben darin die Möglichkeit, eine Superbombe auf ihre Funktionsfähigkeit zu testen, ohne diese Bombe zur Explosion bringen zu müssen. Für schulische Zwecke wurde aus der Superbombe ein „Knaller“.

Hier ist allerdings zu erwähnen, dass die Ideen von Elitzur und Vaidman nicht neu waren, sie haben sie lediglich in einen „werbewirksamen“ Rahmen gepackt. Schon 1960 hat Mauritius Renninger in einem Artikel „Messung ohne Störung des Messobjekts“ auf diese Möglichkeit in der Quantentheorie aufmerksam gemacht [2]. Und schon 1934 schreibt Schrödinger in einem Artikel: *Übrigens ist es auch an sich nichts weniger als einleuchtend, daß das Meßinstrument ohne Wechselwirkung keine Aussage über das Objekt machen kann. Eine Zielscheibe, die nicht getroffen wird, schließt zumindest gewisse Flugbahnen des Projektils aus. Und wenn die Scheibe den Schützen als Hohlkugel umgibt, die bloß ein kleines Loch hat, so gibt sie, ungetroffen, sehr genauen Aufschluß über die Flugbahn* [3]. Damit ist die wesentliche Idee des Artikels von Elitzur und Vaidman schon vorweggenommen.

Elitzur und Vaidman beschreiben folgende Situation: Gegeben ist ein großes Lager von Superbomben, die einen sehr empfindlichen Auslöser haben. Sobald ein Photon auf diesen Auslöser trifft und seine Energie überträgt, explodiert die Bombe. Andererseits hat man festgestellt, dass bei manchen Bomben die Auslöser fehlen und bei diesen Bomben ein Photon, das normalerweise auf einen Auslöser treffen würde, einfach hindurchgeht und an einer Seite wieder austritt. Ein normaler Test, bei dem nachgeschaut wird, ob ein Auslöser vorhanden ist, würde die intakten Bomben explodieren lassen. Man möchte jedoch möglichst viele intakte Bomben behalten können.

Elitzur und Vaidman schlagen nun vor, ein Mach-Zehnder-Interferometer zum Test der Bomben zu verwenden. Die Bomben werden so in den Strahlengang gelegt, dass der Auslöser, sofern vorhanden, ein Hindernis darstellt und, wenn er von einem Photon getroffen wird, die Bombe ausgelöst wird. Bei Bomben, die defekt sind und bei denen kein Auslöser vorhanden ist, ist der Strahlengang

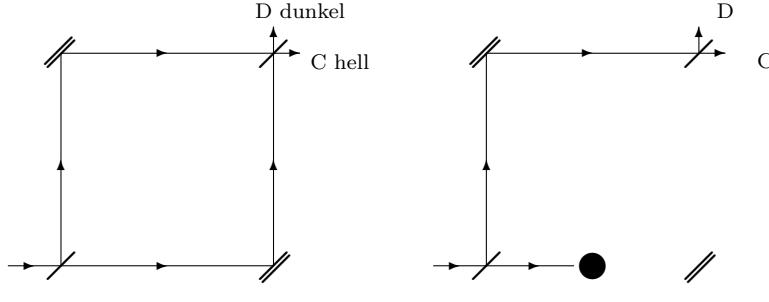


Abbildung 4: Mach-Zehnder Interferometer ohne und mit Hindernis in einem Strahlengang. Ohne Hindernis sollte der Detektor bei D nie ansprechen, mit Hindernis in rund einem Viertel der Fälle.

frei und ein Photon könnte ungehindert durch den fehlenden Auslöser hindurchtreten.

Falls die Bomben defekt sind, handelt es sich bei dem Aufbau um ein normales Mach-Zehnder-Interferometer. Wenn die optischen Weglängen entsprechend eingestellt sind, sollten Photonen letztendlich nur in Detektor C nachgewiesen werden. Man könnte entweder eine ausreichende Anzahl von Photonen durch die Anordnung treten lassen und feststellen, dass immer nur Detektor C anspricht. Man kann aber auch das Photon hinter dem Ausgang C durch entsprechende Spiegel wieder in das Mach-Zehnder-Interferometer leiten und würde feststellen, dass ein Photon „immer im Kreis läuft“. Dies wäre das Zeichen für eine defekte Bombe.

Bei einer intakten Bombe befindet sich ein Hindernis in einem Strahlengang. In rund der Hälfte der Fälle wird ein Photon auf die Bombe treffen und diese auslösen. In der anderen Hälfte der Fälle kann man sich vorstellen, dass das Photon den anderen Strahlengang nimmt und dort hinter dem zweiten Strahlteiler entweder in Detektor C oder Detektor D abgelenkt wird. Landet das Photon in Detektor D weiß man, dass ein Hindernis vorhanden ist, und damit ist auch bekannt, dass die Bombe intakt ist. Da das Photon aber im Detektor nachgewiesen wurde, ist es nicht auf das Hindernis getroffen, es hat keine Wechselwirkung mit dem Hindernis stattgefunden und die Bombe bleibt intakt. Dies ist bei einer intakten Bombe in rund einem Viertel der Fälle der Fall. Schließlich kann das Photon auch in Detektor C landen (ebenfalls ein Viertel der Fälle); in diesem Fall erhält man keine Information über die Bombe. Ein solches Photon kann nochmals in das Mach-Zehnder-Interferometer gelenkt werden, wieder wird in der Hälfte der Fälle die Bombe ausgelöst und in jeweils einem Viertel der Fälle landet das Photon in Detektor D (damit ist bekannt, dass die Bombe intakt ist) oder in Detektor C (keine Information). Auf diese Weise kann man letztendlich rund ein Drittel aller intakten Bomben als intakt erkennen und retten, wohingegen rund zwei Drittel aller intakten Bomben explodieren.

Wir können auch in diesem Fall zur Bestimmung der Wahrscheinlichkeiten die Feynman'sche Summe über alle Möglichkeiten nutzen. Bei einer intakten Bombe gibt es neben den Möglichkeiten, dass ein Photon in Detektor D oder C landet noch die Möglichkeit, auf die Bombe zu treffen und diese auszulösen. Dies geschieht mit Wahrscheinlichkeit 1 für einen Weg, der das Photon auf die Bombe leitet. Zu den Amplituden für die Detektoren trägt nun jeweils nur ein Weg bei:

$$\begin{aligned} \langle D|\gamma\rangle &= |\gamma\rangle \rightarrow ST \rightarrow |r\rangle \rightarrow \text{Spiegel} \rightarrow |r\rangle \rightarrow ST \rightarrow |D\rangle \\ &= \left(\frac{i}{\sqrt{2}}\right)(i)\left(\frac{i}{\sqrt{2}}\right) = -\frac{i}{2} \end{aligned} \quad (16)$$

$$\begin{aligned} \langle C|\gamma\rangle &= |\gamma\rangle \rightarrow ST \rightarrow |r\rangle \rightarrow \text{Spiegel} \rightarrow |r\rangle \rightarrow ST \rightarrow |C\rangle \\ &= \left(\frac{i}{\sqrt{2}}\right)(i)\left(\frac{1}{\sqrt{2}}\right) = -\frac{1}{2} \end{aligned} \quad (17)$$

$$\begin{aligned} \langle \text{Bombe explodiert}|\gamma\rangle &= |\gamma\rangle \rightarrow ST \rightarrow |t\rangle \rightarrow \text{Bombe} \\ &= \left(\frac{1}{\sqrt{2}}\right). \end{aligned} \quad (18)$$

Das Absolutquadrat dieser Amplituden ergibt die Wahrscheinlichkeiten. Damit erhalten wir für die beiden Wahrscheinlichkeiten, dass ein Photon in einem der Detektoren landet, jeweils  $1/4$ , wohingegen die Wahrscheinlichkeit auf die Bombe zu treffen und diese auszulösen gleich  $1/2$  ist.

Für die experimentelle Realisation verwendet man meist ein Michelson-Interferometer (Abb. 5). Die Bombe kann durch einen Spiegel ersetzt werden, der in den Strahlengang geschoben wird und ein Photon, das eigentlich eine intakte Bombe auslösen würde, auf einen Detektor lenkt.

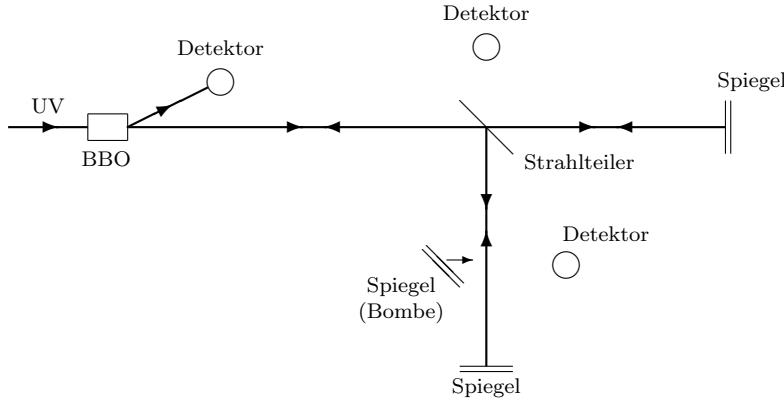


Abbildung 5: Experimentelle Realisierung des „Knallerexperiments“ in einem Michelson-Interferometer. In einem BBO-Kristall werden zwei Photonen erzeugt, von denen eines nachgewiesen wird. Dadurch ist bekannt, dass sich ein zweites Photon in der experimentellen Anordnung befindet. Der bewegliche Spiegel („Bombe“) kann nach rechts in den Strahlengang geschoben werden und lenkt das Photon auf den Detektor. Der obere Detektor (dunkel) sollte nur Ereignisse anzeigen, bei denen ein Hindernis im zweiten Strahlengang vorhanden ist. Der untere Detektor zeigt an, ob das Hindernis getroffen wurde.

## Literatur

- [1] Elitzur, A.C., Vaidman, L.; *Quantum mechanical interaction-free measurements*, Found. of Phys. **23** (1993) 987.
- [2] Renninger, M.; *Messung ohne Störung des Meßobjekts*; Z. Physik 158 (1960) 417.
- [3] Schrödinger, E.; *Über die Unanwendbarkeit der Geometrie im Kleinen*; Die Naturwissenschaften 31 (1934) 518–520.

# BB84 - Quantenkryptographie

Thomas Filk; Universität Freiburg

Quantenkryptographie ist eine der vielen faszinierenden Anwendungen der Quanteninformation. Die Möglichkeit, Nachrichten austauschen zu können, ohne dass diese Nachrichten von dritter Seite abgehört oder entschlüsselt werden können, bietet viele Anwendungen, die - wie immer in solchen Fällen - natürlich auch missbraucht werden können.

Bei den meisten Protokollen zur Quantenkryptographie, unter anderem auch bei dem Protokoll BB84 (benannt nach Charles Bennett und Gilles Brassard, die es 1984 entwickelten [1]), geht es allerdings nur darum, einen idealen Schlüssel - in diesem Fall eine Zufallsfolge von Bits - auszutauschen, der später sowohl für die Verschlüsselung als auch für die Entschlüsselung der Nachricht verwendet werden kann. Wichtig ist, dass man überprüfen kann, ob der Schlüssel außer den beiden Teilnehmern (Sender und Empfänger der Nachricht) tatsächlich niemandem bekannt ist. Genau das leistet das BB84 Protokoll. Die eigentliche Nachricht wird nach klassischen Verfahren ver- und entschlüsselt und auch in klassischer Form verschickt.

Um nicht immer von einer Senderin oder einem Sender *A* und einer Senderin oder einem Sender *B* zu sprechen, haben sich in der Informationstheorie die Bezeichnungen „Alice“ für die Senderin der Nachricht und „Bob“ für den Empfänger der Nachricht etabliert. Den Lauscher bzw. die Lauscherin bezeichnet man meist als „Eve“, abgeleitet von der englischen Bezeichnung „evesdropper“ für Lauscher.

Für die Übertragungssituation des BB84 Protokolls fordert man gewisse Bedingungen, die erfüllt sein sollten, damit das Protokoll sicher ist. So haben sowohl Alice als auch Bob einen abgeschlossenen Bereich, in den kein externer Lauscher eindringen kann. Ein Eingreifen von Eve ist nur während der Übertragung von Daten möglich, solange sich diese Daten in einem offenen Bereich befinden. Außerdem wird vorausgesetzt, dass sich Alice und Bob über einen klassischen Kanal (z.B. Telefon oder Videokanal) austauschen und dabei verifizieren können, dass sie tatsächlich mit der jeweils anderen Person sprechen. Bei einem klassischen Übertragungskanal wird also angenommen, dass sich Eve nicht als Alice oder Bob ausgeben kann. In der Informatik bezeichnet man diese Bedingung als Authentizität. Allerdings kann Eve natürlich einem solchen Gespräch lauschen.

Wir beginnen mit einer kurzen Beschreibung des sogenannten One-Time-Pads, einem Protokoll der klassischen Kryptographie, bei dem einmalig eine Zufallsfolge von Bits sowohl vom Sender zur Verschlüsselung als auch vom Empfänger der Nachricht zur Entschlüsselung verwendet wird. Anschließend wird beschrieben, wie man mit Verfahren der Quantentheorie einen solchen Schlüssel austauschen und gleichzeitig sicherstellen kann, dass außer Alice und Bob niemand den Schlüssel kennt.

## 1 Klassische Kryptographie - das One-Time-Pad

Ein One-Time-Pad ist spezielles Protokoll der klassischen Kryptographie, bei dem eine Nachricht mit Hilfe einer Zufallsfolge von Bits vom Sender verschlüsselt und vom Empfänger entschlüsselt werden kann. Sender und Empfänger müssen diese Zufallsfolge kennen, allerdings sollte niemand sonst Informationen über diese Folge haben. Es gibt auch One-Time-Pads, bei denen dezimale Zahlenfolgen oder Buchstabenfolgen verwendet werden, doch sofern es sich in allen Fällen um wirkliche Zufallsfolgen handelt, sind diese nicht sicherer als eine binäre Bitfolge. Man benötigt allerdings bei einer binären Nachrichtenübertragung mehr Zeichen. Außerdem muss die zu verschlüsselnde Nachricht ebenfalls als

Bitfolge vorliegen, was sich aber immer erreichen lässt. Der Schlüssel des One-Time-Pads sollte den folgenden vier Bedingungen genügen:

1. Die Bitfolge muss mindestens so lang sein wie der zu verschlüsselnde Text.
2. Es muss sich um eine Zufallsfolge von Bits handeln.
3. Die Bitfolge oder Teile von ihr werden kein zweites Mal (weder in der vorliegenden Nachricht noch in anderen Nachrichten) wiederverwendet.
4. Die Bitfolge darf nur dem Sender und dem Empfänger bekannt sein. Auch keine einschränkenden Informationen über die Bitfolge dürfen potenziellen Lauschern bekannt sein.

Unter diesen Bedingungen kann man beweisen, dass eine entsprechend verschlüsselte Nachricht prinzipiell nicht entschlüsselt werden kann. Der Grund ist sehr einfach: Bei der mit einer Zufallsfolge verschlüsselten Nachricht handelt es sich wieder um eine Zufallsfolge, und wenn man alle möglichen Schlüssel ausprobiert, erhält man jede beliebige Bitfolge und somit auch jeden beliebigen Text derselben Länge. Sofern über den Schlüssel nichts bekannt ist, kann man auch keine Information über die ursprüngliche Nachricht gewinnen.

Bedingung 1 und 3 lassen sich im Prinzip sehr leicht erfüllen, sind aber in der Praxis oft ein Problem, da Sender und Empfänger für den Bedarfsfall oft sehr lange Schlüssel austauschen oder vorrätig haben müssen. Bedingung 2 ist etwas schwieriger, da die meisten sogenannten Zufallszahlgeneratoren auf einem deterministischen Algorithmus beruhen und damit für jemanden, der diesen Algorithmus sowie den Anfangszustand kennt, auch reproduzierbar sind. Die eigentliche Problematik ist aber Bedingung 4: Wie kann man sicher sein, dass niemand außer den Teilnehmern die Bitfolge kennt, insbesondere, wenn diese Bitfolge über einen öffentlichen bzw. abhörbaren Kanal ausgetauscht wurde? Wir werden sehen, dass sowohl Bedingung 2 als auch Bedingung 4 im Rahmen der Quantenkryptographie erfüllt werden können.

Ist eine Zufallsbitfolge gegeben, erfolgt die Verschlüsselung durch eine XOR-Operation, also eine Addition modulo 2, mit der zu verschlüsselnden Nachricht. Die Rechenregeln sind:

$$0 + 0 = 1 + 1 = 0 \quad 0 + 1 = 1 + 0 = 1. \quad (1)$$

Die daraus entstandene Bitfolge ist unabhängig von der darin enthaltenen Nachricht nach allen detektierbaren Kriterien ebenfalls wieder eine Zufallsfolge und kann über einen öffentlichen Kanal verschickt werden. Der Empfänger kann diese verschlüsselte Nachricht entziffern, indem er mit demselben Schlüssel nochmals eine XOR Operation durchführt. Da sowohl  $0 + 0$  als auch  $1 + 1$  bezüglich XOR die 0 ergeben, wird durch die insgesamt zweifache XOR Addition des Schlüssels die ursprüngliche Nachricht wieder hergestellt. Dies zeigt folgendes Beispiel, bei dem ein Klartext (abwechselnd dreimal 0 und dreimal 1) mit einer Zufallsfolge zu einer verschlüsselten Nachricht umgewandelt und anschließend mit derselben Zufallsfolge wieder entschlüsselt wird:

Klartext	1 1 1 0 0 0 1 1 1 0 0 0 1 1 1 0 0 0 1 1 1 0 0 0	
Zufallsfolge	0 1 1 0 1 0 0 0 1 0 1 1 1 0 1 1 0 1 0 0 1 0 0 1	
XOR-kodierte Nachricht	1 0 0 0 1 0 1 1 0 0 1 1 0 1 0 1 0 1 1 1 0 0 0 1	(2)
Zufallsfolge	0 1 1 0 1 0 0 0 1 0 1 1 1 0 1 1 0 1 0 0 1 0 0 1	
XOR-Dekodierung=Klartext	1 1 1 0 0 0 1 1 1 0 0 0 1 1 1 0 0 0 1 1 1 0 0 0	

## 2 Das BB84-Protokoll

Wie schon erwähnt geht es bei dem BB84-Protokoll nur um den sicheren Austausch eines Schlüssels, also einer binären Zufallsfolge. Im Prinzip lässt sich das BB84-Protokoll mit jedem QBit, also jedem quantenmechanischen Zwei-Zustandssystem durchführen, allerdings verwendet man meist die linearen Polarisationszustände von Photonen. Für das BB84-Protokoll werden auch keine verschränkten Photonen benötigt. Es gibt allerdings andere Protokolle, die von verschränkten Zuständen Gebrauch machen.

### 2.1 Messung und Präparation der Polarisation von Einzelphotonen

Wichtig für das Verständnis der Quantenkryptographie im Allgemeinen wie auch speziell für das BB84-Protokoll sind folgende Tatsachen:

1. Ein unbekannter Polarisationszustand eines Einzelphotons kann nicht durch Messungen bestimmt werden. Ein polarisationsabhängiger Strahlteiler oder auch ein Polarisationsfilter hat immer eine bestimmte Orientierung. Dadurch wird eine Basis aus zwei orthogonalen Polarisationszuständen ausgezeichnet. Eine „Messung“ findet immer in Bezug auf diese Basis statt und man erhält einen der beiden möglichen Basiszustände als Ergebnis. Ob diese Polarisation auch vorher schon vorlag, ist dabei nicht bekannt.
2. Quantenzustände kann man nicht clonen, d.h., man kann keine Kopien eines Quantenzustands herstellen und dabei das Original behalten. Die Quantenteleportation erlaubt zwar die Erstellung einer Kopie, aber dabei geht das Original verloren, d.h., es gibt immer nur einen Zustand mit den ursprünglichen Eigenschaften.
3. Ist die Polarisationsbasis bekannt, d.h., ist bekannt, bezüglich welcher Orientierung (z.B. eines Polarisationsstrahlteilers) eine Polarisation präpariert wurde, kann diese natürlich auch ausgelernt und kopiert werden.

Das Protokoll für den Schlüsselaustausch besteht aus drei Schritten. In einem weiteren Schritt kann überprüft werden, ob der Schlüssel abgehört wurde. Auf diesen letzten Schritt gehen wir in Abschnitt 2.6 ein, die ersten drei Schritte sind:

1. Alice präpariert zufällig ausgewählte Polarisationszustände von Photonen bezüglich zweier vorab festgelegter Orientierungen (Basisysteme) und verschickt diese an Bob. Diese Polarisationszustände kodieren eine Bitfolge.
2. Bob nimmt Polarisationsmessungen an den Photonen vor, wobei er zwischen den beiden vorab festgelegten möglichen Orientierungen (Basisystemen) zufällig auswählt. Er erhält auf diese Weise ebenfalls eine Bitfolge.
3. Alice und Bob vergleichen über einen klassischen Kanal, welche Basis sie bei den einzelnen Photonen gewählt haben. Die Bitfolge in den übereinstimmenden Fällen ist ihre Zufallsfolge.

Diese Schritte werden im Folgenden eingehender behandelt.

### 2.2 Alice präpariert die Photonen für Bob

Alice (die Senderin) möchte sich mit Bob (dem Empfänger) eine binäre Zufallsfolge teilen, die nur sie beide kennen. Dazu benötigt Alice zunächst eine Quelle von Einzelphotonen, denen sie gezielt

bestimmte Polarisationszustände verleihen kann. Sie wählt dabei zwischen vier möglichen Polarisationszuständen aus, die den Basiszuständen von zwei Orientierungen eines Polarisationsstrahlteilers entsprechen: die Basiszustände bezüglich horizontaler-vertikaler Orientierung und die Basiszustände bezüglich einer  $+45^\circ$ - $-45^\circ$  diagonalen Orientierung (siehe Abb. 1). Wenn im Folgenden von Orientierung die Rede ist, bezieht sich dies immer auf die Wahl eines orthogonalen Basissystems. Für jede Orientierung gibt es dann zwei mögliche Basiszustände bzw. Polarisationszustände.

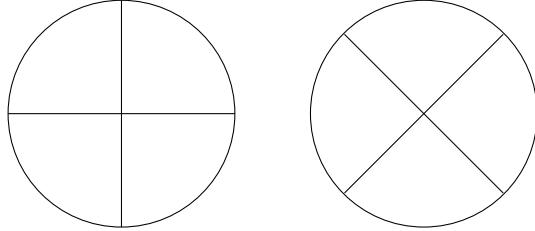


Abbildung 1: Die beiden Orientierungen bzw. Basissysteme für die Präparations- bzw. Messanordnungen beim BB84-Protokoll. (links) horizontal-vertikal ( $h/v$ -Basis), (rechts)  $+45^\circ$ - $-45^\circ$  ( $+/-$ -Basis).

Alice erzeugt nun eine Folge von Einzelphotonen, die jeweils eine zufällige Polarisation bezüglich einer der beiden Polarisationsorientierungen haben, d.h. die zufällig eine der vier Polarisationszustände  $h$ ,  $v$ ,  $+$  oder  $-$  haben. Das kann z.B. folgendermaßen geschehen (es geht hier nur um ein Prinzip, nicht um die in der Praxis verwendete Realisierung): Es gibt nicht-lineare Kristalle (z.B. Bariumborat, oft mit BBO abgekürzt), bei denen ein einfallendes Photon einer bestimmten Energie in zwei Photonen von jeweils der halben Energie (doppelte Wellenlänge) umgewandelt wird. Eines dieser Photonen dient als Signalphoton - es zeigt an, dass ein zweites Photon (das sogenannte Idler-Photon) in diesem Moment in eine bestimmte Richtung emittiert wird. Gewöhnlich hat dieses zweite Photon eine wohldefinierte Polarisation, die Alice in eine der vier genannten Polarisationen drehen kann. Sie kennt also den genauen Polarisationszustand des Photons, das sie an Bob schickt. Alice sollte die Orientierungen, bezüglich der sie die Polarisationen präpariert, zufällig wählen.

Diese Folge von Einzelphotonen schickt Alice an Bob. Sie hat dabei für jedes einzelne dieser Photonen folgende Information, die sie zunächst geheim hält: Sie kennt die Basis, bezüglich der sie die Polarisation der Photonen präpariert hat, und sie kennt den zugehörigen Polarisationszustand dieses Photons. Hierbei verwendet man eine Konvention, die vorher festgelegt wurde, z.B. 0 für den Zustand  $h$  in der  $h/v$ -Basis und 1 für den Zustand  $v$  in der  $h/v$ -Basis, entsprechend 0 für den Zustand  $+$  in der  $+/-$ -Basis und 1 für den Zustand  $-$  in der  $+/-$ -Basis. Alice besitzt also eine Tabelle wie in Tab. 1.

Photon	1	2	3	4	5	6	7	8	9	10	11	12	13
Basis	$+/-$	$+/-$	$h/v$	$+/-$	$h/v$	$h/v$	$h/v$	$+/-$	$+/-$	$h/v$	$+/-$	$+/-$	$+/-$
Zustand	+	-	$h$	+	$h$	$h$	$h$	-	+	$h$	+	+	-
Bit	1	0	0	1	0	0	0	0	1	0	1	1	0

Tabelle 1: Tabelle von Alice. Diese Tabelle enthält die Polarisationszustände der Photonen, die Alice an Bob verschickt.

### 2.3 Bob nimmt an den Photonen von Alice Messungen vor

Bob erhält von Alice die Folge der Photonen und weiß nur, dass jedes einzelne Photon entweder bezüglich der Basis  $h/v$  oder bezüglich der Basis  $+/-$  präpariert wurde. Er nimmt nun an jedem dieser Photonen eine Messung vor, wobei er die Basis dieser Messung, d.h. die Orientierung seines Polarisationsstrahlteilers, ebenfalls zufällig wählt. In ungefähr der Hälfte dieser Messungen wählt

Bob eine Basis, die mit der Präparationsbasis von Alice übereinstimmt. In diesen Fällen stimmt sein Ergebnis mit dem Ergebnis von Alice überein. In allen anderen Fällen ist seine Basis von der Präparationsbasis von Alice verschieden und seine Ergebnisse sind zufällig. Bob erhält dadurch eine ähnliche Tabelle wie vorher Alice (siehe Tab. 2).

Photon	1	<u>2</u>	<u>3</u>	4	5	6	<u>7</u>	8	<u>9</u>	<u>10</u>	11	12	<u>13</u>
Basis	$h/v$	$+/-$	$h/v$	$h/v$	$+/-$	$+/-$	$h/v$	$h/v$	$+/-$	$h/v$	$h/v$	$h/v$	$+/-$
Zustand	$v$	$-$	$h$	$h$	$+$	$-$	$h$	$h$	$+$	$h$	$h$	$v$	$-$
Bit	1	0	0	0	1	0	0	0	1	0	0	1	0

Tabelle 2: Tabelle von Bob. Sie enthält die Polarisationszustände, die Bob an den Photonen von Alice gemessen hat. Stimmt die gewählte Basis mit der von Alice überein (die Nummer der zugehörigen Photonen wurde unterstrichen), sind die Ergebnisse gleich, andernfalls sind sie zufällig.

## 2.4 Alice und Bob vergleichen ihre Basissysteme

Nachdem Alice ihre Photonen an Bob verschickt hat und Bob an diesen Photonen die genannten Messungen vorgenommen hat, vergleichen Alice und Bob über einen klassischen (offenen) Kanal die Polarisationsbasen (also die Orientierungen), die sie jeweils für ihre Photonen bewählt haben. In ungefähr der Hälfte der Photonen werden diese Orientierungen gleich sein.

Dieser klassische Kanal kann ruhig abgehört werden: Die ausgetauschte Information ist nicht mehr verwendbar, da Bob die Photonen schon vermessen hat und wegen des No-Cloning Theorems auch beim Austausch der Photonen keine Kopien angefertigt werden konnten. Alice und Bob müssen nur sicherstellen, dass sie tatsächlich miteinander kommunizieren und die ausgetauschte Information authentisch übertragen wird.

Sie tauschen natürlich nur die jeweils gewählten Basissysteme  $h/v$  bzw.  $+/-$  aus, keine Informationen über die dabei präparierten bzw. gemessenen Werte. Falls die Basis, die Alice zur Präparation verwendet hat, und die Basis, in der Bob die Messung vorgenommen hat, dieselbe ist, sollten die Werte übereinstimmen. Alle anderen Fälle werden verworfen, da in diesen Fällen die Bitwerte zufällig gleich oder verschieden sein können.

Bob verschickt über den klassischen Kanal im Wesentlichen die zweite Zeile seines Messprotokolls. Alice vergleicht diese Zeile mit ihrer zweiten Zeile und schickt an Bob die Nummern der Photonen zurück, für die beide Basen gleich sind. Das sind in obigem Fall die Photonen 2, 3, 7, 9, 10 und 13 (sie wurden in Tabelle 2 unterstrichen). Die Bit-Werte zu diesen Photonen sind beiden bekannt und sie sind gleich. Haben Alice und Bob ihre Basen zufällig gewählt, handelt es sich auch um eine Zufallsfolge. Diese Folge können sie als Schlüssel verwenden.

## 2.5 Eve

Da ein möglicher Lauscher nur in die Übertragung von Photonen oder von Information über öffentliche Kanäle eingreifen kann, bleibt Eve nur eine Möglichkeit: Sie muss die Photonen, die Alice an Bob verschickt, abfangen und ebenfalls Messungen an diesen Photonen vornehmen. Zu diesem Zeitpunkt ist noch nicht bekannt, bezüglich welcher Basis Bob seine Photonen ausmessen wird, da er diese Photonen noch nicht erhalten hat. Also wählt Eve zufällig für jedes Photon eines der beiden Basissysteme. Auch sie erhält so eine Folge von Bits sowie eine Tabelle mit der von ihr gewählten Basis (siehe Tab. 3).

Photon	1	2	3	4	5	6	7	8	9	10	11	12	13
Basis	$h/v$	$h/v$	$+/-$	$+/-$	$h/v$	$+/-$	$h/v$	$h/v$	$h/v$	$+/-$	$+/-$	$h/v$	$+/-$
Zustand	$v$	$v$	$-$	$+$	$h$	$-$	$h$	$v$	$h$	$-$	$+$	$h$	$-$
Bit	1	1	0	1	0	0	0	1	0	0	1	0	0

Tabelle 3: Tabelle von Eve der Photonenzustände, die sie an den Photonen von Alice gemessen hat. Stimmt die Basis mit der von Alice überein, sind die Ergebnisse wieder gleich (unterstrichene Photonenzahlen). Sie schickt Photonen in den von ihr bestimmten Zuständen an Bob. Das ist das Beste, was sie unter diesen Umständen machen kann. In rund der Hälfte der Fälle wird diese Basis mit der von Alice übereinstimmen. In den anderen Fällen verschickt sie die Photonen in einem anderen Polarisationszustand.

In den obigen Tabellen hat Eve für die Photonen 4, 5, 7, 11 und 13 dieselbe Basis gewählt wie Alice. Für die anderen Photonen sind ihre Ergebnisse zufällig. Sie verschielt nun eine Folge von Photonen an Bob, die exakt ihrer Tabelle entspricht, d.h., sowohl die Basissysteme sind für die einzelnen Photonen dieselben als auch die Polarisationen, die sie in der jeweiligen Basis für die Photonen erhalten hat. Sie kann nur hoffen, dass möglichst viele dieser Basissysteme mit den Basen von Alice übereinstimmen.

## 2.6 Überprüfung der Zufallsfolge

Der letzte Schritt, den Bob und Alice ausführen sollten, ist die Überprüfung, ob die ausgetauschten Photonen abgefangen und durch Messungen manipuliert wurden.

Photon	1	2	3	4	5	6	7	8	9	10	11	12	13
Alice Basis	$+/-$	$+/-$	$h/v$	$+/-$	$h/v$	$h/v$	$h/v$	$+/-$	$+/-$	$h/v$	$+/-$	$+/-$	$+/-$
Bit	1	0	0	1	0	0	0	0	1	0	1	1	0
Bob Basis (oE)	$h/v$	$+/-$	$h/v$	$h/v$	$+/-$	$+/-$	$h/v$	$h/v$	$+/-$	$h/v$	$h/v$	$h/v$	$+/-$
Bit	1	0	0	0	1	0	0	0	1	0	0	1	0
Eve Basis	$h/v$	$h/v$	$+/-$	$+/-$	$h/v$	$+/-$	$h/v$	$h/v$	$h/v$	$+/-$	$+/-$	$h/v$	$+/-$
Bit	1	1	0	1	0	0	0	1	0	0	1	0	0
Bob Basis (mE)	$h/v$	$+/-$	$h/v$	$h/v$	$+/-$	$+/-$	$h/v$	$h/v$	$+/-$	$h/v$	$h/v$	$h/v$	$+/-$
Bit	1	1	0	0	1	0	0	1	1	1	0	0	0

Tabelle 4: Die gesamte Tabelle des BB84-Protokolls. Die erste Doppelzeile gibt die Basis und die Bits an, in Bezug auf die Alice ihre Photonen präpariert hat. Die nächste Doppelzeile entspricht dem, was Bob für den Fall gemessen hätte, wenn Eve keine Photonen abgefangen hätte. Die dritte Doppelzeile zeigt die Ergebnisse von Eve und die letzte Doppelzeile die Ergebnisse von Bob, die er erhalten hat, nachdem Eve ihre Photonen so weitergeschickt hat, wie sie bei ihr gemessen wurden.

Tabelle 4 enthält nochmals alle Resultate, die Alice, Bob und Eve erhalten haben, wobei bei Bob unterschieden wird, ob er die Photonen direkt von Alice erhalten hat (oE - ohne Eve), oder ob Eve die Photonen abgefangen und an ihnen Messungen vorgenommen hat (mE - mit Eve).

Nachdem Alice und Bob ihre Basissysteme ausgetauscht haben (dieses Gespräch kann Eve abhören, sie kann zu diesem Zeitpunkt nicht mehr eingreifen) wissen sie, welche Bitfolge sie für ihren Schlüssel nehmen können. Zur Überprüfung verwenden sie nun eine ausreichende Anzahl dieser Bits und vergleichen diese über einen klassischen (offenen) Kanal. In der obigen Liste sind beispielsweise

Alice und Bob zu dem Schluss gekommen, dass sie bei den Photonen 2, 3, 7, 9, 10 und 13 dieselben Werte haben sollten. Vergleichen sie nun diese Bits über einen offenen Kanal stellen sie fest, dass nach dem Eingriff von Eve Photon 2 und 10 nicht zu demselben Bit gehören. Daraus können sie schließen, dass ihr Photonenaustausch abgefangen und manipuliert wurde. Sie werden nun sämtliche Bits ihrer Folge verwerfen und einen neuen Schlüsselaustausch versuchen.

Alice und Bob sollten also deutlich mehr Bits für ihren Schlüssel erstellen, als sie für die Verschlüsselung ihrer Nachricht benötigen. In der Praxis kann man mehrere Tausend Bits des Schlüssels vergleichen und so ziemlich sicher feststellen, ob der Schlüssel durch Eve manipuliert wurde. Eve hat, bei korrekter Durchführung des Protokolls, keine Möglichkeit, die fehlerhaften Bits zu unterdrücken.

Ganz grob kann man sagen, dass rund die Hälfte der Bits, die Alice und Bob mit dem Verfahren generieren, übereinstimmen und für die Verschlüsselung (bzw. einen Teil davon für den Test) verwendet werden können. Falls Eve die Photonen abgefangen und manipuliert hat, hat sie in rund der Hälfte dieser Fälle eine andere Basis gewählt als Alice und Bob, und davon wird in rund der Hälfte der Fälle das Bit bei Bob ein anderes sein als bei Alice. Ganz grob kann man also sagen, dass ungefähr ein Viertel der Bits, die Alice und Bob als gemeinsame Folge identifiziert haben, bei einem Eingriff von Eve andere Werte haben. Verwendet man einige Tausend Bits zur Verifikation des Schlüssels, sollte ein solcher Eingriff auffallen.

### 3 Schulische Teilrealisierung durch Laserlicht

Eine vollständige Realisierung dieses Protokolls scheitert in der Schule schon an dem Problem, dass kaum Experimente mit einzelnen Photonen möglich sein werden. Man kann das Protokoll aber teilweise realisieren, indem man Laserlicht mit Polarisationsfiltern präpariert bzw. misst. Die Zufallselemente, die bei Einzelphotonen bestimmen, welches Bit bei einer bestimmten Basis gemessen wird, kann man durch einen Würfel ersetzen. Im Folgenden werden nochmals die Schritte des Protokolls durchgespielt, wie man sie in der Schule mit einfachen Mitteln (Laserpointer, Polarisationsfilter und geeigneten Würfeln) umsetzen kann. Die folgenden Schritte sollten ausreichend oft wiederholt werden.

1. Alice würfelt für jedes „Photon“ eine Basis. Bei einem normalen Würfel kann man beispielsweise eine gerade Augenzahl für die Basis  $h/v$  wählen und eine ungerade Augenzahl für die Basis  $+/-$ . Es gibt aber auch Würfel, bei denen  $h/v$  und  $+/-$  schon auf den Würfelseiten verteilt sind. Sie würfelt ein zweites Mal und entscheidet damit, auf welche Polarisation der Filter hinter ihrem Laser eingestellt wird (diese Polarisation sollte natürlich mit der vorher gewürfelten Basis verträglich sein). Dann sendet Alice an Bob Laserlicht, das der entsprechenden Polarisation entspricht.
2. Bob entscheidet mit einem Würfel, bezüglich welcher Basis er das Laserlicht von Alice messen möchte. Er wählt nun diese Basis für die Polarisationsfilter, auf die er das Laserlicht von Alice schickt. Bei manchen Aufbauten kann Alice ihr Laserlicht durch einen Strahltreiber aufspalten und dann gleichzeitig auf die beiden orthogonal eingestellten Filter von Bob lenken. Nun gibt es zwei Möglichkeiten: (1) Alice und Bob haben dieselbe Basis gewählt. Dann hat das Laserlicht von Alice eine Polarisation, die nur von einem Filter bei Bob durchgelassen wird. Das zugehörige Bit zu diesem Filter (sowie die gewählte Basis) vermerkt Bob in seiner Liste. (2) Falls Bob eine andere Basis als Alice gewählt hat, erkennt er das daran, dass das Laserlicht von Alice durch beide Filter hindurchgeht (und etwas abgeschwächt wird). In diesem Fall würfelt er ein beliebiges Bit. Bei Einzelphotonen würde Bob in diesem Fall auch nur ein Ergebnis erhalten, d.h., er kann an diesem Punkt nicht feststellen, dass er die falsche Basis gewählt hat. Das gewürfelte Bit

wird später, nachdem die Basisstellungen ausgetauscht wurden, nicht gewertet, da sich dann herausstellt, dass die Orientierungen verschieden gewählt waren.

3. Im letzten Schritt tauschen Bob und Alice ihre Basissysteme aus und sollten nun für die Fälle, in denen die Basis gleich war, dieselben Ergebnisse erhalten haben.
4. Falls Eve in den Prozess eingeschaltet wird, macht sie folgende Schritte: Sie würfelt eine Basis und misst bezüglich dieser Basis die Polarisation des einfallenden Laserlichts. Stimmt ihre Basis mit der von Alice überein, misst sie nur hinter einem ihrer Filter Licht und vermerkt das entsprechende Bit. Sind die Basen von Alice und Eve verschieden, beobachtet sie hinter beiden Filtern Laserlicht und würfelt ein Bit. Sie schickt nun Laserlicht mit der Basis und Polarisation an Bob, die sie verwendet bzw. gemessen oder gewürfelt hat.

Im Wesentlichen an diesem Punkt scheitert das Protokoll in der Realität, wenn man tatsächlich mit Laserlicht statt mit Einzelphotonen arbeiten möchte. Eve kann feststellen, ob sie dieselbe Basis wie Alice gewählt hat oder nicht: Wenn Sie bei einer Basis hinter beiden Filtern Licht beobachtet, ist es die falsche Basis. Sie könnte nun die richtige Basis wählen und die Photonen des Laserlichts mit der richtigen Polarisation weiterleiten. Lauscht sie später dem Vergleich der Basissysteme zwischen Alice und Bob kann sie ebenfalls den Schlüssel ermitteln. Mit Einzelphotonen kann Eve die richtige Basis nicht feststellen.

5. Am Ende vergleichen Alice und Bob ihre Bits, von denen sie glauben, sie seien gleich.

Damit man ein Eingreifen von Eve bemerken kann, sollten insgesamt mindestens 15 bis 20 „Photonen“ ausgetauscht werden. Das kann bei sorgfältiger Durchführung des Experiments eine Weile dauern (insbesondere, wenn man die Schritte von Eve ebenfalls durchführen möchte), sodass leicht eine Doppelstunde mit diesem Protokoll benötigt wird. Andererseits macht es auch Spaß, wenn man am Ende die Bits vergleicht und feststellt, ob bzw. dass Eve eingegriffen hat. Nach dem Austausch von vier oder fünf „Photonen“ kommt auch eine gewisse Routine hinzu und es geht schneller. Außerdem kann man auf diese Weise den Ablauf des Protokolls wirklich miterleben und begreifen.

## 4 Fragen

Dieser Abschnitt enthält einige Fragen, die man zunächst selbst beantworten sollte. Sie eignen sich auch für die Diskussion in der Schule.

1. Weshalb sollte Bob die Orientierungen bei seinen Messungen zufällig wählen? Welche Gefahr besteht, wenn er für die Orientierungen beispielsweise abwechselnd  $h/v$  und  $+/-$  wählt?

Die Gefahr besteht darin, dass Eve diese Vorliebe von Bob kennt. In diesem Fall wählt sie dieselben Orientierungen bei den Messungen wie Bob. Wenn Alice und Bob später ihre gewählten Basissysteme vergleichen, erhalten sie dieselben Übereinstimmungen wie Alice und Eve. Bei den Photonen, bei denen die Orientierungen von Eve und Bob sich von denen von Alice unterscheiden, können die Messwerte verschieden sein, doch diese Bits werden verworfen. Bei den Bits, die behalten werden, stimmen Eve, Bob und Alice überein. Alice und Bob werden also nicht bemerken, dass sie belauscht wurden.

## Literatur

- [1] Bennett, C.H., Brassard, G., *Quantum cryptography: Public key distribution and coin tossing*; in *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, Vol. 175 (1984).

# Quantenradierer

Thomas Filk; Universität Freiburg

Quantenradierer zählen zu der Gruppe der sogenannten „verspätete Wahl“-Experimenten (delayed choice), bei denen die Entscheidung, welche von zwei komplementären Größen gemessen werden soll, erst gefällt wird, nachdem die entscheidende Wechselwirkung zwischen dem Quantensystem und einem Element seiner Umgebung bereits stattgefunden hat. In den meisten Fällen handelt es sich bei Quantenradierern um Abwandlungen des Doppelspaltexperiments, bei denen die „Welcher-Weg“-Information zunächst vorhanden ist, dann aber durch einen weiteren Quanteneffekt unwiderruflich wieder gelöscht werden kann. Die Entscheidung, ob man die „Welcher-Weg“-Information abrufen oder sie endgültig löschen möchte, wird gefällt, nachdem beispielsweise ein Photon den Doppelspalt durchquert hat.

Es gibt verschiedene Versionen von Quantenradierern, die unterschiedliche Eigenschaften von Photonen zur Speicherung der „Welcher-Weg“-Information nutzen. Die „klassische Version“ lässt sich auch mit gewöhnlichem Laserlicht durchführen und im Rahmen einer klassischen Wellentheorie des Lichts und seiner Polarisationseigenschaften beschreiben. Dieser Versuch wird auch gelegentlich in der Schule durchgeführt. Er ist nicht ganz so spektakulär wie andere Versionen des Quantenradierers, da bei ihm die Information wieder gelöscht wird, bevor die Photonen auf einen Schirm treffen und dort, je nachdem ob die „Welcher-Weg“-Information vorhanden ist oder nicht, ein Interferenzmuster zeigen. Bei Experimenten mit Einzelphotonen kann man oft die „Welcher-Weg“-Information abrufen, nachdem die Photonen bereits registriert wurden oder auf einen Schirm getroffen sind. Da zumindest im Prinzip die „Welcher-Weg“-Information vorhanden ist, sieht man zunächst kein Interferenzmuster. Man kann anschließend jedoch entscheiden, ob man die „Welcher-Weg“-Information löschen möchte oder nicht. Entcheidet man sich, im Rahmen einer Messung die „Welcher-Weg“-Information endgültig zu löschen, erhält man als Ergebnis dieser Messung eine andere Information, mit deren Hilfe man eine Postselektion der Photonenspuren auf dem Schirm vornehmen kann, sodass Interferenzmuster sichtbar werden.

Während Einzelphoton-Experimente auf absehbare Zeit für Schulen noch zu kostspielig sind, eignen sich die verschiedenen Einzelphoton-Quantenradierer sehr gut, um bei Schüler\*innen das Gespür, wann „Welcher-Weg“-Information vorliegt und wann nicht, zu stärken. Für die Polarisationszustände von Licht oder Einzelphotonen verwenden wir folgende Darstellung:

$$\text{horizontale und vertikale Polarisation} \quad |h\rangle, |v\rangle \quad (1)$$

$$\text{lineare Polarisation unter } \pm 45^\circ \quad |+\rangle, |-\rangle \quad (2)$$

$$\text{rechts- bzw. linkszirkulare Polarisation} \quad |R\rangle, |L\rangle. \quad (3)$$

## 1 Der „klassische“ Quantenradierer

Das einfachste Modell eines Quantenradierers lässt sich sowohl mit Laserlicht als auch mit Einzelphotonen durchführen. Außerdem kann man den Effekt sowohl im Rahmen einer klassischen Wellentheorie des Lichts und seiner Polarisationseigenschaften erklären als auch im Rahmen einer quantentheoretischen Beschreibung von Einzelphotonen. Interessant und relevant ist in diesem Zusammenhang, dass sich kohärentes Laserlicht in mehrfacher Hinsicht sehr ähnlich verhält wie Einzelphotonen. Insbesondere gilt dies für kohärente Superpositionen. Im Folgenden behandeln wir zunächst den klassischen Fall und die klassische Erklärung, anschließend betrachten wir die gleiche Situation für Einzelphoto-

nen und beschreiben sie ihm Rahmen des quantentheoretischen Formalismus.

## 1.1 Laserlicht und polarisierte elektrische Felder

Laserlicht beschreiben wir klassisch durch ein linear polarisiertes elektrisches Feld  $\vec{E}$  einer festen Wellenlänge  $\lambda$  (die zeitliche Abhängigkeit lassen wir unberücksichtigt; sie ändert nichts an den folgenden Überlegungen). Vertikal polarisiertes Laserlicht treffe auf einen Doppelspalt, wobei sich hinter dem rechten Spalt ein Polarisationsfilter unter  $+45^\circ$  und hinter dem linken Spalt ein Polarisationsfilter unter  $-45^\circ$  befindet. Unmittelbar hinter dem Doppelspalt haben wir somit zwei linear polarisierte elektrische Felder,  $\vec{E}_r^+(x)$  und  $\vec{E}_l^-(x)$ , wobei sich  $r$  und  $l$  auf den rechten bzw. linken Spalt beziehen (siehe Abb. 1).

Das Laserlicht ist nun markiert: Die beiden  $\pm$ -diagonalen Polarisationen zeigen an, welcher Teil des Laserlichts durch den linken und welcher durch den rechten Spalt getreten ist. Etwas weiter hinter dem Doppelspalt überlagern sich die beiden elektrischen Felder, wobei es zwischen dem  $+$ -polarisierten Anteil und dem  $-$ -polarisierten Anteil eine Phasenverschiebung  $\delta$  gibt, die von der jeweiligen Differenz in der optischen Weglänge abhängt:

$$\vec{E}_{\text{ges}}(x) = \vec{E}_l^-(x) + e^{i\delta} \vec{E}_r^+(x) \quad (4)$$

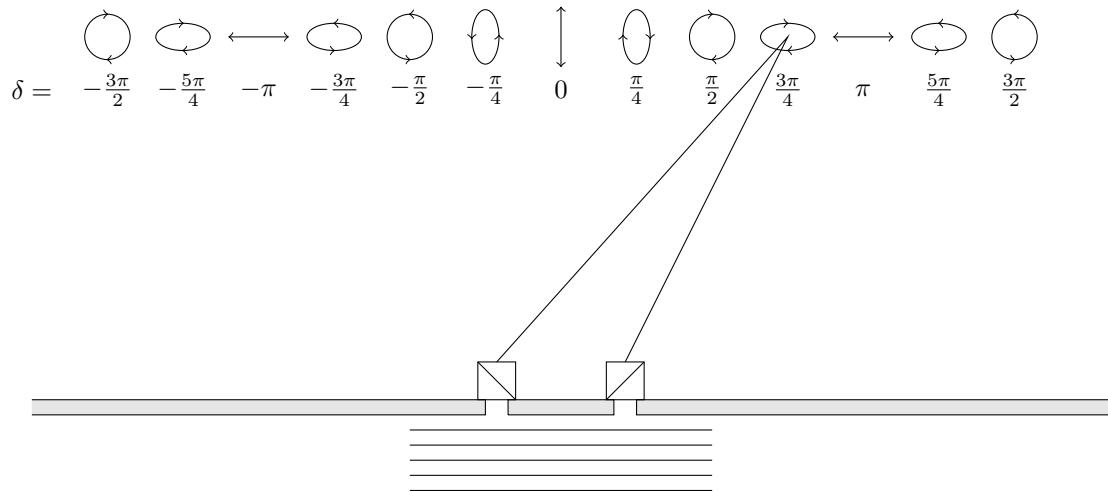


Abbildung 1: Der klassische Radierer: Licht trifft auf einen Doppelspalt, hinter dem sich  $\pm$ -Polarisationsfilter befinden. Dadurch wird der Lichtstrahl „markiert“. Je nach der Differenz in der optischen Weglänge überlagern sich die beiden Anteile zu  $h/v$ -linear polarisiertem Licht oder rechts-links-zirkular polarisiertem Licht bzw. allgemeiner elliptisch polarisiertem Licht.

Trifft dieses Licht auf einen Schirm, beobachtet man kein Interferenzmuster. Im Rahmen der klassischen Physik sagt man, dass Lichtanteile zu orthogonalen Polarisationen nicht miteinander interferieren. Platzieren wir zwischen den Doppelspalt und den Schirm einen Polarisationsfilter unter  $+45^\circ$  oder  $-45^\circ$ , verringert sich die Intensität um die Hälfte, es erscheint aber immer noch kein Interferenzmuster, weil in diesem Fall nur das Licht von einem der Spalte durch den Filter hindurchtritt und auf den Schirm trifft.

Platzieren wir jedoch zwischen Doppelspalt und Schirm einen Polarisationsfilter unter einer  $v$ - oder  $h$ -Orientierung, beobachtet man ein Interferenzmuster. Wie man in Abb. 1 erkennt, treten in regelmäßigen Abständen auf dem Schirm nur horizontale bzw. nur vertikale Polarisationen auf. Dies

liegt daran, dass sich die horizontale bzw. vertikale Polarisation als Linearkombination von  $\pm$ -diagonal polarisiertem Licht auffassen lassen:

$$\vec{E}^{h/v}(x) = \frac{1}{\sqrt{2}}(\vec{E}^+(x) \pm \vec{E}^-(x)). \quad (5)$$

Ist der relative Phasenwinkel zwischen den beiden Anteilen  $\delta = 0$ , erhalten wir vertikal polarisiertes Licht, bei einem relativen Phasenwinkel von  $\pm\pi$  erhalten wir horizontal polarisiertes Licht. Die Projektion der elliptischen Polarisationen auf einen  $h$ - oder  $v$ -Filter ergibt entsprechend geringere Intensitäten und insgesamt beobachtet man Interferenzstreifen.

Die beiden Interferenzmuster zu einem  $h$ -Polarisationsfilter bzw.  $v$ -Polarisationsfilter sind gleich, allerdings um eine halbe Phase versetzt: Dort, wo bei der  $v$ -Orientierung die Maxima sind, liegen beim  $h$ -Filter die Minima und umgekehrt. Insgesamt zeigt sich für die Intensität (ohne Berücksichtigung des Beugungsmusters der Einzelspalte, das streng genommen mit dem Beugungsmuster des Doppelspalts zu falten wäre, bei genügend schmalen Spalten im Vergleich zu ihrem Abstand aber vernachlässigbar ist) die Beziehung:

$$I_{\text{ges}} = I_h + I_v = I(\cos^2 \delta + \sin^2 \delta), \quad (6)$$

wobei  $I_{h/v}$  die Intensitäten sind, die man bei Verwendung eines horizontalen bzw. vertikalen Polarisationsfilters zwischen Doppelspalt und Schirm findet. Die Phasenverschiebung

$$\delta = \frac{\Delta x}{\lambda} \bmod 2\pi \quad (7)$$

( $\Delta x$  der Weglängenunterschied zwischen den beiden Teilstrahlen zu einem bestimmten Punkt am Schirm) ist in eine Koordinate entlang des Schirms umzurechnen (eine rein geometrische Beziehung zwischen dem Abstand Doppelspalt-Schirm,  $L$ , und der  $x$ -Koordinate entlang des Schirms:  $x/L = \tan \delta$ ).

Das Auftreten eines Interferenzmusters am Schirm kann man auch so deuten, dass keine „Welcher-Weg“-Information vorhanden ist. Die „Welcher-Weg“-Information, die unmittelbar hinter dem Doppelspalt durch die  $\pm$ -Polarisationsfilter dem Strahl mitgegeben wurde, wurde durch den zweiten Polarisationsfilter ( $h/v$ ) wieder gelöscht. Dem Licht hinter dem  $h/v$ -Filter kann man nicht mehr entnehmen, ob es ursprünglich mal  $++$  oder  $--$ -polarisiert war. Ein ähnliches Ergebnis würden wir erhalten, wenn wir statt der  $h/v$ -Polarisation des zweiten Filters eine  $L/R$ -Polarisation (links-rechts-zirkular polarisiert) herausfiltern würden. Auch in diesem Fall würde man ein Interferenzmuster beobachten (die beiden Muster zu einem  $R$ - bzw. einem  $L$ -Filter sind wieder um eine Phase von  $180^\circ$  verschoben; relativ zu den  $h/v$ -Mustern findet man eine Phasenverschiebung von  $90^\circ$ ).

Da man im Prinzip den Filter, den man zusätzlich vor den Schirm bringt, wählen kann, nachdem das Licht durch den Doppelspalt getreten ist, spricht man von einem “delayed choice” Experiment. Allerdings wird in diesem Fall die Information gelöscht, bevor das Licht auf den Schirm tritt.

## 1.2 Der klassische Quantenradierer mit Einzelphotonen

Wir können den klassischen Quantenradierer natürlich auch im Rahmen eines quantentheoretischen Formalismus beschreiben und beispielsweise das Verhalten von einzelnen Photonen verfolgen.

Ein Photon im Zustand  $|v\rangle$  treffe auf den Doppelspalt. Hinter dem Spalt und den beiden  $\pm$ -Polarisationsfiltern beschreiben wir dieses Photon durch den Zustand

$$|v\rangle \rightarrow \text{Doppelspalt} + \pm\text{-Polfilter} \rightarrow \frac{1}{\sqrt{2}}(|+\rangle \otimes |r\rangle + |-\rangle \otimes |l\rangle) \equiv \frac{1}{\sqrt{2}}(|+,r\rangle + |-,l\rangle). \quad (8)$$

Die Polarisationsfreiheitsgrade des Photons sind hinter dem Doppelspalt mit den räumlichen Freiheitsgraden (linker oder rechter Spalt) verschränkt. Der Zustand des Photons besteht aus zwei Anteilen, die sich jeweils auf einen Spalt und die dazugehörige Polarisation beziehen. Trifft dieses Photon nun ohne weitere optische Elemente auf den Schirm, besteht zwischen den beiden Anteilen eine vom Auftreffpunkt abhängige optische Weglängendifferenz in Form einer Phase. Für die Wahrscheinlichkeitsamplitude, dass das Photon an einer bestimmten Stelle  $x$  auftrifft, erhalten wir:

$$\langle x|\gamma\rangle = \frac{1}{\sqrt{2}}(|+\rangle\langle x|r\rangle + e^{i\delta(x)}|-\rangle\langle x|l\rangle) \quad (9)$$

Hier wurde die Wahrscheinlichkeitsamplitude, bei  $x$  aufzutreffen, auf die räumlichen Freiheitsgrade geschoben, da diese Amplitude nicht von der Polarisation abhängt. Allerdings hängt die Auftreffwahrscheinlichkeit auch nicht von  $r$  oder  $l$  ab und kann durch eine  $x$ -unabhängige Intensitätsdichte  $I$  ersetzt werden. Eine  $x$ -Abhängigkeit steckt lediglich in der Phasendifferenz  $\delta(x)$ . Da jedoch die Zustände  $|+\rangle$  und  $|-\rangle$  orthogonal sind, erhalten wir für die Wahrscheinlichkeitsdichte ein Photon bei  $x$  aufzutreffen:

$$|\langle x|\gamma\rangle|^2 = \frac{1}{2}I|1+1|^2 = I. \quad (10)$$

Es gibt keine Interferenzterme und keine  $x$ -Abhängigkeit, d.h., wir erhalten auf dem Schirm eine konstante Intensitätsverteilung. Wird ein Filter bezüglich der  $+$ - oder  $-$ -Orientierung vorgeschaltet, filtert dieser einen der beiden Terme heraus und wir erhalten die halbe Wahrscheinlichkeit für das Auftreffen eines Photons.

Wählen wir jedoch vor dem Schirm einen  $h$ -Polfilter, ändert sich das Ergebnis. Zunächst entwickeln wir die  $|\pm\rangle$ -Zustände nach der  $h/v$ -Basis:

$$|+\rangle = \frac{1}{\sqrt{2}}(|h\rangle + |v\rangle) \quad |-\rangle = \frac{1}{\sqrt{2}}(|h\rangle - |v\rangle). \quad (11)$$

Damit wird aus Gleichung 8:

$$|\gamma\rangle \rightarrow \frac{1}{\sqrt{2}}(|+,r\rangle + e^{i\delta(x)}|-,l\rangle) \rightarrow \frac{1}{2}((|h,r\rangle + |v,r\rangle + e^{i\delta(x)}|h,l\rangle - e^{i\delta(x)}|v,l\rangle). \quad (12)$$

Ein  $h$ -Polfilter eliminiert die Terme zur vertikalen Polarisation und wir erhalten

$$|\gamma\rangle \rightarrow h\text{-Polfilter} \rightarrow \frac{1}{2}(|h,r\rangle + e^{i\delta(x)}|h,l\rangle) = \frac{1}{2}|h\rangle \otimes (|r\rangle - e^{i\delta(x)}|l\rangle). \quad (13)$$

Die Information über den Spalt, durch welchen das Photon getreten ist, ist nun nicht mehr mit der Polarisation verschränkt. Es handelt sich um einen separierbaren Zustand. Für die Wahrscheinlichkeitsamplitude, ein Photon auf dem Schirm am Ort  $x$  anzutreffen, folgt:

$$|\langle x|\gamma\rangle|^2 = \frac{1}{4}|\langle h|h\rangle|^2I|1+e^{i\delta(x)}|^2 = \frac{I}{2}(1+\cos\delta). \quad (14)$$

In diesem Fall beobachtet man ein Interferenzmuster. Wählen wir statt eines  $h$ -Filters einen  $v$ -Filter wird aus dem Pluszeichen ein Minuszeichen. Man erhält ebenfalls ein Interferenzmuster, allerdings um eine halbe Periode verschoben. Die Summe der beiden Intensitäten ist  $x$ -unabhängig. Wir erhalten also dieselben Ergebnisse wie aus der klassischen Überlegung.

## 2 Quantenradierer nach Scully

Die Idee eines Quantenradierers geht auf Marlan O. Scully zurück [5]. Das Experiment in seiner ursprünglichen Version wurde 1999 von einer Gruppe um Scully durchgeführt [2]. Die Idee besteht

darin, dass ein Photon von einem Doppelspalt auf einen Schirm bzw. Detektor trifft. Ist keine „Welcher Weg“-Information vorhanden, sollten die Signale dort ein Interferenzmuster zeigen. Ist jedoch „Welcher Weg“-Information vorhanden, sollte eine breite Verteilung der Photonen resultieren. Die Entscheidung, ob man die „Welcher Weg“-Information auswertet oder sie unwiederbringlich löscht wird in diesem Experiment erst gefällt, nachdem das Photon auf dem Schirm aufgetroffen ist. Da zunächst die „Welcher Weg“-Information im Prinzip vorhanden ist, sieht man kein Interferenzmuster. Löscht man jedoch durch eine quantenmechanische Wechselwirkung (Messung einer zur „Welche Weg“-Information komplementären Größe) die „Welcher Weg“-Information, erhält man bei diesem Prozess eine Zusatzinformation, mit der man im Nachhinein die Photonen in zwei Klassen einteilen kann. Beide Klassen zeigen ein gegeneinander verschobenes Interferenzmuster.

Der Aufbau des Experiments ist in Abb. 2 skizziert. Ein Pumpplaser trifft auf einen Doppelspalt hinter dem sich ein BBO-(Bariumborat)-Kristall befindet. Der Laser kann über die sogenannte parametrisierte Fluoreszenz (parametric down-conversion) die Emission von zwei verschränkten Photonen in den Bereichen A oder B in dem Kristall induzieren. Die beiden Photonen verlassen den Kristall in unterschiedliche Richtungen, tragen zunächst aber beide die Information, ob sie aus Bereich A oder B stammen (der Zustand ist eine Superposition dieser beiden Möglichkeiten). Eines der Photonen wird über eine Linse in Richtung eines Detektors  $D_0$  gelenkt, wo die beiden Anteile des Zustands von A und B interferieren können.

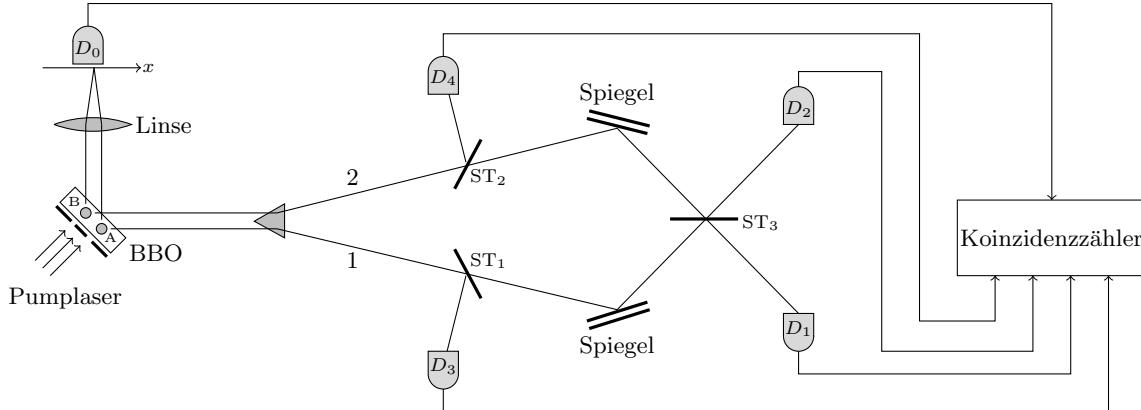


Abbildung 2: Quantenradierer nach Scully [2]. Ein Pumpplaser regt unmittelbar hinter einem Doppelspalt in einem BBO-Kristall die Emission von zwei Photonen aus dem Zentrum A oder dem Zentrum B an. Der Zustand besteht aus einer Superposition dieser beiden Prozesse. Eines der Photonen trifft über eine Linse auf Detektor  $D_0$ , der entlang der  $x$ -Achse verschoben werden kann. Das andere Photon folgt entweder Weg 1 oder Weg 2 zum Strahlteiler ST<sub>1</sub> bzw. ST<sub>2</sub>. Dort wird es entweder in Detektor 3 oder 4 abgelenkt oder von einem Spiegel auf einen zweiten Strahlteiler ST<sub>3</sub> gelenkt, hinter dem es entweder in Detektor 1 oder 2 landet. Ein Koinzidenzzähler vergleicht, welche Ereignisse bei  $D_0$  mit welchen Ereignissen bei  $D_1 - D_4$  zusammenfallen. Bei  $D_3$  und  $D_4$  ist die Weginformation bekannt, bei Detektor  $D_1$  und  $D_2$  wurde die Weginformation bei ST<sub>3</sub> ausgelöscht.

Das andere Photon wird auf einen von zwei Strahlteilern gelenkt (ST<sub>1</sub> und ST<sub>2</sub>) und dort entweder in einen Detektor gelenkt ( $D_3$  oder  $D_4$ ), oder aber durchgelassen und über einen Spiegel auf einen weiteren Strahlteiler ST<sub>3</sub> geleitet und von diesem entweder in Detektor  $D_1$  oder Detektor  $D_2$  geleitet. Trifft ein Photon auf die Detektoren  $D_3$  oder  $D_4$  ist bekannt, ob es von dem Bereich A oder B stammt. Trifft ein Photon jedoch auf Detektor  $D_1$  oder  $D_2$ , wurde diese Information bei ST<sub>3</sub> gelöscht. In diesem Fall erhält man aber eine Zusatzinformation, nämlich ob das Photon in Detektor  $D_1$  oder

in Detektor  $D_2$  gelandet ist, die man mit den korrespondierenden Ereignissen bei  $D_0$  verbinden kann.

Die Photonen bei  $D_0$ , die mit den Photonen in den Detektoren  $D_3$  und  $D_4$  koinzidieren, zeigen kein Interferenzmuster, da hier bekannt ist, von welchem Zentrum, A oder B, sie stammen. Die Photonen bei  $D_0$ , die mit Signalen in Detektor  $D_1$  korrelieren, zeigen ein Interferenzmuster, ebenso die Photonen bei  $D_0$ , die mit Signalen in Detektor  $D_2$  korrelieren, allerdings sind die beiden Interferenzmuster um eine Phase von  $180^\circ$  gegeneinander verschoben.

In der Version von Scully werden alle Informationen bzw. Situationen in demselben Experiment erfasst. Der Aufbau ist so, dass die Photonen bei  $D_0$  etwas früher (8 Nanosekunden) eintreffen als die anderen Photonen bei den Detektoren  $D_1 - D_4$ . Mit anderen Worten, die Photonen bei  $D_0$  können nicht wissen, ob letztendlich „Welcher Weg“-Information vorhanden sein wird oder nicht.

Das Experiment lässt sich in mehrfacher Hinsicht abwandeln: Man kann z.B. die Strahlteiler ST<sub>1/2</sub> weglassen, dann treffen alle Photonen auf ST<sub>3</sub> und die „Welcher Weg“-Information wird gelöscht. Man kann die beiden Strahlteiler auch durch Spiegel ersetzen. In diesem Fall wird die „Welcher Weg“-Information immer gewonnen. Man kann die Entscheidung, ob man diese Spiegel in den Strahlengang bringt, auch erst dann fällen, wenn das erste Photon schon in Detektor  $D_0$  registriert wurde.

### 3 Quantenradierer nach Walborn

Die folgende Darstellung eines Quantenradierers folgt [6] und ist teilweise [1] entnommen. Ähnlich wie bei dem Quantenradierer in Abschnitt 2 werden verschränkte Photonen verwendet, sodass man an einem Photon Information über das andere erhalten kann. Allerdings steckt die „Welcher-Weg“-Information nun im Polarisationsfreiheitsgrad der Photonen.

Abbildung 3 zeigt die experimentelle Anordnung dieses Quantenradiers. Aus einer Photonenquelle treffen Photonen auf einen BBO-Kristall, an dem durch Down-Conversion zwei in einem EPR-Zustand verschränkte Photonen der halben Energie erzeugt werden. Eines der Photonen (in der Abbildung oben) trifft auf einen Polarisationsdetektor (1), d.h. einen Polarisationsstrahlteiler, hinter dessen beiden Strahlgängen Detektoren stehen, sodass wir die Polarisation des Photons bezüglich einer voreingestellten Basis (beispielsweise horizontal/vertikal, also  $|h\rangle$  und  $|v\rangle$ , oder  $\pm 45^\circ$ , d.h.  $|+\rangle$  und  $|-\rangle$ ) messen können. Dieser Strahlteiler kann auch sehr weit hinter der Apparatur stehen, d.h., die entsprechende Information über das Photon kann theoretisch nach einer beliebig langen Zeit eingeholt werden.

Das zweite Photon trifft auf einen Doppelspalt. Hinter jedem der beiden Spalte befindet sich jeweils ein  $\lambda/4$ -Plättchen, wobei die ‚schnellen‘ Achsen der beiden Plättchen orthogonal zueinander sind. Photonen, die vor dem Doppelspalt bezüglich der  $+/-45^\circ$ -Achse polarisiert sind (also im Zustand  $|+\rangle$  oder  $|-\rangle$ ), erfahren durch die  $\lambda/4$ -Plättchen keine Änderung ihres Polarisationszustands sondern lediglich (je nach Spalt) eine Phasenverschiebung um  $\pm\pi/2$ . Sind die Photonen vor dem Doppelspalt jedoch bezüglich der  $h$ - oder  $v$ -Achsen polarisiert (also im Zustand  $|h\rangle$  oder  $|v\rangle$ ), beschreiben wir sie hinter den Plättchen durch einen  $|R\rangle$ - oder  $|L\rangle$ -Zustand. Diesen Zustand kann Polarisationsdetektor 2 bestimmen, d.h., dieser Detektor misst nicht nur, an welcher Stelle ein Photon ankommt, sondern auch, ob es links- oder rechtszirkular polarisiert ist.<sup>1</sup> In allen Fällen misst der Detektor 2 zunächst eine breite, unstrukturierte Verteilung ohne Anzeichen einer Interferenz.

Wir können nun entscheiden, ob wir die Information über den Spalt, durch den ein Photon

---

<sup>1</sup>Ein solches Nachweisgerät lässt sich im Prinzip aus einem  $\lambda/4$ -Plättchen und einem Polarisationsstrahlteiler für planare Polarisation – orientiert entsprechend der schnellen und langsamen Achse des  $\lambda/4$ -Plättchens – mit dahinter platzierten Detektoren herstellen. Wichtig ist nicht die konkrete Realisation, sondern dass diese Information tatsächlich gewonnen werden kann.

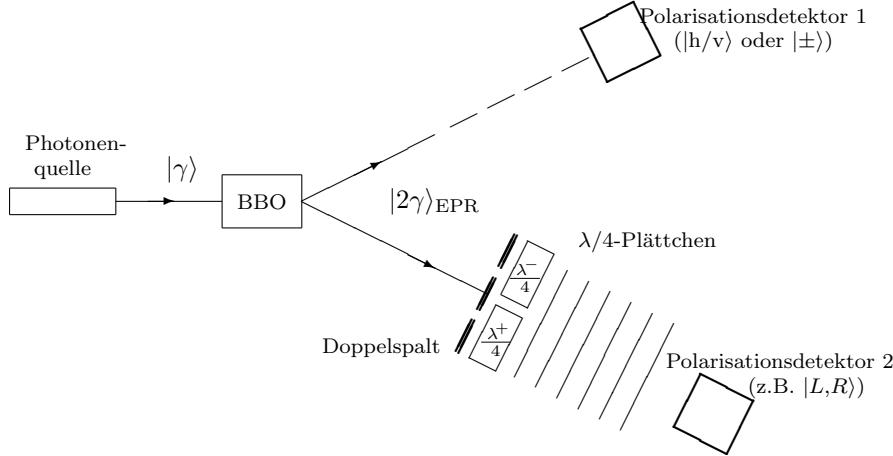


Abbildung 3: Aufbau eines Quantenradierers. Photonen aus einer Photonenquelle treffen auf einen BBO-Kristall, der zwei im EPR-Zustand verschränkte Photonen der halben Energie erzeugt. Eines der Photonen trifft auf einen Doppelspalt, hinter dem  $\lambda/4$ -Plättchen eine „Markierung“ eines Photons ermöglichen, die es im Prinzip erlaubt, später festzustellen, durch welchen Spalt es getreten ist.

getreten ist, messen wollen oder nicht. Wenn wir für das erste Photon die Basis des Polarisationsdetektors 1 auf  $h$  bzw.  $v$  einstellen, ist wegen der Verschränkung auch das zweite Photon, das durch den Spalt tritt, in dieser Basis polarisiert. Die  $\lambda/4$ -Plättchen machen aus dieser Polarisation eine zirkulare Polarisation, die für den rechten und linken Spalt jeweils entgegengesetzt ist. Kenn man also die  $h/v$ -Polarisation vor dem Spalt und misst die  $L/R$ -Polarisation hinter dem Spalt an Detektor 2, kann man von jedem Photon angeben, durch welchen Spalt es getreten ist. Die „Welcher-Weg“-Information ist also vorhanden und die Photonen zeigen kein Interferenzmuster.

Doch die Messung an Photon 1 (oben) kann sehr spät erfolgen (theoretisch Jahre später). Trotzdem ist die Information „irgendwo in unserem Kosmos“. Daher findet man auch kein Interferenzmuster für Photon 2, auch wenn die Messung der zirkularen Polarisation allein, ohne die zusätzliche Information der Polarisation vor dem Spalt, noch keinen Rückschluss auf den Spalt zulässt, durch den ein Photon getreten ist.

Angenommen, wir messen an Detektor 1 (möglicherweise wieder „Jahre später“) nicht die Polarisation bezüglich  $h$  und  $v$ , sondern bezüglich der Basis  $+$  bzw.  $-$ . Wegen der Verschränkung der beiden Photonen wissen wird damit auch, welche Photonen bei Detektor 2 vor ihrem Eintritt in den Doppelspalt im Zustand  $|+\rangle$  bzw.  $|-\rangle$  waren. In diesem Fall ist die „Welcher-Weg“-Information zwar endgültig verloren, doch nun können wir die Ereignisse, die von Detektor 2 aufgenommen wurden, hinsichtlich der Zustände  $+$  bzw.  $-$  nachträglich trennen (wir sortieren also den gesamten Datensatz nachträglich entsprechend der gewonnenen Information in zwei Klassen). Für jede der so gewonnenen Klassen finden wir nun das Interferenzmuster, denn für jede dieser Klassen ist die „Welcher-Weg“-Information gelöscht. Auf diese Weise können wir nachträglich die Interferenzmuster sichtbar machen. Da die beiden Interferenzmuster zu den beiden Klassen von Ereignissen jedoch um eine halbe Wellenlänge relativ zueinander verschoben sind, ist ihre Summe eine breite Verteilung ohne Interferenzstreifen.

Abbildung 4 fasst diese Situation nochmals in stilisierter Form zusammen. Teil a zeigt die gemessene zirkulare Polarisation der Photonen, nachdem sie durch den Doppelspalt mit den  $\lambda/4$ -Plättchen getreten sind. Die Verteilung der  $R$ - bzw.  $L$ -zirkular polarisierten Photonen ist zufällig und zeigt keinerlei Interferenz. Entscheiden wir uns, an dem Detektor für Photon (1) die  $h/v$ -Polarisation

zu messen, kennen wir auch die  $h/v$ -Polarisation der Photonen in Strahl 2, bevor sie auf den Doppelspalt getreten sind (diese Information ist in Abb. 4 b wiedergegeben). Aus diesen beiden Informationen können wir den Spalt bestimmen, durch den jedes einzelne Photon getreten ist. Hatte ein Photon vor dem Spalt eine  $h$ -Polarisation und wurde es nach dem Spalt mit einer  $R$ -Polarisation gemessen, wissen wir, dass das entsprechende Photon durch den rechten Spalt getreten ist (entsprechend bei einer  $L$ -Polarisation durch den linken Spalt). War es vorher  $v$ -polarisiert, ist die Situation umgekehrt (Abb. 4 c). Man beachte, dass erst beide Informationen zusammengenommen ( $h/v$ -Polarisation vor dem Spalt und  $R/L$ -Polarisation hinter dem Spalt) die ‚Welcher-Weg‘-Information liefern.

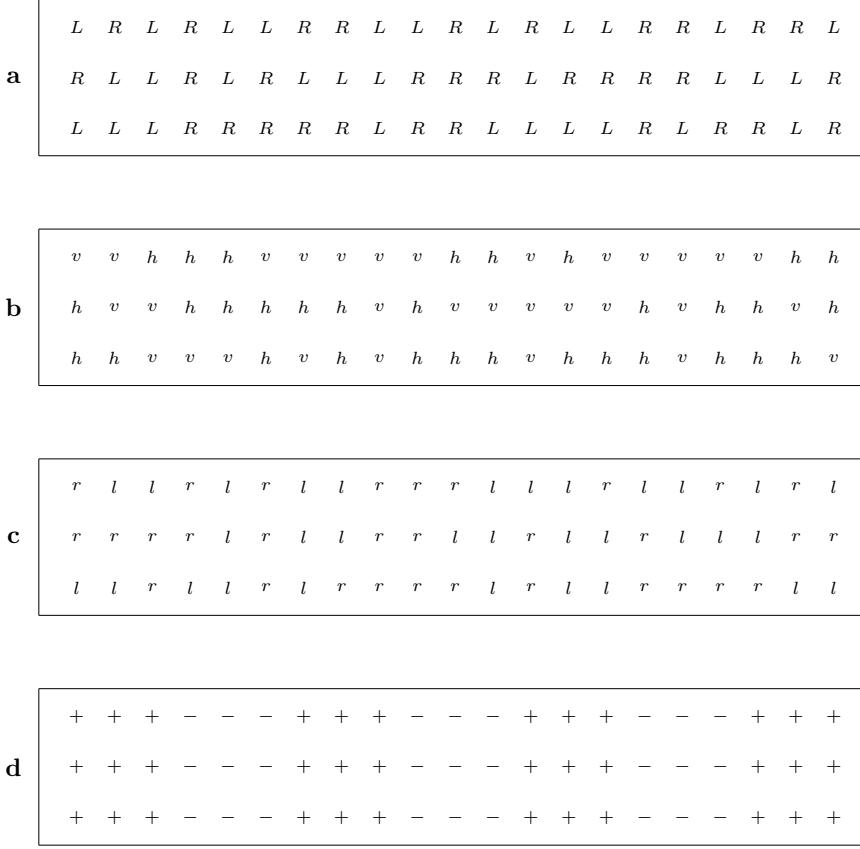


Abbildung 4: Stilisierte Darstellung der möglichen Ereignisse beim Quantenradierer. Jeder Buchstabe bzw. jedes Symbol repräsentiert ein gemessenes Ereignis. Gleiche Orte auf dem ‚Schirm‘ entsprechen auch gleichen Ereignissen.

Wird jedoch an Photon (1) die Polarisation bezüglich einer  $+/-$ -Basis gemessen, ändern die  $\lambda/4$ -Plättchen die Polarisation nicht und wir erhalten aus einer Messung der  $L/R$ -Polarisation hinter dem Spalt keine ‚Welcher-Weg‘-Information. Stattdessen zeigen sowohl die  $+$ - als auch die  $-$ -polarisierten Photonen jeweils ein Interferenzmuster (Abb. 4 d), die jedoch gegeneinander um eine halbe Interferenzbreite verschoben sind, sodass alle Photonen zusammen eine interferenzfreie Verteilung haben.

## 4 Quantenradierer nach Küblbeck

Josef Küblbeck hat zusammen mit Rainer Müller ein Buch geschrieben „Die Wesenszüge der Quantenphysik - Modelle, Bilder und Experimente“ [3], das sich in erster Linie an Physiklehrkräfte richtet. Kapitel [Interferometer](#) behandelt Experimente am Mach-Zehnder-Interferometer, unter anderem auch eine Version des Quantenradierers, die auf eine Arbeit von Ou, Wang, Zou und Mandel zurückgeht [4]. Küblbeck hat auch Unterrichtsmaterialien zum Quantenradierer erstellt.<sup>2</sup>

Die Aufgabe der Schüler\*innen besteht im Wesentlichen darin zu entscheiden, unter welchen Bedingungen die „Welcher Weg“-Information vorhanden ist und somit kein Interferenzmuster zu sehen ist, und wann die „Welcher Weg“-Information gelöscht wurde, sodass ein Interferenzmuster beobachtbar sein sollte. Der Aufbau des Quantenradierers ist in Abb. 5 wiedergegeben. In dieser Abbildung sind allerdings alle optischen Elemente für den Quantenradierer eingetragen. Man kann auch einige Elemente (z.B. die Strahlteiler ST<sub>2</sub> und ST<sub>3</sub>) weglassen und dann fragen, ob die Information über den Weg vorhanden ist oder nicht.

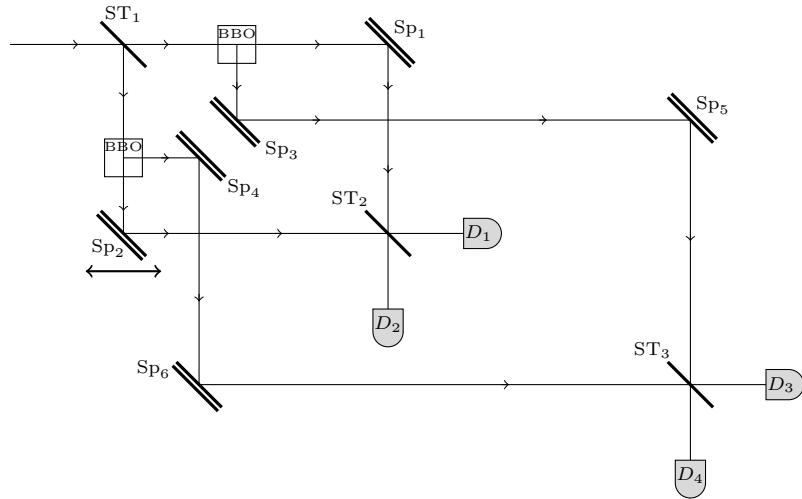


Abbildung 5: Der Quantenradierer nach Küblbeck [3]. Im Wesentlichen handelt es sich um zwei Mach-Zehnder-Interferometer für die beiden Photonen, die in den BBO-Kristallen von dem einfallenden Photon erzeugt werden. Die Situation ist symmetrisch: Jedes Photon enthält die Weginformation über das jeweils andere Photon. Sp<sub>i</sub> sind gewöhnliche Spiegel, ST<sub>i</sub> sind Strahlteiler und D<sub>i</sub> sind Detektoren für die Photonen.

Ein Photon tritt oben links in die Apparatur und trifft auf einen Strahlteiler ST<sub>1</sub>. Es kann nun zwei Wegen folgen. In beiden Wegen trifft es zunächst auf einen BBO-Kristall, wo das eine Photon in zwei Photonen umgewandelt wird, die nun in verschiedene Mach-Zehnder-Interferometer gelenkt werden. Über verschiedene Spiegel werden die Photonen auf einen zweiten Strahlteiler geleitet (ein Photon auf Strahlteiler ST<sub>2</sub>, das andere auf Strahlteiler ST<sub>3</sub>). Hinter beiden Strahlteiler befinden sich Detektoren - einmal die Detektoren D<sub>1</sub> und D<sub>2</sub> und einmal die Detektoren D<sub>3</sub> und D<sub>4</sub>. Durch Verschieben des Spiegels Sp<sub>2</sub> kann man die wechselnden Helligkeiten in den Detektoren und damit die Interferenz beobachten.

Das Besondere bei diesem Quantenradierer ist, dass ein Photon innerhalb eines Mach-Zehnder-Interferometers in einem BBO-Kristalle „verdoppelt“ wird (die beiden Photonen haben in

---

<sup>2</sup>Ich danke Herrn Küblbeck, dass er mir diese Materialien während einer Lehrkräftefortbildung in Dillingen zur Verfügung gestellt hat.

der Summe die Energie des ursprünglichen Photons). Beide Photonen tragen die Information des Weges, den das ursprüngliche Photon genommen hat. Sofern nur einer der beiden Strahlteiler  $ST_2$  oder  $ST_3$  vorhanden ist, sollte daher kein Interferenzmuster nachweisbar sein, auch nicht in den Detektoren, die hinter dem noch vorhandenen Strahlteiler sind. Sind jedoch beide Strahlteiler vorhanden, wie in Abb. 5, wird durch die Strahlteiler die „Welcher Weg“-Information gelöscht. Die Information, welcher der beiden Detektoren  $D_3$  oder  $D_4$  das Photon nachgewiesen hat, kann man nun nutzen um die bei  $D_1$  und  $D_2$  nachgewiesenen Photonen in zwei Gruppen zu unterteilen, die jeweils ein Interferenzmuster zeigen (und umgekehrt). Die beiden Interferenzmuster sind wiederum um  $180^\circ$  gegeneinander verschoben.

## Literatur

- [1] T. Filk; *Quantenmechanik (nicht nur) für Lehramtsstudierende*; Springer-Verlag 2019.
- [2] Y.-H. Kim, R. Yu, S. P. Kulik, Y. H. Shih, M. O. Scully; *A Delayed Choice Quantum Eraser* Phys. Rev. Lett. 84 (2000) 1–5.
- [3] J. Kübler, R. Müller; *Die Wesenszüge der Quantenphysik - Modelle, Bilder und Experimente*; Aulis-Verlag Deubner, 3. Auflage, 2007.
- [4] Z.Y. Ou, L.J. Wang, X.Y. Zou, L. Mandel; *Evidence for phase memory in two-photon down conversion through entanglement with the vacuum*; Phys. Rev. A 41 (1990) 566–568.
- [5] M. O. Scully, K. Drühl; *Quantum eraser: A proposed photon correlation experiment concerning observation and “delayed choice” in quantum mechanics*; Phys. Rev. A 25(4) (1982) 2208–2213.
- [6] S. P. Walborn, M. O. Terra Cunha, S. Pádua, C. H. Monken; *Double-slit quantum eraser*; Phys. Rev. A 65 (2002) 033818.

# Si-Einheiten

Thomas Filk, Universität Freiburg

Auf ihrer 26. Versammlung im November 2018 hat die CGPM (Conférence générale des poids et mesure - General Conference on Weights and Measures) ein neues Einheitensystem (SI - Système international d'unités) beschlossen, das am 20. Mai 2019 in Kraft trat. Dieses System basiert im Wesentlichen auf der Festlegung einer Zeiteinheit sowie der Festlegung bestimmter Naturkonstanten, mit denen ausgehend von der Zeiteinheit andere Grundeinheiten - Länge, Masse, Temperatur, Stromstärke, Mengeneinheit - definiert werden können.

## 1 Das neue System der Grundeinheiten

Das neue System der Grundeinheiten beruht auf den Festlegungen der in Tabelle 1 angegebenen Naturkonstanten.

Bezeichnung	Symbol	Wert und Einheit
Frequenz des Hyperfeinstrukturübergangs in Cäsium-133	$\Delta\nu_{\text{Cs}}$	9 192 631 770 Hz
Lichtgeschwindigkeit im Vakuum	$c$	299 792 458 m/s
Planck'sche Konstante	$h$	$6,626\,070\,15 \cdot 10^{-34}$ Js
Elementarladung	$e$	$1,602\,176\,634 \cdot 10^{-19}$ C
Boltzmann-Konstante	$k_B$	$1,380\,649 \cdot 10^{-23}$ J/K
Avogadro-Konstante	$N_A$	$6,022\,140\,76 \cdot 10^{23}$ mol <sup>-1</sup>
Photometrisches Strahlungsäquivalent (Monochromatische Strahlung von 540 THz)	$K_{cd}$	683 lm/W

Tabelle 1: Die fundamentalen Konstanten zur Festlegung der SI-Basis.

Das photometrische Strahlungsäquivalent ist eine technische Konstante, die eine physikalische Größe (eine Strahlungsleistung, ausgedrückt in Watt) mit einer physiologischen Größe, einer wahrgenommenen Helligkeit - ausgedrückt durch die Einheit Lumen - in Verbindung bringt. Hierauf werden wir nicht weiter eingehen.

Die Definition einer fundamentalen Frequenz  $\Delta\nu_{\text{Cs}}$  über einen bestimmten Übergang (dem Hyperfeinstrukturübergang im Grundzustand) in einem bestimmten Atom (Cäsium-133) gibt gleichzeitig eine physikalische Realisierung dieser Frequenz an. Atomuhren, bei denen die Frequenz des Cäsiumübergangs gemessen wird, bezeichnet man als „primäre Zeitnormale“. Es gibt andere Realisierungen (Übergänge im Wasserstoff oder in Rubidium etc.), die aber vorher am Cäsiumübergang geeicht werden müssen und mit mindestens derselben Genauigkeit und Stabilität reproduzierbar sein sollten. Solche Realisierungen bezeichnet man auch als „sekundäre Zeitnormale“.

Die fundamentalen Einheiten - Sekunde, Meter, Kilogramm, Ampere, Kelvin, Mol (und das Candela) - erhält man nun als Kombination dieser Größen. Wie diese fundamentalen Einheiten zu bestimmen sind, ist mit Ausnahme der Sekunde nicht festgelegt. Bei der Boltzmann-Konstanten und der Avogadro-Konstanten handelt es sich um Proportionalitätsfaktoren zwischen historisch festgelegten Größen, die man heute nicht mehr unterscheiden müsste. Die Planck'sche Konstante und die Lichtgeschwindigkeit im Vakuum sind universelle Naturkonstanten, die in vielen Naturgesetzen auf-

treten und somit auf unterschiedliche Weisen bestimmt werden können. Das neue SI-System lässt diese Bestimmung offen.

## 1.1 Die Sekunde

Eine Frequenz gibt an, wie oft sich ein periodisch wiederkehrendes Ereignis in einer gewissen Zeit-einheit wiederholt. Die Einheit Hertz (Hz) legt diese Zeiteinheit auf eine Sekunde fest. Eine Sekunde ist somit die Zeitdauer, in der sich die Schwingungen, die dem Übergang im Cäsium entsprechen, 9 192 631 770 mal wiederholen. Oder anders ausgedrückt: Die Periodendauer einer Schwingung entspricht

$$T = \frac{1}{\Delta\nu_{\text{Cs}}} = \frac{1}{9\,192\,631\,770} \text{s}. \quad (1)$$

Damit ist

$$1 \text{s} = 9\,192\,631\,770 T = 9\,192\,631\,770 \frac{1}{\Delta\nu_{\text{Cs}}}. \quad (2)$$

Diese Gleichung kann man auch so interpretieren, dass die Definition

$$\Delta\nu_{\text{Cs}} = 9\,192\,631\,770 \text{s}^{-1} \quad (3)$$

nach der Einheit Sekunde aufgelöst wird.

## 1.2 Das Meter

Bereits 1975 wurde der Wert der Lichtgeschwindigkeit im Vakuum als Naturkonstante, wie in Tabelle 1 angegeben, festgelegt. Damit lässt sich die Einheit für eine Länge - das Meter - auf die Einheit der Zeit zurückführen:

$$1 \text{ m} = \frac{c}{299\,792\,458} \cdot 1 \text{s} = \frac{9\,192\,631\,770}{299\,792\,458} \frac{c}{\Delta\nu_{\text{Cs}}} \approx 30,6633189885 \frac{c}{\Delta\nu_{\text{Cs}}}. \quad (4)$$

Diese Festlegung hat zwei äquivalente Interpretationen: Zum einen ist ein Meter der  $1/299\,792\,458$  Teil der Strecke, die das Licht im Vakuum in einer Sekunde zurücklegt, bzw. das  $30,6633\dots$  fache der Strecke, die das Licht im Vakuum in der Zeitdauer einer Periode der Cs-Schwingung zurücklegt. Andererseits entspricht das Verhältnis  $c/\Delta\nu$  der Wellenlänge einer Strahlung, die sich mit der Geschwindigkeit  $c$  ausbreitet und eine Frequenz  $\Delta\nu$  hat. Also ist 1 Meter das  $30,6633\dots$ -fache der Wellenlänge der Strahlung zu dem Cs-Übergang im Vakuum.

Damit erhalten wir auch eine Vorstellung von der Wellenlänge der Strahlung zu dem Cs-Hyperfeinstrukturübergang:  $1/30,6633\dots$  Meter oder ungefähr 3,26 cm. Es handelt sich also um eine Strahlung im Radiobereich. Per Definition sind Radiowellen alle Formen von Wellen, deren Frequenz unter 3000 GHz liegt, was hier offensichtlich der Fall ist.

## 1.3 Das Kilogramm

Das Produkt aus Planck'schem Wirkungsquantum und Frequenz, also  $h\nu$ , ist eine Energie. Es ist die Energie eines einzelnen Photons mit der Frequenz  $\nu$ . Über die Einstein'sche Gleichung  $E = mc^2$  können wir eine Energie mit einer Masse in Beziehung setzen. Offenbar hat  $h\nu/c^2$  die Einheit einer Masse, und wenn man die Naturkonstanten in SI-Einheiten ausdrückt, ist die Einheit dieser Masse das Kilogramm. Wenn wir den Ausdruck:

$$\frac{h\Delta\nu_{\text{Cs}}}{c^2} = \frac{6,626\,070\,15 \cdot 10^{-34} \cdot 9\,192\,631\,770}{(299\,792\,458)^2} \text{kg} \quad (5)$$

nach der Einheit Kilogramm auflösen, erhalten wir:

$$1 \text{ kg} = \frac{(299\,792\,458)^2}{6,626\,070\,15 \cdot 10^{-34} \cdot 9\,192\,631\,770} \frac{h\Delta\nu_{\text{Cs}}}{c^2} \approx 1,475\,5214 \cdot 10^{40} \frac{h\Delta\nu_{\text{Cs}}}{c^2}. \quad (6)$$

Man kann diese Gleichung folgendermaßen interpretieren: Ein Kilogramm entspricht über die Beziehung  $E = mc^2$  der Energie von  $1,475\dots \cdot 10^{40}$  Photonen, von denen jedes einzelne zu dem Strahlungsübergang im Cs-Atom gehört.

## 1.4 Das Ampere

Das neue SI-System definiert die Elementarladung als fundamentale Naturkonstante. Die Einheit - das Coulomb - ist gleich  $1 \text{ C} = 1 \text{ A} \cdot \text{s}$ . Damit hat das Produkt aus der Elementarladung und der Cs-Frequenz die Einheit Ampere:

$$e \cdot \Delta\nu_{\text{Cs}} = 1,602\,176\,634 \cdot 10^{-19} \cdot 9\,192\,631\,770 \text{ A} \approx 1,4728219827 \cdot 10^{-9} \text{ A}. \quad (7)$$

Dieses Produkt ist gleich der Stromstärke, die man erhält, wenn eine Elementarladung in der Periodendauer einer Cs-Schwingung durch eine vorgegebene Fläche tritt. Damit folgt für die Einheit Ampere:

$$1 \text{ A} = \frac{1}{1,602\,176\,634 \cdot 10^{-19} \cdot 9\,192\,631\,770} e \cdot \Delta\nu_{\text{Cs}} \approx 6,789\,6868 \cdot 10^8 e \cdot \Delta\nu_{\text{Cs}}. \quad (8)$$

Die Stromstärke von 1 Ampere entspricht also entweder dem Fluss von  $1/1,602\dots \cdot 10^{19} \approx 0,624 \cdot 10^{19}$  Elementarladungen pro Sekunde oder dem Fluss von  $6,789\dots \cdot 10^8$  Elementarladungen pro Periodendauer einer Cs-Schwingung.

## 1.5 Das Kelvin

Theoretisch könnte man auf eine eigene Temperaturskala verzichten. In allen relevanten Fällen, in denen die Temperatur  $T$  mit anderen physikalischen Größen in Beziehung gesetzt wird, tritt das Produkt  $k_B T$  mit der Boltzmann-Konstanten  $k_B$  auf. Dieses Produkt hat die Dimension einer Energie, also dieselbe Dimension wie  $h\nu$ . Es sind hauptsächlich historische Gründe, dass man der Temperatur nicht die Dimension der Energie gegeben hat.

Da die Boltzmann-Konstante in den Einheiten J/K angegeben ist, hat

$$\frac{h\Delta\nu_{\text{Cs}}}{k_B} = \frac{6,626\,070\,15 \cdot 10^{-34} \cdot 9\,192\,631\,770}{1,380\,649 \cdot 10^{-23}} \text{ K} \quad (9)$$

die Dimension einer Temperatur. Lösen wir diese Gleichung nach der Einheit Kelvin auf, folgt:

$$1 \text{ K} = \frac{1,380\,649 \cdot 10^{-23}}{6,626\,070\,15 \cdot 10^{-34} \cdot 9\,192\,631\,770} \frac{h\Delta\nu_{\text{Cs}}}{k_B} \approx 2,266\,6653 \frac{h\Delta\nu_{\text{Cs}}}{k_B}. \quad (10)$$

Die thermische Energie zu einem Kelvin entspricht also ungefähr der Energie des Strahlungsübergangs bei Cäsium. Etwas anders ausgedrückt: Wenn man jedem thermischen Freiheitsgrad eines Systems die Energie  $h\Delta\nu$  eines Photons aus dem Cs-Strahlungsübergang zuführt, erhöht sich seine Temperatur um ungefähr 1 Kelvin.

Hierbei handelt es sich um „ungefähr“-Werte, also Größenordnungen. Zum einen hängt die Beziehung zwischen der zugeführten Energie und der Temperaturerhöhung von der spezifischen Wärme ab und kann an Phasenübergängen sogar überhaupt keine Temperaturerhöhung zur Folge haben (latente Wärme), andererseits gibt die Beziehung

$$\langle \epsilon_{\text{kin}} \rangle = \frac{3}{2} k_B T \quad (11)$$

eine klare Beziehung zwischen dem Erwartungswert der kinetischen Energie eines Bestandteils einer Substanz und der Temperatur an, die lediglich die Boltzmann-Konstante als Faktor enthält.

## 1.6 Das Mol

Die Avogadro-Konstante ist eine historisch gewählte Konstante zwischen der Menge einer Substanz und der Anzahl ihrer elementaren Bestandteile (Atome, Moleküle, Ionen, etc.). Bei „Substanzen“ aus makroskopischen Bestandteilen (z.B. einer Gruppe von Menschen oder Billiardkugeln) würde man einfach deren Anzahl angeben. Zu einer Zeit, als die mikroskopische Natur der Materie noch nicht bekannt war, definierte man die Menge einer Substanz über bestimmte chemische Reaktionen, bei denen diese Substanz mit einer wohldefinierten Menge einer anderen Substanz reagierte. Willkürlich hat man dann eine bestimmte Menge an Kohlenstoff, ausgedrückt in Gramm, als die Einheit mol definiert. Später hat man dann durch Messungen die Anzahl der Kohlenstoffmoleküle bestimmt, die dieser Menge entspricht, also die Avogadro-Zahl.

Heute ist ein Mol einer Substanz genau die Menge, die aus  $N_A = 6,022\,140\,76 \cdot 10^{23}$  elementaren Bestandteilen dieser Substanz besteht.

# 2 Zur Geschichte der Grundeinheiten

## 2.1 Die Sekunde

Die Einteilung eines Tages in 24 Stunden finden wir schon im antiken Babylon bzw. Mesopotamien. Bis ins Mittelalter wurden Tag und Nacht in jeweils 12 Stunden unterteilt, was allerdings je nach Jahreszeit zu unterschiedlich langen Tag- und Nachtstunden führte. Diese sogenannten „Temporalstunden“ (horae inequaes) konnten sich deutlich unterscheiden und waren nur bei den Tag-und-Nacht-Gleichen - also den Äquinoktien um den 20. März und den 23. September - gleich. Erst mit dem Aufkommen von Räderuhren zu Beginn des 14. Jahrhunderts setzten sich die sogenannten horae aequinoctiales, also gleichlange Tag- und Nachtstunden, durch.

Den Begriff der Minute und Sekunde finden wir erst im Mittelalter. Dabei ist - fern aller Logik - die Minute eine Abkürzung von „pars minuta prima“ (einmal verminderter Teil) und die Sekunde eine Abkürzung von „pars minuta secunda“ (zweimal verminderter Teil).

Bis 1956 war die Sekunde definiert als der  $1/86\,400$ -ste Teil eines mittleren Sonnentages.<sup>1</sup> Nachdem man aber festgestellt hatte, dass auch die mittleren Sonnentage Schwankungen unterworfen sind, legte man 1956 die Sekunde als den Bruchteil  $1/31\,556\,925,9747$  des tropischen Jahres 1900 fest, wobei man hier ein im Rahmen einer Theorie hochgerechnetes Jahr 1900 aus den Bewegungsdaten der Erde um die Sonne am 31. Januar 1899, 12 Uhr, gewählt hat. Die so definierte Sekunde bezeichnet man als Ephemeridensekunde.

Die Definition der Sekunde über den Hyperfeinstrukturübergang in Cäsium-133 löste 1967 die Ephemeridensekunde ab. Es wird nicht ausgeschlossen, dass bis 2030 eine Neudefinition der Sekunde über atomare Übergänge im optischen Bereich erfolgt, da mit diesen eine deutlich höhere Genauigkeit erreicht werden kann.

---

<sup>1</sup>Die sogenannten wahren Sonnentage - beispielsweise die Zeitspanne zwischen zwei Sonnenhöchstständen an aufeinanderfolgenden Tagen - sind nicht immer gleich lang, siehe das Kapitel zur [Zeitgleichung](#). Der mittlere Sonnentag ist die über ein Jahr gemittelte Dauer eines Tages.

## 2.2 Das Meter

Bis in die frühe Neuzeit waren sehr unterschiedliche Längenmaßstäbe in Gebrauch. Elle, Fuß, Zoll (Daumenbreite) oder Schritt (heute noch im englischen Yard) konnten sich beispielsweise auf den jeweiligen Herrscher beziehen. Eintausend Doppelschritte (Lateinisch „mille passus“, hiervon leitet sich die Bezeichnung Meile ab) entsprachen im Römischen Reich einer Meile von etwas über 1,5 km. Später setzte sich die nautische Meile bzw. Seemeile mit 1852 Metern (das entspricht ungefähr einer Bogenminute am Äquator) als meist verwendete Form der Meile durch.

Ende des 18. Jahrhunderts definierte man die Einheit Meter als den 10 000 000-sten Teil des Meridianbogens (also des Längengrads) durch Paris vom Nordpol zum Äquator. Mittlerweile war durch genaue Vermessungen des Geoids (der Erdform) bekannt, dass die Erde keine Kugelform hat und somit der Umfang oder auch die Länge eines Meridianbogens davon abhängen, wo diese Größen gemessen werden. Man realisierte das so definierte Meter durch ein Urmeter, einen Platinstab, der in Paris gelagert wurde. Obwohl sich später herausstellte, dass die Messungen der Länge des Meridians durch Paris fehlerhaft waren, wurde 1889 international die Länge des Meters nach den alten Prototypen festgelegt und als Urmeter einer Platin-Iridium-Legierung angefertigt.

Das Meter als die Länge des in Paris gelagerten Urmeters war bis 1960 in Gebrauch. Zwischen 1960 und 1975 wurde das Meter über die Wellenlänge einer bestimmten Linie einer Krypton-Lampe definiert. Durch die Festlegung eines Zahlenwerts für die Lichtgeschwindigkeit im Vakuum wurde ab 1975 das Meter über die Sekunde definiert.

## 2.3 Das Kilogramm

Ursprünglich sollte das Kilogramm der Masse von einem Kubikdezimeter Wasser entsprechen, wobei die Temperatur von manchen Definitionen auf 0°C, von anderen auf die Temperatur maximaler Dichte (rund 4°C) festgelegt wurde. Da diese Definitionen nur schwer mit der erforderlichen Genauigkeit (im Bereich von Milligramm pro Kilogramm) reproduziert werden konnten, stellte man später Prototypen des Kilogramms aus Platin-Iridium-Legierungen her. Die Definition eines Kilogramms über solch einen Prototyp hatte bis 2018 Gültigkeit. Allerdings zeigte sich schon seit mehreren Jahren, dass die Masse des in Paris in einem Safe bei konstanter Temperatur gelagerten Prototypen und die Massen von ursprünglich gleich schweren Kopien signifikante Unterschiede aufwiesen. Daher wurde 2018 festgelegt, dass das Kilogramm über die Planck'sche Konstante und die Lichtgeschwindigkeit auf die Definition der Sekunde zurückgeführt werden soll.

## 2.4 Das Ampere

Streng genommen müsste man für die Gesetze des Elektromagnetismus keine neuen Einheiten einführen. Über die Coulomb-Kraft könnte man z.B. der Ladung eine Einheit in Bezug auf die drei Grundeinheiten Kilogramm, Meter und Sekunde geben. Dazu definiert man beispielsweise die Ladung über das Kraftgesetz in folgender Form:

$$F = \frac{q_1 q_2}{r^2}. \quad (12)$$

Die Ladung erhält dadurch eine Einheit  $\text{kg}^{\frac{1}{2}} \text{m}^{\frac{3}{2}} \text{s}^{-1}$ . Die freie Naturkonstante, die die Ladungen  $q_1$  und  $q_2$  im Abstand  $r$  mit einer Kraft  $F$  verbindet, wird zu 1 definiert. Experimentell überprüft werden kann nur, dass die Kraft proportional zum Produkt der beiden Ladungen und umgekehrt proportional zum Quadrat des Abstands ist. Das sogenannte CGS-System (bzw. die Gauß'schen Einheiten) beruht auf dieser Konstruktion. CGS steht für Zentimeter-Gramm-Sekunde, d.h. es gibt in diesem System zusätzlich noch Zehnerpotenzen zwischen den CGS-Einheiten und den MKS-Einheiten (Meter-Kilogramm-Sekunde). Eine weitere Freiheit gibt es bei der Proportionalität zwischen Ladung

und elektrischen Feld sowie bei der Proportionalität zwischen Stromstärke und magnetischem Feld. Neben der Einheit für die Ladung besteht also eine Freiheit für die Einheiten des elektrischen und magnetischen Feldes.

Während im 19. Jahrhundert das CGS-System bzw. das Gauß'sche System verbreitet waren, entschloss man sich gegen Ende des 19. Jahrhunderts, der Stromstärke eine eigene Einheit, das Ampere, zu geben. Man hätte ebenso gut (wie es im Prinzip heute der Fall ist) der Ladung eine eigene Einheit geben können. Auf diese Weise vermied man nicht ganzzahlige Exponenten für die Einheiten mancher Größen (wie beispielsweise für die Ladung im CGS-System). Mitte des 20. Jahrhunderts definierte man das Ampere über die Kraft zwischen zwei stromdurchflossenen Leitern. Diese Definition war im Wesentlichen bis 2019 gültig.

## 2.5 Das Kelvin

Die Definition einer Temperaturskala galt historisch als ein Problem. Rein subjektiv können wir zwischen kälteren und wärmeren Systemen unterscheiden und es ist eine Erfahrungstatsache, dass es eine Eigenschaft gibt, die mit diesem Empfinden korreliert, und die für Systeme, die längere Zeit in Kontakt sind, gleich wird. Diese Eigenschaft nennen wir Temperatur. Verschiedene Systeme reagieren jedoch sehr unterschiedlich auf Temperaturänderungen (Festkörper ändern beispielsweise ihre lineare Ausdehnung, Metalle ihren elektrischen Widerstand, Gase ihr Volumen oder ihren Druck etc.). Die Festlegung einer Skala erscheint daher zunächst willkürlich. Erst durch Experimente im 17. und 18. Jahrhundert fand man heraus, dass für Gase bei hohen Temperaturen und hohen Verdünnungen ein von dem jeweiligen Gas nahezu unabhängiges Ausdehnungsverhalten vorliegt. Die Gesetze von Boyle-Mariotte (um 1670; für eine feste Temperatur ist das Produkt aus Druck und Volumen konstant), von Gay-Lussac (um 1800; für einen festen Druck ist das Volumen proportional zur Temperatur, manchmal bezeichnet man dieses Gesetz auch als das Gesetz von Charles, der es 1786 entdeckte) und von Amontons (um 1700; für ein konstantes Volumen ist der Druck proportional zur Temperatur) erlaubten eine Definition der Temperatur, die in mehrfacher Hinsicht ausgezeichnet war: Sie war unabhängig von dem jeweiligen Gas sowie von anderen Parametern, sofern die Dichte klein genug und die Temperatur hoch genug war, sodass man verschiedene Systeme miteinander vergleichen konnte. Das ideale Gasgesetz in seiner heutigen Form als Zusammenfassung der drei oben genannten Gesetze wurde allerdings erst 1834 von Benoît Clapeyron formuliert.

1824 erschien eine Schrift von Nicolas Léonard Sadi Carnot, in der er im Wesentlichen den Carnot-Prozess beschrieb und eine Schranke für den Wirkungsgrad von Wärmemaschinen (also Maschinen, bei denen ein Temperaturunterschied zwischen zwei Systemen und der damit verbundene Wärmefluss, wenn man diese Systeme in Kontakt bringt, genutzt wird, um mechanische Arbeit zu leisten). Eine wichtige Folgerung dieses Prozesses war, dass man eine Temperaturskala definieren konnte, ohne sich auf eine spezielle Realisierung zu beziehen. Mathematisch kann man sagen: Die Temperatur ist der integrierende Faktor zwischen der nicht exakten Einsform Wärme und der exakten Einsform zur Entropie (d.h. dem Gradienten der Zustandsgröße Entropie).

Die heutige Celsius-Skala wurde 1742 von Anders Celsius formuliert: Er unterteilte die Temperatur zwischen dem Gefrier- und dem Siedepunkt von Wasser bei Normaldruck in 100 Teile. Allerdings ordnete Celsius dem Siedepunkt die Temperatur 0 und dem Gefrierpunkt die Temperatur 100 zu. Zwei Jahre später (nach dem Tod von Celsius) wurde diese Zuordnung umgedreht. Schon einige Zeit vor der Celsius-Skala führte Daniel Gabriel Fahrenheit um 1714 eine Temperaturskala ein, die auf drei Fixpunkten basierte:  $0^{\circ}\text{F}$  entspricht  $-17,8^{\circ}\text{C}$  und war damals die tiefste Temperatur, die man mit einer sogenannten Kältemischung (aus Eis, Wasser und Salmiak) erzeugen konnte. Fahrenheit glaubte, auf diese Weise negative Temperaturwerte vermeiden zu können. Der Gefrierpunkt von Wasser wurde zu  $32^{\circ}\text{F}$  festgelegt, und die Körpertemperatur eines gesunden Menschen zu  $96^{\circ}\text{F}$ , was mit

35,6 °C etwas niedrig ist.

William Thomson (First Baron Kelvin) schlug 1848 vor, die Celsius-Skala „zu verschieben“ und den Nullpunkt auf den absoluten Nullpunkt festzulegen, den man z.B. aus den idealen Gasgesetzen mit dem universellen Volumenausdehnungskoeffizienten  $\gamma = 1/T$  zurückrechnen konnte. Die neue Skala wurde später Kelvin genannt. Sie basierte auf dem absoluten Nullpunkt mit 0 K (bis 1967 sagte man auch „0 Grad Kelvin“) sowie dem Tripelpunkt von Wasser, der zu 273,16 K festgelegt wurde (also bei 0,01 °C).

Seit 2019 ist die Einheit Kelvin direkt über die Energieeinheit Joule definiert. Die Boltzmann-Konstante ist streng genommen keine Naturkonstante sondern verknüpft die historische Kelvin-Skala mit der Energieskala. Die Temperatur ist direkt proportional zur thermischen Energie, d.h. die thermische Energie ist zu  $k_B T$  definiert.

## 2.6 Das Mol

Dass es überhaupt eine Einheit für die Substanzmenge gibt, geht vermutlich auf die Chemie zurück. Ohne ihre Gesetze hätte man Anfang des 19. Jahrhunderts Mengen verschiedener Substanzen kaum vergleichen können.

John Dalton entdeckte um 1800 das Gesetz der konstanten Proportionen und das Gesetz der multiplen Proportionen. Ersteres besagt, dass in einer Substanz (also einer aus Molekülen bestehenden chemischen Zusammensetzung) das Massenverhältnis der Bestandteile (chemischen Elemente) immer gleich ist. Das zweite Gesetz besagt: Wenn zwei chemische Elemente mehrere Verbindungen eingehen können (z.B. die Stickoxide NO, NO<sub>2</sub>, N<sub>2</sub>O, N<sub>2</sub>O<sub>3</sub> etc.), geschieht dies immer in Massenverhältnissen, die relativ zu einander kleine Zahlen annehmen. Diese Gesetze unterstützten die Atomtheorie Daltons, allerdings konnte man zu dieser Zeit keine absoluten Größenordnungen angeben. Die Vermutung war aber, dass sich zwei Substanzen A und B im Prinzip in der Form AB oder AB<sub>2</sub> oder A<sub>2</sub>B etc. verbinden können. Wenn man somit willkürlich eine bestimmte Menge (im Sinne einer bestimmten Masse) der Substanz A als „Mengeneinheit“ bezeichnet, kann man aus den Massenverhältnissen in solchen chemischen Verbindungen die Masse der Substanz B bestimmen, die nach Daltons Atomhypothese dieselbe Anzahl von atomaren Einheiten haben sollte.

Kurze Zeit später - um 1812 - entdeckte Amedeo Avogadro das Avogadro'sche Gesetz (oftmals spricht man auch von der Avogadro'schen Vermutung, da die Atomhypothese damals nicht als erwiesen galt), wonach (ideale) Gase bei gleicher Temperatur, gleichem Druck und gleichem Volumen auch dieselbe Anzahl von Molekülen enthalten. Seine Vermutung wiederum ging auf eine Beobachtung von Gay-Lussac um 1800 zurück, die darin bestand, dass sich Gase, wenn diese sich zu einer chemischen Substanz verbinden, immer in entsprechenden Volumenverhältnissen miteinander reagierten (z.B. zwei Volumeneinheiten Wasserstoff mit einer Volumeneinheit Sauerstoff zu Wasser, H<sub>2</sub>O).

Es folgte ein Sprung bis in die Mitte des 19. Jahrhunderts: Josef Loschmidt fand 1865 eine interessante Beziehung zwischen der mittleren Weglänge  $l$  von Luftmolekülen bei einer bestimmten Temperatur  $T$ , dem Durchmesser  $d$  der Luftmoleküle und dem Verhältnis  $V_l/V_g$  - dem Volumen  $V_l$  von Luft in seiner flüssigen Form und dem Volumen  $V_g$  derselben Masse an Luft im gasförmigen Zustand bei derselben Temperatur  $T$  (er nannte dieses Verhältnis Kondensationskoeffizient):

$$d = 8l \frac{V_l}{V_g}. \quad (13)$$

Loschmidt wusste, dass diese Beziehung wegen mehrerer Näherungen, die er gemacht hatte, nicht exakt gilt, ihm ging es aber auch nur um eine Bestimmung der Größenordnung von  $d$ . Das Volumen flüssiger Luft war damals zwar noch nicht bekannt, allerdings verwendete er theoretische Überlegungen, dieses Volumen aus der chemischen Zusammensetzung und dem Vergleich mit

ähnlichen Gasen (z.B. Wasser, wo dieses Verhältnis bekannt war) abzuschätzen. Maxwell hatte in seiner Arbeit von 1860 unter Bezug auf Messungen von Stokes die mittlere freie Weglänge der Luftmoleküle bei Zimmertemperatur ziemlich gut auf etwas über 60 Nanometer (heute gibt man meist 68 nm an) abgeschätzt. Damit konnte Loschmidt den Durchmesser der Luftmoleküle mit rund 1 nm angeben. Es war dann wieder Maxwell, der erkannte, dass man aus dem Wert von Loschmidt für die Größe der Luftmoleküle auch deren Anzahl in einem bestimmten Volumen bzw. einer bestimmten Masse (und damit die heutige Avogadro-Zahl) bestimmen konnte. (Näheres in Abschnitt 3.2.)

Auch wenn sich die Verfahren zur Bestimmung der Avogadro-Zahl im Verlauf der Zeit verbesserten, blieb sie mit vergleichsweise großen Ungenauigkeiten behaftet. Daher behielt man bis 2018 die historische Definition für ein Mol bei: Das Mol ist die Stoffmenge eines Systems, das aus ebensoviel Einzelteilchen besteht, wie Atome in 0,012 Kilogramm des Kohlenstoffnuklids  $^{12}\text{C}$  enthalten sind. Durch die Festlegung der Avogadro-Zahl wurde die Einheit Mol eigentlich überflüssig, sie ist jedoch aus praktischen Gründen immer noch sinnvoll.

### 3 Historische Kuriositäten

Im Zusammenhang mit der Festlegung der fundamentalen Einheiten gibt es einige interessante und teilweise amüsante historische Anmerkungen.

#### 3.1 Jean Picard und das Meter

Eine interessante Idee, eine von körperlichen Gliedmaßen unabhängige Definition des Meters zu finden, stammt von Jean Picard, der 1668 vorschlug, als Längeneinheit die Länge eines Pendels mit der Halbperiode von einer Sekunde zu wählen. Diesem Vorschlag ging die Beobachtung Galileis um 1630 voraus, dass ein Pendel für kleine Auslenkungen eine wohldefinierte Periode (unabhängig von dem Grad der Auslenkung) hat; außerdem hatte Christiaan Huygens um 1656 das Prinzip der Pendeluhr erfunden. Nach der Formel für die Periodenlänge  $T$  eines Pendels der Länge  $l$  im Schwerkraftfeld der Erde mit der Erdbeschleunigung  $g$  bei kleinen Auslenkungen,

$$T = 2\pi \sqrt{\frac{l}{g}}, \quad (14)$$

folgt aus  $T/2 = 1\text{ s}$  mit  $g = 9,81\text{ m/s}^2$  die Länge  $l = 0,99396\text{ m}$ . Dieser Vorschlag Picards ist insfern interessant, weil hier ein Längenmaßstab über eine „Naturkonstante“ ( $g$ ) auf eine Zeiteinheit zurückgeführt wurde. Vermutlich trug die schlechte Reproduzierbarkeit der Sekunde dazu bei, dass sich dieser Vorschlag nicht durchsetzte. Dass  $g$  an verschiedenen Orten der Welt unterschiedliche Werte hat, stellte sich erst später heraus, wobei schon Picard als Referenzort Paris vorgeschlagen hatte.

#### 3.2 $\sqrt{2}$ oder $4/3$ - wenn große Geister uneins sind

Wenn sich zwei große Wissenschaftler - in diesem Fall James Clerk Maxwell und Rudolf Clausius - um einen Faktor streiten, kann man davon ausgehen, dass ein interessantes Problem dahintersteckt.

1858 formulierte Rudolf Clausius eine Beziehung zwischen der mittleren freien Weglänge  $l$  von Atomen bzw. Molekülen in Gasen und ihrer Größe [1]. Die Herleitung dieser Formel erfolgt heute meist in folgender Form: Angenommen, wir haben einen Behälter vom Volumen  $V$  mit  $N$  kugelförmigen Teilchen ( $N$  sehr groß, Teilchendichte  $\rho = N/V$ ), die in diesem Volumen zufällig verteilt sind und zunächst als in Ruhe angenommen werden. Für ein weiteres Teilchen, das sich mit der

Geschwindigkeit  $v$  durch diesen Behälter bewegt, ist  $s = vt$  die in der Zeit  $t$  zurückgelegte Strecke. Die Frage ist: Wie oft stößt dieses Teilchen auf dieser Strecke mit anderen Teilchen zusammen, wobei der Wirkungsquerschnitt für einen solchen Zusammenstoß  $\sigma$  (Dimension einer Fläche) sein soll?  $\sigma \cdot vt$  ist das Volumen des Zylinders, den das Teilchen mit seinem Wirkungsquerschnitt für einen Zusammenstoß mit einem ruhenden Teilchen in der Zeit  $t$  überstrichen hat. In diesem Zylinder befinden sich  $\rho\sigma \cdot vt$  ruhende Teilchen, mit denen das fliegende Teilchen im Prinzip zusammengestoßen ist; dies ist also die Anzahl der Zusammenstöße. Die mittlere freie Weglänge  $l$  ist nun die zurückgelegte Strecke  $vt$  dividiert durch die mittlere Anzahl der Zusammenstöße, also

$$l = \frac{vt}{\rho\sigma \cdot vt} = (\rho\sigma)^{-1}. \quad (15)$$

Wenn sich die anderen Teilchen nicht in Ruhe befinden, sondern eine mittlere Geschwindigkeit  $u$  haben, zählt für das in der Zeit  $t$  überstrichene Volumen die mittlere Relativgeschwindigkeit zwischen  $v$  und  $u$ ; diese ist  $v_{\text{rel}} = \sqrt{v^2 + u^2}$  (das Skalarprodukt  $\mathbf{u} \cdot \mathbf{v}$  verschwindet im Mittel). Da  $u$  und  $v$  im Mittel gleich sind, erhalten wir einen Faktor  $\sqrt{2}$ :

$$l = \frac{vt}{\rho\sigma \cdot v_{\text{rel}}t} = (\sqrt{2}\rho\sigma)^{-1}. \quad (16)$$

Clausius hatte in seiner Arbeit von 1858 [1] statt des Faktors  $\sqrt{2}$  ohne eine weitere Begründung den Faktor  $4/3$  verwendet. James Clerk Maxwell hatte in seiner Arbeit von 1860, in der er auch die heute als Maxwell'sche Geschwindigkeitsverteilung bekannte Verteilungsformel für die Geschwindigkeiten der Moleküle bzw. Atome bei einer bestimmten Temperatur herleitete [3], den Faktor  $\sqrt{2}$  nach obiger Argumentation hergeleitet und in einer Notiz vermerkt, dass Clausius statt dessen den Faktor  $4/3$  verwendet habe. Daraufhin fühlte sich Clausius genötigt, seinen Faktor  $4/3$  zu begründen [2]. Bis zum ersten Gleichheitszeichen in Gl. 16 stimmen beide Autoren überein. Doch Claudius berechnet die mittlere relative Geschwindigkeit, indem er über die Richtungen aller Geschwindigkeiten  $u$  der anderen Teilchen mittelt:

$$v_{\text{rel}} = \frac{1}{2} \int_0^\pi \sqrt{v^2 + u^2 - 2uv \cos \theta} \sin \theta d\theta. \quad (17)$$

Das Integrationsmaß  $\frac{1}{2} \sin \theta d\theta$  begründet Clausius mit dem Argument, dass jede Richtung gleichwahrscheinlich sei und daher proportional zum Flächenelement zu dieser Richtung. Die relative Häufigkeit, in die Richtung  $\theta$  abgelenkt zu werden, ist somit  $2\pi \sin \theta d\theta / 4\pi$  (Flächenelement zum Winkel  $d\theta$  relativ zur Kugeloberfläche). Heute würden wir das Maß in Winkelkoordinaten für die Kugeloberfläche schreiben,  $\frac{1}{4\pi} \sin \theta d\theta d\varphi$ , und da der Integrand nicht von  $\varphi$  abhängt, können wir das Integral über  $\varphi$  ausführen (es ergibt einen Faktor  $2\pi$ ) und wir erhalten dasselbe Ergebnis.

Das Integral lässt sich mit  $\sin \theta d\theta = d \cos \theta$  leicht berechnen und man erhält:

$$v_{\text{rel}} = \frac{1}{6uv} (v^2 + u^2 - 2uv \cos \theta)^{3/2} \Big|_0^\pi = \frac{1}{6uv} (|v+u|^3 - |v-u|^3) = \begin{cases} v + \frac{u^2}{3v} & u \leq v \\ u + \frac{v^2}{3u} & u > v \end{cases}. \quad (18)$$

Für  $u = v$  ist somit  $v_{\text{rel}} = \frac{4}{3}v$ . Diese Argumentation klingt sehr vernünftig.

In den „Collected Papers“ von Maxwell aus dem Jahre 1890 ist auch die 1860er Arbeit wiedergegeben [4], allerdings folgt der Bemerkung, dass Clausius den Faktor  $4/3$  verwendet, eine Fußnote, die der Herausgeber hinzugefügt hat. Dort heißt es, dass das Ergebnis von Clausius zwar im Prinzip richtig ist, Clausius aber immer nur von einer mittleren Geschwindigkeit  $u$  bzw.  $v$  der Teilchen ausgegangen ist und für  $u = v$  annimmt, dass alle Teilchen dieselbe Geschwindigkeit haben und lediglich ihre Richtungen gleichverteilt sind. Er ersetzt nun den konstanten Geschwindigkeitsbetrag für  $u$  und  $v$  (mit der Mittelung über alle Richtungen - hier verwendet der Herausgeber die Formeln von Clausius)

durch die Maxwell'sche Geschwindigkeitsverteilung und behauptet (ohne Beweis), dass dies auf den gewünschten Faktor  $\sqrt{2}$  statt  $4/3$  führe.

Loschmidt verwendete die Clausius'sche Beziehung zwischen der mittleren freien Weglänge und dem Durchmesser (sowie der Dichte) der Moleküle. Die „heutige“ Formel beruht auf dem von Maxwell korrigierten Faktor. Allerdings sind beide Gleichungen ohnehin nur als Näherungen aufzufassen, da sie beispielsweise den Leerraum zwischen Molekülen in der flüssigen Phase vernachlässigen.

## Literatur

- [1] Rudolf Clausius; *Über die mittlere Länge der Wege, welche bei der Molecularbewegung gasförmiger Körper von den einzelnen Molekülen zurückgelegt werden; nebst einigen anderen Bemerkungen über die mechanische Wärmetheorie*; Annalen der Physik 181 (1958), p. 239–258.  
<https://era-prod11.ethz.ch/zut/ch19/content/zoom/15344023>
- [2] Rudolf Clausius; *On the dynamical theory of gases*; Philosophical Magazine 19, 4. Series (1860); p. 434–436.
- [3] James Clerk Maxwell; *Illustrations of the dynamical theory of gases – Part I. on the motions and collisions of perfectly elastic spheres*; Philosophical Magazine 19; 4. Series (1860); p. 19–32.
- [4] James Clerk Maxwell; *The Scientific Papers of James Clerk Maxwell*; Cambridge University Press, 1890.
- [5] Johann Josef Loschmidt; *Zur Größe der Luftmoleküle*; Sitzungsbericht der kais. Akad. der Wissenschaften, Band 52 (1866) Abt. II; p. 395–413. <https://mpoweruk.com/timekeepers.htm>

# Kalendersysteme

Thomas Filk, Universität Freiburg

Der (Sonnen-)Tag als die Zeitdauer zwischen zwei aufeinanderfolgenden Sonnenhöchstständen, der (synodische) Monat als die Zeitdauer zwischen zwei aufeinanderfolgenden Voll- oder Neumonden und das (tropische) Jahr als die Zeitdauer zwischen zwei Sonnendurchgängen durch den Frühlingspunkt gaben insbesondere in der Antike die wichtigen Perioden der Zeitrechnung vor. Dabei hat die Tatsache, dass weder die Mondphasen noch ein ganzzahliges Vielfaches eines Sonnentages über einen längeren Zeitraum mit den Jahreszeiten und den damit zusammenhängenden Erscheinungen (z.B. die regelmäßig im Juli bis September auftretenden Nilschwemmen in Ägypten) übereinstimmen, zu teilweise sehr komplizierten Kalendersystemen geführt. Tabelle 1 enthält die wichtigsten Zahlen, die in diesem Kapitel benötigt werden.

tropisches Jahr in Tagen	365,24219 d
Julianisches Jahr in Tagen	365,25 d
Gregorianisches Jahr in Tagen	365,2425 d
synodischer Monat	29,5306 d

Tabelle 1: Die wichtigsten physikalischen Größen im Zusammenhang mit den Kalendersystemen.

Außerdem benötigen wir noch die folgenden Begriffe, die ausführlicher in Kapitel [Nachthimmel](#) erläutert werden:

Unter der *Ekliptik* versteht man einen Großkreis am Nachthimmel, den man erhält, wenn man die Sonne von der Erde aus an den Himmel projiziert. Umgekehrt kann man auch die Projektion der Erde vom Mittelpunkt der Sonne aus an den Nachthimmel als Ekliptik definieren. Die Umlaufbahn der Erde um die Sonne liegt dann in der Ekliptikebene. Der *Himmelsäquator* ist die Projektion des Erdäquators an den Himmel vom Mittelpunkt der Erde aus betrachtet. Man kann ihn auch als den Großkreis am Himmel definieren, der senkrecht zum Himmelsnordpol (der Projektion der Erdachse an den Himmel) steht. Diese beiden Großkreise (Ekliptik und Himmelsäquator) schneiden sich in zwei Punkten: dem *Frühlingspunkt* und dem *Herbstpunkt*. Befindet sich die Sonne im Frühlings- oder Herbstpunkt, sind Nacht und Tag gleich lang, daher spricht man auch von den *Äquinoktien*.

## 1 Tage, Monate und Jahre

Wir alle wissen, was im Alltag gemeint ist, wenn wir von „Tag“, „Monat“, oder „Jahr“ sprechen, und doch erweisen sich diese Begriffe als recht komplex und vieldeutig, wenn man versucht, sie präziser zu definieren.

### 1.1 Tage

Unter einem Tag verstehen wir im Allgemeinen den Zeitraum zwischen zwei gleichen Sonnenständen, z.B. zwei Sonnenhöchstständen, also von Mittag bis zum Mittag des nächsten Tages. Bei Sonnenhöchststand steht die Sonne für jeden Beobachter nördlich des nördlichen Wendekreises (dem Breitengrad bei  $23,4^\circ$ ) exakt im Süden, für einen Beobachter südlich des südlichen Wendekreises exakt im Norden. Für Beobachter zwischen diesen beiden Wendekreisen hängt der Sonnenhöchststand

von der Jahreszeit ab. In jedem Fall befindet sich die Sonne zum Zeitpunkt „12 Uhr mittags (wahre Zeit)“ auf einer gedachten Linie, die den Längengrad des Beobachters (also den Nord-Süd-Meridian, der durch den Ort des Beobachters verläuft) vom Erdmittelpunkt aus an den Himmel projiziert. Ein solcher Tag heißt *wahrer Sonnentag*.

Da der wahre Sonnentag aus verschiedenen Gründen im Laufe eines Jahres in seiner Länge schwanken kann (siehe das Kapitel zur [Zeitgleichung](#)), definiert man einen sogenannten *mittleren Sonnentag*, das ist ein über das Jahr genommenes Mittel der wahren Sonnentage. Dieser mittlere Sonnentag wird in 24 Stunden bzw. 86 400 Sekunden unterteilt.

Versteht man unter einem Tag, dass sich die Erde einmal um ihre Achse gedreht hat, muss man einen Bezugspunkt angeben, der „einmal rum“ spezifiziert. Ist dieser Bezugspunkt die Sonne, erhalten wir den oben beschriebenen Sonnentag. Handelt es sich bei diesem Bezugspunkt aber um den Sternenhimmel, also z.B. einen bestimmten Fixstern, dessen Eigenbewegung wir vernachlässigen können, erhalten wir einen *Sternentag* oder *siderischen Tag*.

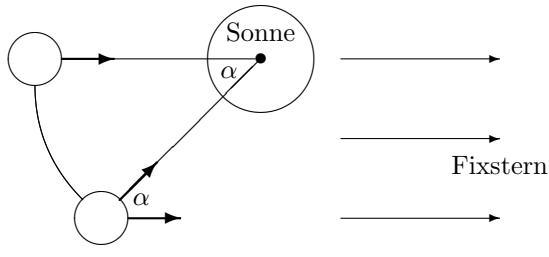


Abbildung 1: Siderischer Tag und Sonnentag.  
Da sich die Erde im Verlauf eines Tages um den Winkel  $\alpha$  weiterbewegt hat (hier übertrieben dargestellt), muss sie sich relativ zum Fixsternhimmel um diesen Winkel weiter drehen, damit ein bestimmter Punkt wieder in Richtung Sonne zeigt.

Siderischer Tag und (mittlerer) Sonnentag unterscheiden sich um ein paar Minuten. Grund ist, dass sich die Erde im Laufe eines Tages etwas weiter um die Sonne bewegt hat und daher die Sonne nach einem Tag nicht mehr exakt unter derselben Richtung steht (im Vergleich zum Fixsternhimmel) wie vorher. Ganz grob kann man den Unterschied folgendermaßen abschätzen: Die Erde bewegt sich in rund 365 Tagen einmal um die Sonne, d.h. an einem Tag bewegt sie sich um den Winkel  $\alpha = 365/360 \approx 1$  Grad weiter. Da der Tag  $24 \times 60$  Minuten hat, dreht sich die Erde in 4 Minuten um ein Grad weiter. Der Sonnentag (24h) ist somit im Mittel um rund 4 Minuten länger als der siderische Tag (23h56m). Eine genauere Rechnung ergibt als Differenz 3 Minuten und 56,6 Sekunden.

## 1.2 Monate

Für den Monat gibt es gleich mehrere Definitionen (die Zahlenangaben beziehen sich auf die Bewegung des Mondes am 1. Januar des Jahres 2000):

1. *Synodischer Monat*: Der synodische Monat ist die Zeitspanne zwischen zwei gleichen Stellungen des Monds relativ zu Erde und Sonne, also beispielsweise die Zeitspanne zwischen zwei Vollmonden oder Neumonden. Bei Vollmond spricht man auch von Opposition, bei Neumond von Konjunktion. Dieser Monat ist am längsten, da sich im Verlauf eines Monats das Erde-Mond-System weiter um die Sonne bewegt hat. Ein synodischer Monat dauert rund 29,5306 Tage oder 29 Tage, 12 Stunden, 44 Minuten und 3 Sekunden.
2. *Siderischer Monat*: Beim siderischen Monat hat sich der Mond relativ zum Fixsternhimmel einmal um die Erde gedreht. Der siderische Monat ist deutlich kürzer als der synodische Monat: In einem Monat dreht sich das Erde-Mond-System um etwas weniger als 30 Grad um die Sonne. Das Verhältnis von synodischem zu siderischem Monat ist somit ungefähr  $(360 + 30)/360 = 13/12$ . Genauer erhält man für den siderischen Monat 27,3217 Tage oder 27 Tage, 7 Stunden, 43 Minuten und 12 Sekunden.

3. *Tropischer Monat*: Beim tropischen Monat bezieht sich „einmal rum“ nicht auf den Fixsternhimmel sondern auf den Frühlingspunkt der Erde. Wegen der Präzession - der langsamsten Drehung der Rotationsachse der Erde - verschiebt sich dieser Frühlingspunkt im Vergleich zum Fixsternhimmel um rund 7 Sekunden im Monat. Um diese 7 Sekunden ist ein tropischer Monat kürzer als ein siderischer Monat.
4. *Anomalistischer Monat*: Die Mondbahn um die Erde (genauer um den gemeinsamen Schwerpunkt) ist eine Ellipse. Bei einem idealen gravitativen Zwei-Körper-Problem (ohne relativistische Korrekturen) wäre diese Ellipse stabil, d.h., das Perigäum (der erdnächste Punkt dieser Bahn) wäre relativ zum Fixsternhimmel immer derselbe. Durch verschiedene Störfaktoren (insbesondere den Einfluss der Sonne aber auch relativistische Korrekturen) verschiebt sich dieser Punkt jedoch im Laufe der Zeit relativ zum Fixsternhimmel. Als anomalistischen Monat bezeichnet man die Zeitspanne zwischen zwei aufeinanderfolgenden Durchläufen des Mondes durch das Perigäum. Diese Definition bezieht sich somit ausschließlich auf die Bahnperiode der Mondbahn um die Erde und bedarf keines äußeren Fixpunkts. Ein anomalistischer Monat dauert 27,55455 Tage oder 27 Tage, 13 Stunden, 18 Minuten und 33 Sekunden.
5. *Drakonitischer Monat*: Die Mondbahn liegt in einer Bahnebene, die relativ zur Ekliptik (also der Bahnebene der Erde um die Sonne) um ungefähr 5 Grad geneigt ist. Als Mondknoten bezeichnet man die beiden Punkte der Mondbahn, die in der Ebene der Ekliptik liegen. Die Zeitspanne zwischen zwei Durchgängen (von Süd nach Nord) des Mondes durch einen Mondknoten bezeichnet man als drakonitischen Monat. Eine Sonnen- bzw. Mondfinsternis kann nur dann von einem Punkt der Erde aus beobachtet werden, wenn dieser Punkt auf der Erde, der Mond und die Sonne auf einer Linie liegen. Dazu muss sich der Mond in diesem Augenblick in der Nähe eines Mondknoten befinden, da sonst der Mond ober- bzw. unterhalb der Sonne (bzw. des Erdschattens der Sonne bei einer Mondfinsternis) vorbeizieht. Der drakonitische Monat ist somit für die Berechnung von Sonnen- und Mondfinsternissen von Bedeutung. Ein drakonitischer Monat dauert 27,21222 Tage oder 27 Tage, 5 Stunden, 5 Minuten und 36 Sekunden.

### 1.3 Jahre

Auch beim Jahr kann man wieder mehrere Definitionen unterscheiden (auch hier beziehen sich die Zahlenangaben auf die Bewegung der Erde um die Sonne am 1. Januar 2000; die Dauer eines Jahres ist aus dieser Bewegung rechnerisch extrapoliert):

1. *Siderisches Jahr*: Ein siderisches Jahr bezeichnet die Zeitspanne, in der sich die Erde relativ zum Fixsternhimmel einmal um die Sonne bewegt hat. Es dauert 365 Tage, 6 Stunden, 9 Minuten und 9,54 Sekunden oder 365,2563604167 Tage.
2. *Tropisches Jahr*: Für das tropische Jahr gibt es zwei Definitionen. Die ältere Definition bezieht sich auf den Durchgang der Erde durch den Frühlingspunkt. Der Frühlingspunkt ist dabei der Zeitpunkt, bei dem die Sonne vom Erdmittelpunkt aus betrachtet genau über dem Äquator steht.<sup>1</sup> Zu diesem Zeitpunkt sind an einem idealisierten Tag die Sonnenstunden (der helle Tag) und die Nachtstunden gleich lang. Daher spricht man auch von der Tagundnachtgleiche bzw. dem Äquinoktium. Eine zweite Interpretation (dieser ersten Definition) ist: Zu diesem Zeitpunkt befindet sich die Erde in einem der beiden Schnittpunkte ihrer Bahnebene (der Ekliptik) mit ihrer Äquatorebene, d.h. die Rotationsachse der Erde steht senkrecht zur Verbindungsline

---

<sup>1</sup>Es gibt zwei solche Punkte im Jahr: der Frühlingspunkt und der Herbstpunkt. Man sollte daher spezifizieren, dass die Sonne vom Erdmittelpunkt aus betrachtet in diesem Augenblick den Äquator von Süd nach Nord durchläuft.

Erde-Sonne. Das so definierte tropische Jahr kann sich in unterschiedlichen Jahren um mehrere Minuten (bis zu einer Viertelstunde) unterscheiden. Die Einflüsse anderer Planeten auf die Präzession der Erde oder auch die Tatsache, dass der Frühlingspunkt (wegen der Präzession) immer an einem anderen Punkt der elliptischen Bahn der Erde ist, spielen hier eine wesentliche Rolle.

Die Internationale Astronomische Union (IAU) hat daher 1955 beschlossen, diese zwar sehr anschauliche aber durch keine präzisen Wert angebbare Definition der Länge eines tropischen Jahres durch eine zweite Definition zu ersetzen. Danach bestimmt man die Länge eines tropischen Jahres in Bezug auf einen bestimmten Augenblick. In diesem Augenblick wird die Winkelgeschwindigkeit einer mittleren Sonne - die periodische Schwankung der Winkelgeschwindigkeit der Sonne aufgrund der elliptischen Bahn der Erde wird hierbei durch den Bezug auf eine mittlere Sonne ausgeglichen - bestimmt und extrapoliert, wie lange es dauert, bis bei dieser Winkelgeschwindigkeit  $360^\circ$  zurückgelegt werden. Dies bezeichnet man dann als momentanes tropisches Jahr. Diese Winkelgeschwindigkeit der mittleren Sonne wird auf die Drehachse der Erde, ein sogenanntes „mittleres Äquinoktium des Datums“, bezogen. Diese Definition ist zwar unanschaulich, hat aber den Vorteil, dass man einem tropischen Jahr zu jedem Augenblick einen präzisen Wert zuordnen kann.

Am 1. Januar 2000 dauerte ein tropisches Jahr nach dieser Definition 365,24219052 SI-Tage.

3. *Anomalistisches Jahr*: Ähnlich wie beim anomalistischen Monat bezeichnet ein anomalistisches Jahr die Zeitspanne zwischen zwei Periheldurchgängen der Erde, wobei das Perihel der sonnennächste Punkt der Erdbahn ist. Aufgrund des Einflusses anderer Planeten sowie relativistischer Korrekturen verschiebt sich das Perihel jedes Jahres um ungefähr 5 Minuten relativ zum siderischen Jahr. Das anomalistische Jahr (1. Januar 2000) dauert 365,259635864 Tage oder 365 Tage, 6 Stunden, 13 Minuten und 52,54 Sekunden.

## 2 Der Menton-Zyklus und die Struktur von Mondkalendern

Viele antike Kalendersysteme sind Mondkalender, d.h. bei ihnen ist der Monat die natürliche Einheit und zur ungefähren Anpassung an die Jahreslänge wurden gelegentlich zusätzliche Tage oder gar zusätzliche Monate eingefügt. Schon im antiken Babylon war bekannt, dass 19 Jahre ziemlich genau 235 Monaten entsprechen. Wählt man die Julianische Schaltjahrregelung, nach der ein Jahr 365,25 Tage hat, entsprechen 19 Jahre 6939,75 Tage. Andererseits entsprechen 235 synodische Monate 6939,691 Tagen. Auf 19 Jahre ein Fehler von 0,059 Tagen (oder 1 Stunde, 24 Minuten und 58 Sekunden) bedeutet, dass sich in  $19 \times 1/0,059 = 322$  (Julianischen) Jahren die Beziehung zwischen Mondphasen und Sonnenjahr um einen Tag verschiebt. Definiert man andererseits die Länge eines Jahres als  $1/19$ .tel von 235 Monaten (das entspricht 365,2469 Tagen), so liegt diese Zeitspanne zwischen dem tropischen Jahr und dem Jahr nach dem Julianischen Kalender. Die Dauer von 235 (synodischen) Monaten bezeichnet man auch als Menton-Zyklus.

Im Folgenden geht es meist nur um die größten Regeln eines Kalenders, sodass Jahre in Monate und Monate in Tage unterteilt werden können. Die meisten Kalender enthalten weitere Ausnahmeregelungen, auf die hier nicht eingegangen wird.

### 2.1 Der Jüdische Kalender

Der Jüdische Kalender ist ein reiner Mondkalender. Monate mit 29 und 30 Tagen wechseln sich im Wesentlichen ab. Ein Jahr besteht aus 12 Monaten. Damit hat ein Jahr rund 354 Tage. Da dies etwas

zu kurz ist, werden gelegentlich Schaltmonate mit meist 30 Tagen eingefügt. Insgesamt richtet sich diese Einteilung nach dem Merton-Zyklus, d.h. 19 Jahre bestehen aus 235 Monaten. Da 19 Jahre mit 12 Monaten nur 228 Monaten entsprechen, werden in den 19 Jahren insgesamt 7 Monate als Schaltmonate eingefügt, und zwar in den Jahren 3, 6, 8, 11, 14, 17 und 19. Auf diese Weise erreicht man, dass der Jahresanfang mehr oder weniger gleich bleibt; im Jüdischen Kalender im September oder Oktober.

## 2.2 Der Islamische Kalender

Es gibt mehrere verschiedene islamische Kalender, aber der Kalender, nachdem sich auch heute noch die Festtage (oder beispielsweise der Beginn des Monats Ramadan) bestimmen, umfasst 12 Monate mit jeweils 29 oder 30 Tagen. Während im altarabischen Kalender alle zwei oder drei Jahre ein Schaltmonat eingefügt wurde (dieser Kalender also dem Jüdischen Kalender ähnelte), wurde im Islam dieser Schaltmonat abgeschafft. Ein Jahr hat nun also rund 354 Tage. Damit verschieben sich der Jahresanfang und auch die wichtigsten Feiertage jährlich um rund 10-12 Tage nach vorne und wandern im Verlauf der Zeit durch das ganze Jahr.

## 3 Die Wochentage

Schon in der Schöpfungsgeschichte (Genesis) des alten Testaments, die vermutlich auf das 5. bis 6. Jahrhundert vor Christus zurückgeht, ist davon die Rede, dass Gott die Welt in sechs Tagen erschuf und am siebten Tage ruhte. Die Zeiteinheit „Woche“ als sieben Tage war schon in Babylon in Gebrauch und vermutlich hat die jüdische Tradition diese Einheit während des babylonischen Exils übernommen.

Die Zahl sieben für die Anzahl der Tage in einer Woche geht vermutlich auf astronomische Beobachtungen zurück: Die damals bekannten sieben beweglichen Himmelskörper waren (in aufsteigender Reihenfolge ihrer Umlaufzeiten): Mond (1 Monat), Merkur (3 Monate), Venus (7 Monate), Sonne (1 Jahr), Mars (2 Jahre), Jupiter (12 Jahre) und Saturn (30 Jahre). Wie man heute noch an den Bezeichnungen in einigen europäischen Sprachen ablesen kann, wurden die Wochentage ursprünglich nach diesen sieben Himmelskörpern benannt (Tab. 2)

Deutsch	Englisch	Französisch	Lateinisch	Himmelskörper
Sonntag	Sunday	dimanche	Solis dies	Sonne
Montag	Monday	lundi	Lunae dies	Mond
Dienstag	Tuesday	mardi	Martis dies	Mars
Mittwoch	Wednesday	mercredi	Mercurii dies	Merkur
Donnerstag	Thursday	jeudi	Iovis dies	Jupiter
Freitag	Friday	vendredi	Veneris dies	Venus
Samstag	Saturday	samedi	Saturni dies	Saturn

Tabelle 2: Die Wochentage in verschiedenen europäischen Sprachen und die zugehörigen Planeten.

Auch in den germanischen Sprachen sind diese Ursprünge teilweise erkennbar. Der Donnerstag ist der Tag des Donnergottes Thor (im Englischen Thursday erkennbar), der wiederum dem römischen Gott Jupiter entsprach. Und der Freitag ist vermutlich der Freyastag, der Tag der Götting Freya, die wiederum der römischen Gottheit Venus entsprach. Hier gibt es aber unterschiedliche Theorien.

## 4 Die Kalenderreform von 1582

Einen relativ genauen Sonnenkalender hat Julius Cäsar um 45 v. Chr. eingeführt: Er sah vor, dass ein Jahr 365 Tage hat und dass alle vier Jahre ein sogenanntes Schaltjahr eingefügt wird, bei dem ein Jahr 366 Tage hat. Der zusätzliche Tag ist der 29. Februar. Diese Schaltjahrregelung war schon vorher in Ägypten in Gebrauch, und Cäsar hat sie für das römische Reich übernommen und angepasst.

Damit ergibt sich für die Jahreslänge im Julianischen Kalender

$$(4 \cdot 365 + 1)/4 = 365,25 \text{ Tage.} \quad (1)$$

Diese Formel folgt aus folgender Überlegung: Ein Zeitraum von 4 Jahren bildet eine Periode des Julianischen Kalenders, d.h. nach vier Jahren wiederholt sich das Schema. Diese vier Jahre haben  $4 \times 365$  Tage plus einen weiteren Tag wegen des Schaltjahrs. Teilt man diese Anzahl von Tagen wieder durch die Anzahl der Jahre einer Periode (also 4) so erhält man die mittlere Dauer eines Jahres in Tagen.

Unsere Jahreszeiten werden durch das tropische Jahr bestimmt, d.h. durch den Zeitraum zwischen zwei aufeinanderfolgenden Frühlingspunkten. Ein tropisches Jahr dauert 365,24219 (Sonnen)Tage (genau gilt dies für ein von der Bewegung der Sonne am 1. Januar 2000 extrapoliertes Jahr). Die Differenz zum Julianischen Kalender sind 0,00781 Tage (oder 11 Minuten und 14,8 Sekunden) pro tropischem Jahr, bzw. alle 128 Jahre verschiebt sich im Julianischen Kalender der Frühlingspunkt um einen Tag (nach vorne). Diese Differenz war bereits im 9. Jahrhundert bekannt.<sup>2</sup> Da der Frühlingspunkt, d.h. der Beginn des Frühjahrs, im christlichen Jahr eine besondere Bedeutung hat - Ostern (und damit viele beweglichen Feiertage im Kirchenjahr) bestimmt sich aus dem ersten Vollmond im Frühjahr - wollte man natürlich nicht, dass beispielsweise Ostern irgendwann im Herbst stattfindet und Weihnachten im Hochsommer. Daher gab Papst Gregor XIII. die Ausarbeitung einer Kalenderreform in Auftrag, die er dann im Jahre 1582 per päpstlichem Dekret verkündete.

Die Kalenderreform bestand aus zwei Anteilen:

1. Die 10 Tage zwischen dem 4. Oktober und dem 15. Oktober 1582 wurden ausgelassen, d.h. Donnerstag, dem 4. Oktober 1582, folgte Freitag, der 15. Oktober 1582. Die dazwischenliegenden Tage gibt es im Gregorianischen Kalender nicht. Damit wurde der Frühlingsanfang wieder auf den 21. März gelegt, wie im Jahr 325, als auf dem Konzil von Nicäa die Bestimmung des Osterfestes in Abhängigkeit vom Frühlingsanfang festgelegt wurde.
2. Die Schaltjahrregelung wurde leicht abgeändert: Alle vier Jahre (in den Jahren, die glatt durch 4 teilbar sind) gibt es ein Schaltjahr mit 366 Tagen, allerdings gibt es alle hundert Jahre (in den Jahren, die glatt durch 100 teilbar sind) kein Schaltjahr, jedoch gibt es alle 400 Jahre (in den Jahren, die glatt durch 400 teilbar sind) trotzdem ein Schaltjahr. Somit gab es im Jahr 1600 und 2000 ein Schaltjahr, in den Jahren 1700, 1800 und 1900 jedoch nicht.

Damit ergibt sich für die Jahreslänge im Gregorianischen Kalender:

$$(400 \cdot 365 + 100 - 4 + 1)/400 = 365,2425 \text{ Tage.} \quad (2)$$

Hier gilt wieder: Eine Periode des Gregorianischen Kalenders dauert 400 Jahre, in diesen 400 Jahren finden  $100 - 4 + 1 = 97$  Schaltjahre mit 366 Tagen statt, alle anderen Jahre haben 365 Tage. Die Differenz zwischen gregorianischem Jahr und tropischem Jahr beträgt nun 0,00031 Tage (oder 26,8

<sup>2</sup>Bereits Mitte des 8. Jahrhunderts wusste man, dass sich die Mondphasen relativ zu den theoretischen Überlegungen (nach denen sich die Mondphasen alle 235 Jahre wiederholen - dies wusste man schon im alten Babylonien) verschoben hatten.

Sekunden). Das heißt, alle 3226 Jahre verschiebt sich dieser Kalender relativ zum tropischen Jahr (des Jahres 2000) um einen Tag. Damit kann man zunächst einmal leben. Weitere Kalenderreformen sind in die ferne Zukunft verschoben.

Vom Kalendersystem unabhängig ist die Festlegung des Jahresbeginns. Obwohl schon der Kalender von Julius Caesar den Jahresbeginn auf den 1. Januar festgelegt hatte, waren unterschiedliche Konventionen in Gebrauch. Erst im 16. und 17. Jahrhundert setzte sich der 1. Januar als Jahresbeginn weitgehend durch.

## 5 Kuriositäten der Kalenderreform

Die Kalenderreform von Papst Gregor XIII. wurde nicht überall gleich angenommen. Selbst katholische Gebiete sträubten sich teilweise, und protestantische bzw. reformierte Gebiete verweigerten sich der Reform schon alleine deshalb, weil sie vom Papst ausging. So wurde der Kalender in England erst 1752 übernommen (zu diesem Zeitpunkt mussten bereits 11 Tage ausfallen), in Sowjetrussland erst 1918 (hier wurden 13 Tage gestrichen) und beispielsweise in Griechenland erst 1923. In manchen orthodoxen Ostkirchen gilt heute noch der Julianische Kalender. Mit diesen unterschiedlichen Daten der Übernahme des Gregorianischen Kalenders sind einige Kuroositäten verbunden.

1. William Shakespeare und Miguel de Cervantes starben offiziell beide am 23. April 1616. Allerdings liegt der Todestag der beiden Dichter 10 Tage auseinander: In England verwendete man Anfang des 17. Jahrhunderts noch den Julianischen Kalender, in Spanien bereits den Gregorianischen. Das bedeutet, Cervantes starb tatsächlich 10 Tage vor Shakespeare.
2. Für die Lebensdaten von Newton findet man zwei Versionen: 25. 12. 1642 bis 20. 3. 1726 und 4. 01. 1643 bis 31. 03. 1727. Der Unterschied im Geburtsdatum ist offensichtlich: Das eine Geburtsdatum (Dezember 1642) bezieht sich auf den Julianischen Kalender, der zu diesem Zeitpunkt in England noch gültig war, das andere Datum (Januar 1643) auf den Gregorianischen Kalender. Überraschend am Todesdatum ist weniger der Tag (20.3. bzw. 31.3.), der sich durch die Differenz zwischen Julianischem und Gregorianischem Kalender ergibt, als vielmehr das Jahr: 1726 bzw. 1727. In England war Neujahr 1727 (der Tag es Jahreswechsels) der 25. März und nicht der 1. Januar (diese Konvention bezeichnet man als *Annuntiationsstil*, weil der 25. März als die Verkündigung (Annunciation) der Empfängnis von Maria galt). Das bedeutet, nach dem damals in England gültigen Kalender starb Newton noch vor dem Jahreswechsel (also im Jahr 1726), nach dem Gregorianischen Kalender nach dem Jahreswechsel (also im Jahr 1727).
3. Ebenfalls kurios ist Newtons Beerdigung, die nach dem Julianischen Kalender am 28. März 1727 stattfand. Das bedeutet, Newton ist nach dem Julianischen Kalender am 20. März 1726 gestorben und am 28. März 1727 beerdigt worden. Das klingt zunächst so, als ob zwischen dem Sterbedatum und der Beerdigung über ein Jahr liegt. Doch wie schon erwähnt fand in England damals der Jahreswechsel am 25. März statt.
4. Die sogenannte Oktoberrevolution in Russland fand nach dem Julianischen Kalender am 25. Oktober 1917 statt, nach dem Gregorianischen Kalender am 7. November 1917. Im Jahre 1917 galt in Russland jedoch noch der Julianische Kalender, daher „Oktober-“Revolution. Nach dem heutigen in Russland (bzw. der früheren Sowjetunion) gültigen Kalender handelte es sich eigentlich um eine Novemberrevolution. In der Sowjetunion wurde der Jahrestag der Oktoberrevolution immer am 7. November gefeiert.

# Die Zeitgleichung

Thomas Filk, Universität Freiburg

Die sogenannten *wahren Sonnentage*, hier ist die Zeitspanne von Sonnenhöchststand bis zum nächsten Sonnenhöchststand gemeint, sind nicht immer gleich lang. Dafür sind in erster Linie zwei Einflüsse verantwortlich: die elliptische Form der Erdbahn und die Neigung der Erdachse gegen die Ekliptik, also die Ebene, in der die Erdbahn um die Sonne verläuft. Dies führt zu einer Differenz zwischen der wahren Sonnenzeit (bei der die Sonne um 12 Uhr ihren Höchststand in Richtung Süden erreicht) und der sogenannten mittleren Sonnenzeit, wie sie auf einer gleichmäßig gehenden Uhr angezeigt wird.<sup>1</sup> Diese Differenz, die sogenannte Zeitgleichung, war schon im Altertum bekannt. Die Zeitgleichung  $ZG(t)$  gibt an, wie viele Minuten man zur gemessenen wahren Ortszeit addieren muss, um die gemittelte Ortszeit zu erhalten:

$$ZG(t) = \text{Mittlere Ortszeit}(t) - \text{Wahre Ortszeit}(t). \quad (1)$$

Sie konnte auf der Basis der Kepler'schen Gesetze Anfang des 16. Jahrhunderts auch theoretisch begründet werden.

In Tabelle 1 sind die wichtigsten Größen zusammengefasst, die in diesem Kapitel benötigt werden.

große Halbachse Erde-Sonne	$a = 149\,598\,023 \text{ km}$
Datum des Perihels	derzeitig am $(3 \pm 2)$ . Januar
Neigung der Erdachse zur Ekliptik	$\varepsilon = 23,44^\circ$
numerische Exzentrizität der Erdbahn	$\epsilon = 0,0167$
Abstand Erde-Sonne im Perihel	$r_{\min} = 147,10 \cdot 10^6 \text{ km}$
Abstand Erde-Sonne im Aphel	$r_{\max} = 152,10 \cdot 10^6 \text{ km}$
Dauer eines tropischen Jahres	365,24219 Tage

Tabelle 1: Die wichtigsten physikalischen Größen im Zusammenhang mit der Zeitgleichung. Je nach Zusammenhang und erforderlicher Genauigkeit werden die Werte unterschiedlich gerundet.

## 1 Zeitsysteme

Wie schon erwähnt, entspricht der Sonnenstand, die sogenannte wahre Sonnenzeit, im Allgemeinen nicht dem, was man als Ortszeit bezeichnet. Zum einen hat es sich als sinnvoll erwiesen, sogenannte Zeitzonen zu definieren, innerhalb deren die Zeitsysteme gleich sind, zum anderen ist der scheinbare Gang der Sonne am Himmel nicht vollkommen gleichförmig.

### 1.1 Zeitzonen

Die wahre Sonnenzeit an einem bestimmten Ort auf der Erde ist definiert durch den Sonnenhöchststand, der den Zeitpunkt 12 Uhr festlegt. Auf der nördlichen Halbkugel (nördlich des  $23,5^\circ$

<sup>1</sup>Hinzu kommt noch, dass die Ortszeit heute nicht mehr die wahre Ortszeit ist, sondern einer Zeitzone zugeteilt wurde. Je nach Längengrad, an dem sich ein Ort befindet, muss zur lokalen Zeit entsprechend der Zeitzone noch ein Korrekturterm hinzugefügt werden werden. Diese Korrektur wird in Abschnitt 1.2 kurz erwähnt und ansonsten nicht berücksichtigt.

Breitengrads) befindet sich die Sonne in diesem Augenblick genau im Süden, auf der südlichen Halbkugel (südlich des  $-23,5$ . Breitengrads) ist sie im Norden.<sup>2</sup> Etwas anders ausgedrückt handelt es sich bei 12 Uhr mittags (wahre Sonnenzeit) um den Augenblick, an dem die Verbindungsline Erdmittelpunkt-Sonnenmittelpunkt den Längengrad des jeweiligen Orts schneidet. Ein wahrer Sonnentag ist dann der Zeitraum von einem Sonnenhöchststand zum nächsten Sonnenhöchststand.

Die wahre Uhrzeit (bezüglich der Sonne) hängt vom Längengrad eines Orts ab. Aus diesem Grund ist eine solche Zeitrechnung sehr kompliziert, da sie lokal von Ort zu Ort verschieden sein kann. Bis zum Ende des 19. Jahrhunderts war das tatsächlich der Fall: Jede größere Stadt hatte ihre eigene Ortszeit. Insbesondere als Folge des zunehmenden Eisenbahnverkehrs wurde ein solches System jedoch zu umständlich, und so wurde per Erlass im Jahre 1893 in Deutschland als gesetzliche Zeit die Zeit am 15. Längengrad eingeführt. Diese Zeit bezeichnet man allgemein auch als „Mitteleuropäische Zeit (MEZ)“, da sie in den meisten europäischen Ländern gilt.

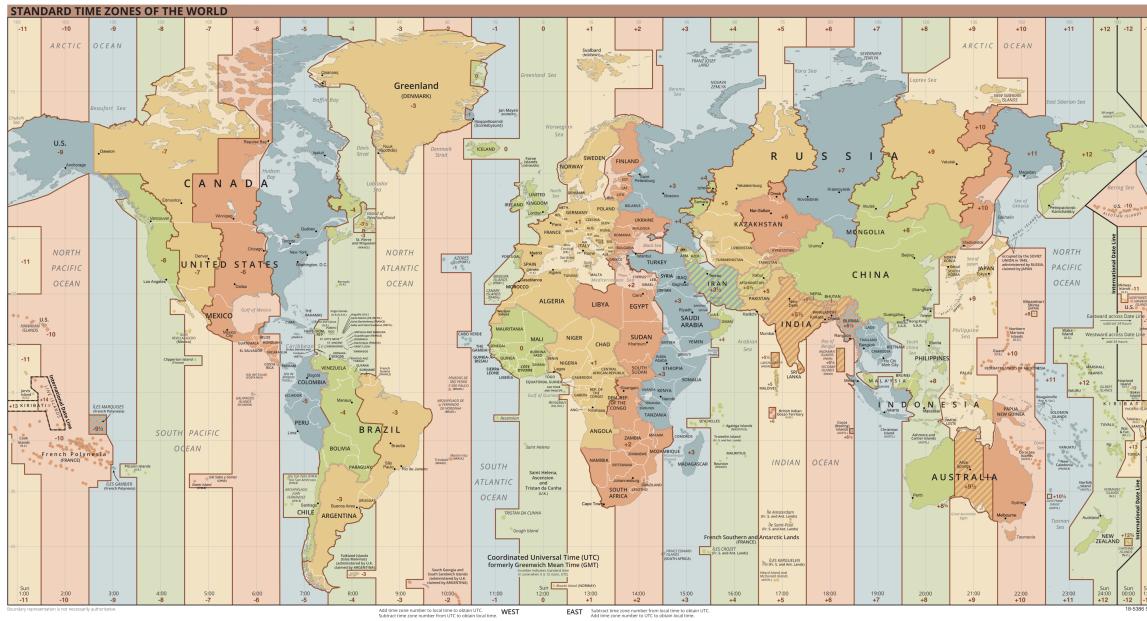


Abbildung 1: Die Zeitzonen der verschiedenen Länder. (aus [7])

Die Erdkugel ist in 24 Zeitzonen eingeteilt (Abb. 1), jede Zeitzone umfasst somit  $360/24 = 15$  Längengrade. Befindet sich ein Ort nicht auf dem Längengrad seiner Zeitzone, muss man pro Längengraddifferenz 4 Minuten addieren bzw. subtrahieren, je nachdem ob sich der Ort westlich oder östlich von dem Längengrad seiner Zeitzone befindet. Die Mitteleuropäische Zeit, die in Deutschland und den meisten westeuropäischen Kontinentalgebieten (außer Portugal) gültig ist, hat ihren Längengrad bei  $15^\circ$  Ost. Das ist nahe des östlichsten Punkts der deutschen Grenze. Selbst Frankfurt an der Oder liegt noch etwas weiter westlich. Freiburg liegt beim Längengrad  $7,85^\circ$ , also  $7,15^\circ$  westlich vom  $15$ . Längengrad. Dementsprechen muss man zur MEZ rund 28,6 Minuten addieren, um die wahre Ortszeit zu erhalten (in der Sommerzeit nochmals eine Stunde mehr). Der Sonnenhöchstand im Sommer in Freiburg ist gegen 13:30.

<sup>2</sup>Zwischen diesen Breitengräden kann die Sonne zu diesem Zeitpunkt je nach Jahreszeit im Norden oder Süden stehen. In jedem Fall entspricht ihre Lage einem Extrempunkt bezüglich des Winkels über dem Horizont.

## 1.2 Wahrer und mittlerer Sonnentag

Unabhängig von den bisher erwähnten Korrekturen, die zur Bestimmung der lokalen Ortszeit notwendig sind, kann der wahre Sonnentag, also die Zeitdauer zwischen zwei Sonnenhöchstständen an einem Ort, im Laufe eines Jahres um bis zu fast einer halben Minute schwanken, was sich zu manchen Zeiten kumulativ bis zu einer Viertelstunde aufaddieren kann. Die beiden Hauptursachen dafür - die elliptische Bahn der Erde und die Neigung der Erdachse relativ zur Ekliptik - werden in den Abschnitten 3 und 4 besprochen. Aus diesem Grund definiert man einen mittleren Sonnentag als die Dauer eines Tages, sodass ein Jahr genau 365,2422 mittlere Sonnentage hat. Dies hat jedoch zur Folge, dass eine Uhr, welche die Uhrzeit nach dem mittleren Sonnentag anzeigt, nicht mit einer Uhr übereinstimmt, welche die wahre Sonnenzeit anzeigt, also z.B. eine Sonnenuhr. Manche Sonnenuhren berücksichtigen diese Effekte, indem die Stundenlinien die Form eines sogenannten Analemmas haben (Abschnitt 6.2).

## 2 Zur Mathematik einer elliptischen Bahnkurve

Es gibt zwei bekannte Definitionen einer Ellipse: (1) als die Menge aller Punkte, bei denen die Summe der Abstände zu zwei gegebenen Punkten, den Brennpunkten der Ellipse, konstant ist, und (2) als Schnitt eines Kegelmantels mit einer Ebene. Dass die Lösungen des Zwei-Körper-Kepler-Problems (also das Zwei-Körper-Problem mit einem Potenzial proportional zu  $1/r$ ) Kegelschnitte sind, ist ebenfalls bekannt. Die gebundenen Lösungen bilden dabei Ellipsen, wobei der Kreis ein Spezialfall einer Ellipse ist.

Die sogenannte *lineare Exzentrizität*  $e$  einer Ellipse ist gleich dem Abstand vom Mittelpunkt der Ellipse zu einem der Brennpunkte. Sie hat die Dimension einer Länge. Die *numerische Exzentrizität*  $\epsilon$  ist gleich dem Verhältnis von  $e$  zur großen Halbachse  $a$  und ist somit dimensionslos (vgl. Abb. 2, links):

$$\epsilon = \frac{e}{a}. \quad (2)$$

Außerdem gilt

$$2a = r_{\max} + r_{\min} \quad \text{und} \quad 2e = r_{\max} - r_{\min}, \quad (3)$$

wobei  $r_{\max}$  der maximale Abstand zwischen einem Brennpunkt und der Ellipse ist (also der Abstand zwischen einem Brennpunkt und dem gegenüberliegenden Punkt der Ellipse) und  $r_{\min}$  der minimale Abstand. Bei der Erdumlaufbahn um die Sonne bezeichnet man den Punkt mit dem maximalen Abstand zur Sonne als Aphel, den Punkt mit dem minimalen Abstand als Perihel.<sup>3</sup>

Aus den Gleichungen 2 und 3 folgt

$$\epsilon = \frac{r_{\max} - r_{\min}}{r_{\max} + r_{\min}} \quad \text{oder} \quad \frac{r_{\min}}{r_{\max}} = \frac{1 - \epsilon}{1 + \epsilon}. \quad (4)$$

Nach dem zweiten Kepler'schen Gesetz („In gleichen Zeiten werden gleiche Flächen überstrichen“) folgt, dass sich die Geschwindigkeiten im Perihel (dem sonnennächsten Punkt),  $v_{\max}$ , zur Geschwindigkeit im Aphel (dem sonnenfernsten Punkt),  $v_{\min}$ , wie die Abstände der Bahn zur Sonne verhalten (siehe Abb. 2, rechts):

$$\frac{v_{\min}}{v_{\max}} = \frac{r_{\min}}{r_{\max}}. \quad (5)$$

Das zweite Kepler'sche Gesetz ist äquivalent zur Drehimpulserhaltung. ([Herleitung](#))

---

<sup>3</sup>Bei der Mondbahn um die Erde bezeichnet man den Punkt der Mondbahn mit dem größten Abstand zur Erde als Apogäum und den Punkt mit dem kleinsten Abstand als Perigäum.

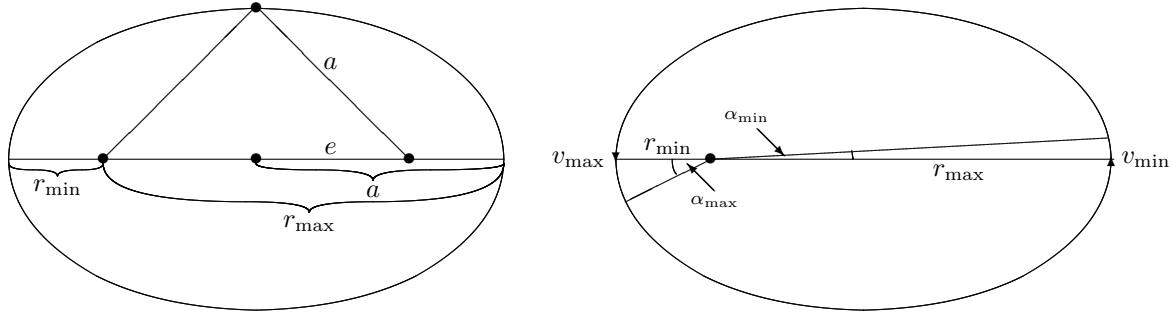


Abbildung 2: (links) Die geometrischen Größen in einer Ellipse:  $a$  (große Halbachse),  $e$  (Exzentrizität),  $r_{\min}$  und  $r_{\max}$  (kleinster und größter Abstand der Ellipse zu einem Brennpunkt); (rechts) die dynamischen Größen in einer Ellipse:  $v_{\min}$  und  $v_{\max}$  (kleinste und größte Geschwindigkeit eines umlaufenden Objekts),  $\alpha_{\max}$  und  $\alpha_{\min}$  (in gleichen Zeiten überstrichene Winkel im Perihel und Aphel).

Für die Winkel im Perihel  $\alpha_{\max}$  und Aphel  $\alpha_{\min}$ , die in gleichen Zeiten überstrichen werden, bedeutet dies:

$$\frac{\alpha_{\min}}{\alpha_{\max}} = \left( \frac{v_{\min}}{r_{\max}} \right) / \left( \frac{v_{\max}}{r_{\min}} \right) = \left( \frac{r_{\min}}{r_{\max}} \right)^2. \quad (6)$$

### 3 Der Einfluss der elliptischen Bahnkurve

In diesem Abschnitt vernachlässigen wir die Neigung der Erdachse zur Ekliptik (diesen Effekt berücksichtigen wir im nächsten Abschnitt) und nehmen an, die Rotationsachse der Erde sei senkrecht zu ihrer Bahnebene.

Würde sich die Erde auf einer Kreisbahn um die Sonne bewegen, wäre der Winkel, den sie auf ihrer Bahn um die Sonne täglich überstreicht, immer derselbe. Wir bezeichnen diesen Winkel mit  $\alpha_0$ . Sein Wert berechnet sich aus der Tatsache, dass ein Jahr rund 365,2422 Tage dauert und in dieser Zeit ein Winkel von  $360^\circ$  überstrichen wird:

$$\alpha_0 = \frac{360^\circ}{365,2422 \text{ d}} = 0,98565^\circ/\text{d}. \quad (7)$$

Die Erde dreht sich in 24 Stunden einmal um ihre Achse von Sonnenhöilstand zu Sonnenhöchstand. Dabei überstreicht sie einen Winkel von  $360^\circ + \alpha$ . Der zusätzliche Winkel  $\alpha$  röhrt daher, dass sich die Erde im Verlauf eines Tages auf ihrer Bahn weiter um die Sonne bewegt hat, und sich somit die Sonne relativ zur Erde um den Winkel  $\alpha$  weiterbewegt hat. Auf einer Kreisbahn wäre dieser Winkel  $\alpha = \alpha_0$ . Um sich um diesen Winkel  $\alpha$  zu drehen, benötigt sie ungefähr 4 Minuten.<sup>4</sup> Um diese Zeit ist der ideale Sonnentag länger als der siderische Tag (der sich auf eine Umdrehung relativ zum Fixsternhimmel bezieht).

Tatsächlich bewegt die Erde sich aber auf einer elliptischen Bahn und ihre Winkelgeschwindigkeit ist im Perihel größer als im Aphel. Das bedeutet, dass an einem Tag in Perihelnähe (dieser Tag liegt derzeit um den 3. Januar) ein größerer Winkel von ihr relativ zur Sonne überstrichen wird als an einem Tag im Aphel (rund 3. Juli). Für die Zeitabhängigkeit dieses Winkels können wir folgenden Ansatz machen

$$\alpha(t) = \alpha_0 \left( 1 + A \cos \left( \frac{2\pi}{T} t \right) \right), \quad (8)$$

<sup>4</sup>Da sich die Erde in 24 Stunden um rund  $360^\circ$  dreht, dreht sie sich in einer Stunde um  $15^\circ$  und somit in 4 Minuten um  $1^\circ$ .

wobei  $T$  die Periodendauer der Erdbahn um die Sonne ist - also ein Jahr - und die Zeit  $t$  am Perihel beginnen soll (also  $\alpha(t=0) = \alpha_{\max}$ ). Nach Gl. 6 besteht folgende Beziehung zwischen der Amplitude  $A$  und der Exzentrizität  $\epsilon$ :

$$\frac{\alpha_{\max}}{\alpha_{\min}} = \left( \frac{1+\epsilon}{1-\epsilon} \right)^2 \approx 1 + 4\epsilon. \quad (9)$$

Andererseits gilt nach Gl. 8:

$$\frac{\alpha_{\max}}{\alpha_{\min}} = \frac{\alpha(0)}{\alpha(T/2)} = \frac{\alpha_0(1+A)}{\alpha_0(1-A)} \approx 1 + 2A. \quad (10)$$

Wir erhalten somit die Beziehung  $A = 2\epsilon$ .

Für die Winkelgeschwindigkeit der Eigendrehung der Erde können wir nach obiger Argumentation schreiben:

$$\omega = \frac{360^\circ + \alpha_0}{86400 \text{ s}}. \quad (11)$$

Andererseits ist die Zeitdauer  $T(\alpha)$ , die die Erde benötigt, um sich um den Winkel  $\alpha$  zu drehen, gegeben durch

$$T(\alpha) = \frac{\alpha}{\omega}, \quad (12)$$

und damit dauert ein Tag (in Abhängigkeit von der Jahreszeit  $t$ ):

$$T(t) = \frac{360^\circ + \alpha(t)}{\omega} = \left( \frac{360^\circ + \alpha_0(1 + 2\epsilon \cos \frac{2\pi}{T} t)}{360^\circ + \alpha_0} \right) T_0 \quad (13)$$

$$= T_0 + 2\alpha_0 \epsilon \frac{\cos \frac{2\pi}{T} t}{360^\circ + \alpha_0} T_0, \quad (14)$$

wobei  $T_0 = 86\,400 \text{ s}$  einem mittleren Sonnentag entspricht. Die Zeitdifferenz zwischen einem mittleren Sonnentag und dem tatsächlichen Sonnentag ist somit

$$\Delta T(t) = T(t) - T_0 = 2\alpha_0 \epsilon \frac{\cos \frac{2\pi}{T} t}{360^\circ + \alpha_0} T_0. \quad (15)$$

Setzen wir Werte ein, so erhalten wir für die Amplitude dieser Schwankung:

$$\Delta T = 2\alpha_0 \epsilon \frac{1}{360^\circ + \alpha_0} 86400 \text{ s} \approx 7,88 \text{ s}. \quad (16)$$

Um diesen Wert kann ein einzelner Tag also länger oder kürzer sein als ein mittlerer Sonnentag (sofern man nur die elliptische Bahn der Erde berücksichtigt). Ein solcher Effekt für sich genommen wäre vermutlich im Altertum noch nicht aufgefallen, wäre dieser Effekt nicht kumulativ, d.h., wenn viele aufeinanderfolgende Tage um ähnliche Werte zu lang oder zu kurz sind, addieren sich die Effekte auf. Je nach Jahreszeit geht die tatsächliche Zeit der gemittelten Zeit voraus bzw. nach. Relevant für die Zeitgleichung ist dieser kumulative Effekt, d.h. die Differenz zwischen der wahren Zeit, die an einem Ort am Stand der Sonne abgelesen wird, und der gemittelten Zeit, die eine Uhr anzeigt.

Bilden wir von Gl. 15 die Summe, erhalten wir:

$$ZG_1(t) = \sum_{t'=0}^t \Delta T(t') \approx \frac{T}{2\pi} \Delta T \sin \frac{2\pi}{T} t. \quad (17)$$

Hierbei verläuft die Summe über die Tage  $t$  vom Perihel gerechnet und  $T$  ist die Dauer eines Jahres in Tagen (also  $T = 365,2422$ ). Die Summe selbst wurde durch ein Integral approximiert. Die Amplitude dieser periodischen Funktion ist schon beträchtlich höher: Sie beträgt rund 7,63 Minuten. Dies ist ein Maß für den kumulativen Effekt: Im Verlauf eines Jahres kann die wahre Zeit an einem Ort der gemittelten Zeit um  $\pm 7,63$  Minuten vor bzw. nach gehen. Dieser Effekt ist Anfang April (drei Monate nach dem Perihel) und nochmals Anfang Oktober (sechs Monate später) am größten. Zunächst hinkt die wahre Zeit der gemittelten Zeit hinterher.

## 4 Die Neigung der Erdachse zur Ekliptik

Für die separate Beschreibung des zweiten Einflusses auf die Tageslänge nehmen wir an, dass sich die Erde gleichmäßig auf einer Kreisbahn um die Sonne bewegt. Wir berücksichtigen allerdings nun, dass die Rotationsachse der Erde um  $23,44^\circ$  relativ zur Normalen der Ekliptik (der Bahnebene der Erde) geneigt ist. Außerdem ist es sinnvoll, nicht ein heliozentrisches Koordinatensystem zu verwenden, sondern ein geozentrisches, allerdings soll dieses Koordinatensystem relativ zum Frühlingspunkt (also der Schnittlinie zwischen Äquatorebene und Ekliptik) fest sein (siehe Abb. 3). Die  $z$ -Achse dieses Systems sei die Rotationsachse der Erde, die  $x,y$ -Achsen liegen dann in der Äquatorebene (z.B. sei die  $x$ -Achse die Schnittlinie der Ekliptik und Äquatorebene, sie zeige also in Richtung Frühlingspunkt). Es handelt sich also nicht um ein erdfestes Koordinatensystem, das die tägliche Drehung der Erde mitmacht, sondern um ein Koordinatensystem, das bezüglich des Frühlingspunkts fest ist. Dies ist auch das Koordinatensystem, das man in der Astronomie häufig verwendet.

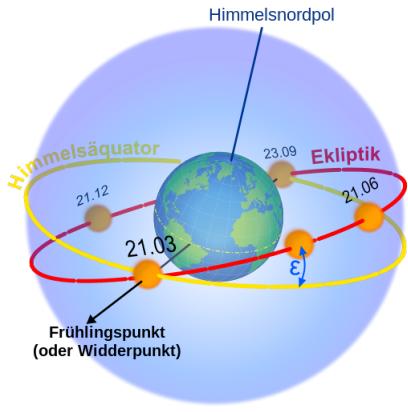


Abbildung 3: Äquatorebene und Ekliptik sind um  $23,44^\circ$  gegeneinander geneigt (hier mit  $\varepsilon$  bezeichnet). Eine gleichmäßige Bewegung der Sonne entlang der Ekliptik wird zu einer ungleichmäßigen Bewegung in Bezug auf die Längengrade.

In Bezug auf ein solches Koordinatensystem bewegt sich die Sonne im Verlauf eines Jahres gleichförmig auf einem Kreis, der um  $23,44^\circ$  relativ zur Äquatorebene geneigt ist. Die Gleichförmigkeit dieser Bewegung folgt aus der gleichförmigen Bewegung der Erde um die Sonne (da wir eine Kreisbewegung annehmen). Die Längengrade auf der Erde schneiden die scheinbare Bahn der Sonne, die relativ zum Äquator geneigt ist, jedoch nicht gleichförmig. Das bedeutet, dass die scheinbare Bewegung der Sonne relativ zu den Längengraden nicht gleichförmig ist, und das wiederum bedeutet, dass ein Sonntag zu manchen Zeiten kürzer bzw. länger ist. Genauer sind die Tage um den Frühlings- und Herbbspunkt die kürzesten und die Tage um die Sommer- und Wintersonnenwende die längsten.

Zur Berechnung dieses Effekts für die Zeitgleichung muss man einen gleichmäßig unterteilten Vollkreis, der unter  $23,44^\circ$  geneigt ist, entlang der Längengrade auf die Äquatorebene projizieren. Die auf der Ekliptik gleichen Bogenlängen werden auf der Äquatorebene ungleich - etwas dichter am Frühlingspunkt (derzeit 21.3.) und Herbbspunkt (derzeit 23.9.) und etwas weniger dicht bei der Sommersonnenwende (derzeit 21.6.) und der Wintersonnenwende (derzeit 21.12).

Die folgende Herleitung dieses Beitrags geht davon aus, dass sich der Effekt wieder als eine trigonometrische Funktion schreiben lässt, was in guter Näherung auch erfüllt ist. Diese Funktion hat im Verlaufe eines Jahres zwei Maxima (an den Sonnenwendepunkten) und zwei Minima (an den Äquinoktien). Damit müssen wir nur noch die Amplitude dieser Funktion bestimmen.

Wie aus Abb. 4 ersichtlich, ändert sich eine Bogenlänge  $a$ , wenn sie von der Ekliptik entlang der Längengrade auf den Äquator projiziert wird. In der Nähe der Äquinoktien ist die Beziehung durch  $a' = a \cos \varepsilon$  gegeben. In der Nähe der Sonnenwendepunkte am  $23,44$ .ten Breitengrad gilt ungefähr  $a = a' \cos \varepsilon$  (genauer ist dies die Beziehung zwischen den Bogenlängen entlang der beiden Breitengrade; der Abstand zwischen den beiden Punkten entlang der Ekliptik ist etwas kleiner, dies kann man

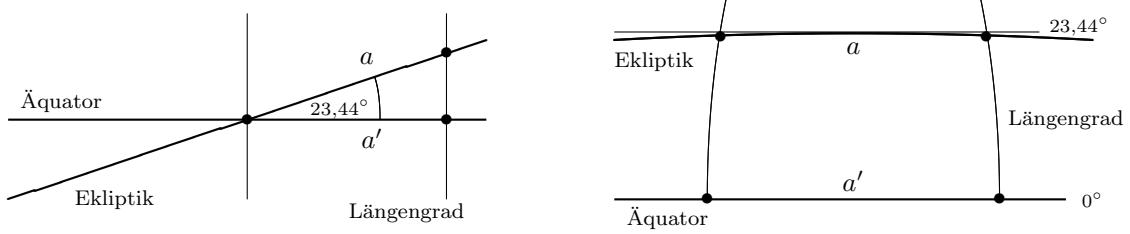


Abbildung 4: Projektionen von der Ekliptik auf die Äquatorebene in der Nähe der Äquinoktien (Schnittpunkte zwischen Äquator und Ekliptik; links) und der Sommer- bzw. Wintersonnenwende (rechts). Zwei Punkte, die entlang der Ekliptik eine Bogenlänge  $a$  haben, werden entlang der Längengrade auf zwei Punkte auf dem Äquator projiziert, die dort eine Bogenlänge  $a'$  haben. Man vergleiche diese Abbildung auch mit Abb. ?? und ??.

aber in führender Ordnung und wenn die Punkte genügend nahe beieinander liegen vernachlässigen). Machen wir wieder einen Ansatz der Form

$$\alpha(t) = \alpha_0 \left( 1 + B \cos \left( \frac{4\pi}{T} t \right) \right) \quad (18)$$

folgt

$$\frac{\alpha_{\min}}{\alpha_{\max}} = \frac{1 - B}{1 + B} = \cos^2 \varepsilon \quad (19)$$

oder

$$B = \frac{1 - \cos^2 \varepsilon}{1 + \cos^2 \varepsilon} = 0,0859153. \quad (20)$$

Dies führt auf maximale Schwankungen in der Tageslänge von  $\pm 20,27$  s und auf eine kumulative Zeitverschiebung von

$$ZG_2(t) = \sum_{t=0}^T \Delta T(t) \approx \frac{T}{4\pi} \Delta T \sin \frac{4\pi}{T} t \approx (9,82 \text{ min}) \cdot \sin \frac{4\pi}{T} t. \quad (21)$$

Wiederum ist  $T$  die Zeitdauer von einem Jahr, und da in der Summe die Diskretisierung in Tagen  $t$  vorgenommen wurde, sollte auch  $T$  in Tagen angegeben werden.  $t$  ist die Anzahl der Tage vom Frühlingspunkt (dort ist nun  $t = 0$ ) gemessen.

## 5 Die Zeitgleichung

Die Zeitgleichung ist die Summe dieser beiden Beiträge. Dabei ist noch zu berücksichtigen, dass die Zeitpunkte  $t = 0$  unterschiedlich gewählt wurden. Gibt man  $t$  in Tagen an, sollte die Periode für den ersten Anteil (elliptische Erdbahn) am 3. Januar beginnen, also am 3. Tag des Jahres, und die Periode für den zweiten Anteil (Neigung der Erdachse zur Ekliptik) am 21. März, in Jahren, die kein Schaltjahr sind, also am 80. Tag eines Jahres. Abb. 5 zeigt die Summe der beiden Beiträge.

Insgesamt sind die Minuten der Zeitgleichung zur wahren Ortszeit zu addieren, sodass man die mittlere Ortszeit erhält. Diese Wahl des Vorzeichens geht darauf zurück, dass man früher beispielsweise mithilfe eines Sextanten auf hoher See die wahre Ortszeit gemessen hat und nun die mittlere Ortszeit berechnen musste. Ist die Zeitgleichung negativ (beispielsweise zu Beginn des Jahres) müssen die entsprechenden Minuten (Absolutwerte) der Zeitgleichung also von der wahren Ortszeit abgezogen werden, um zur mittleren Ortszeit zu gelangen.

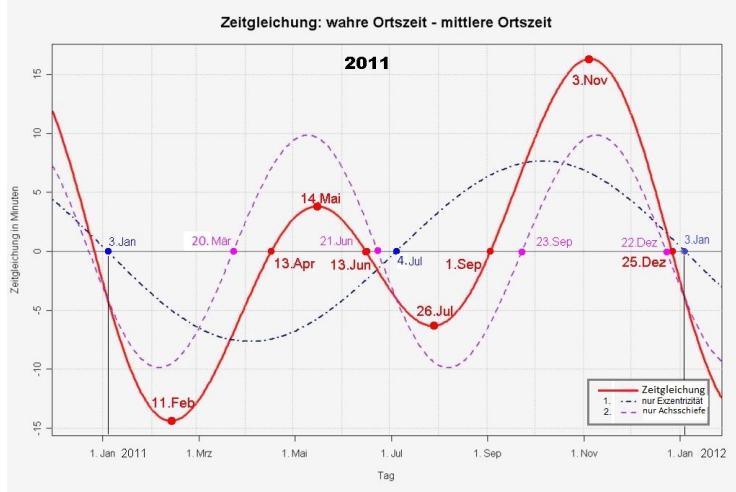


Abbildung 5: Die Zeitgleichung. Im Verlaufe eines Jahres kann die wahre Ortszeit von der mittleren Ortszeit um über eine Viertelstunde abweichen. Da sich das Perihel langsam verschiebt (in rund 21 000 Jahren einmal um  $360^\circ$ ) verschoben sich im Verlauf der Zeit auch die beiden Beiträge relativ zueinander. (aus [6])

## 6 Kuriositäten

### 6.1 Die Ekliptik auf alten Landkarten und Globen

Auf manchen alten Landkarten und Globen ist die Projektion der Ekliptik auf die Erde eingezeichnet. Ein Beispiel ist die bekannte Weltkarte von Hendrik Hondius aus dem Jahr 1630 (Abb. 6, oben).

Sowohl die Äquatorlinie als auch die Ekliptik sind in 360 äquidistante Einheiten von jeweils  $1^\circ$  unterteilt. Für die Äquatorlinie gilt das auch auf der Karte. Da der Großkreis der Ekliptik in seiner Projektion auf der Karte bei unterschiedlichen Breitengraden verläuft und der Maßstab der Karte vom Breitengrad abhängt (siehe auch das Kapitel [Landkarten](#)), erscheint die Unterteilung der Ekliptik nicht mehr gleichförmig. In der Nähe der Wendepunkten der Sonne, also in den Bereichen, wo die Ekliptik in der Nähe des  $\pm 23,5^\circ$  Breitengrads ist, sind die Unterteilungen auf der Ekliptik etwas breiter als am Äquator (siehe Abb. 6, unten rechts), da der Maßstab auf der Karte kleiner wird, wenn man sich vom Äquator entfernt. In den Ausschnitten (Abb. 6, unten) erkennt man, dass beim Schnittpunkt der Ekliptik mit dem Äquator (also den Äquinoktien) die Projektion der Ekliptik auf den Äquator die Bogenlänge schrumpfen lässt (Abb. 6, unten links). An diesen Stellen schneidet die  $10^\circ$  Bogenlänge der Ekliptik den Äquator links von der  $10^\circ$  Marke bzw. dem  $10^\circ$  Längengrad. Andererseits vergrößert wegen des veränderten Maßstabs die Projektion entlang der Längengrade in der Nähe der Wendepunkte die Bogenlänge (Abb. 6, unten rechts): Hier liegt die  $10^\circ$  Marke der Ekliptik ( $10^\circ$  neben dem Wendepunkt) rechts neben dem  $100^\circ$  Längengrad (der  $90^\circ$  Längengrad markiert den Wendepunkt) am Äquator.

### 6.2 Das Analemma

Das Analemma ist eine achtförmige Figur, die ungefähr den Schattenverlauf der Spitze eines Stabes (z.B. eines Obelisken oder der Spitze des Zeigers einer Sonnenuhr) zur selben mittleren Ortszeit (z.B. um 12 Uhr mittags) im Verlauf eines Jahres beschreibt (siehe Abb. 7, links). Die beiden Koordinaten dieser Figur entsprechen einmal dem unterschiedlichen Höchstand der Sonne im Sommer und Winter: Im Winter steht die Sonne auch zur Mittagszeit sehr tief und somit ist der Schatten vergleichsweise lang, im Sommer steht die Sonne hoch am Himmel und daher ist der Schatten entsprechend kurz.

Die dazu (fast) orthogonale Richtung entspricht der Zeitgleichung: Da die Position der Schattenspitze immer zur selben mittleren Ortszeit bestimmt wird, geht die wahre Sonne dieser Zeit entsprechend der Zeitgleichung entweder etwas vor oder nach, d.h. der Schatten ist etwas nach links oder



Abbildung 6: (Oben) Weltkarte von Hendrik Hondius aus dem Jahr 1630. Deutlich erkennbar neben der Äquatorlinie ist die Projektion der Ekliptik. (Unten) Ausschnitt aus der Weltkarte von Hendrik Hondius. (links) Ausschnitt beim Schnittpunkt von Äquator und Ekliptik, der hier auf einen der damals gebräuchlichen Nullmeridiane durch die Azoren gelegt wurde. Links unten sind Teile von Südamerika erkennbar, rechts oben Teile von Afrika. (rechts) Ausschnitt in der Nähe des südlichen Wendepunkts der Ekliptik (unser Winterpunkt). Der Kontinent links ist Afrika, die Insel rechts daneben Madagaskar; rechts oben erkennt man die Südspitze von Indien und Sri Lanka (damals Ceylon). ([3])

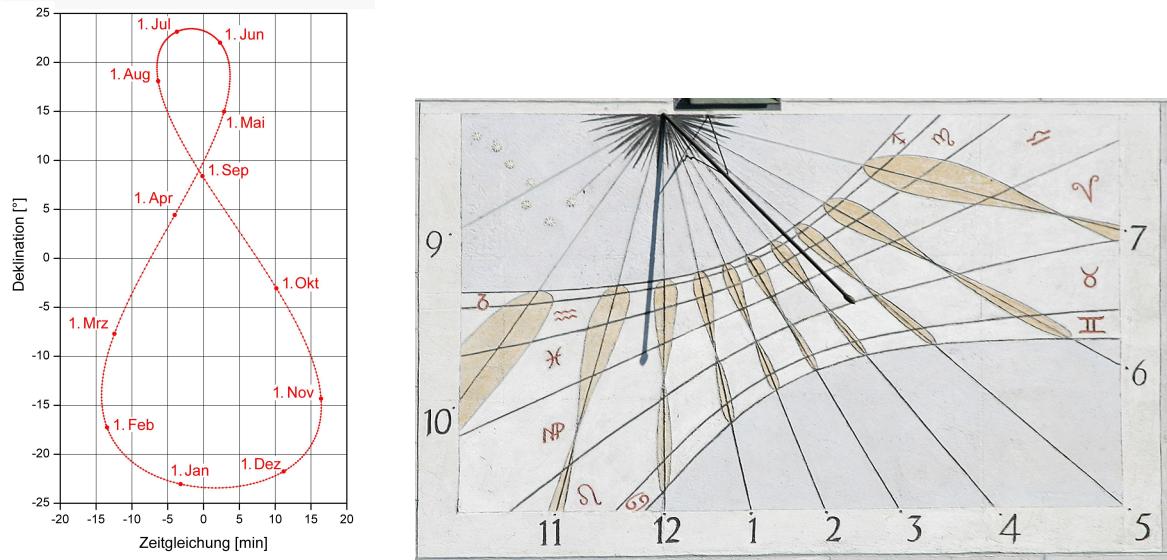


Abbildung 7: Das Analemma. (links) Die derzeitige Figur des Analemmas. Die obere Spitze der 8 beschreibt den Sonnenstand im Sommer, wenn die Sonne hoch steht und der Schatten entsprechend kurz ist. Der Zeiger oder Obelisk steht oberhalb der Figur und die Südrichtung ist nach oben (aus [1]). (rechts) Die Sonnenuhr am Alten Rathaus in München mit einem Analemma zur Korrektur der wahren Sonnenzeit auf die mittlere Sonnenzeit. (aus [1])

rechts versetzt.

Manche Sonnenuhren korrigieren den Effekt der Zeitgleichung. Dazu gibt es unterschiedliche Verfahren. Man kann z.B. die Schattenlinien zu einer festen Uhrzeit nicht als Geraden zeichnen (die einem festen Längengrad entsprechen würden) sondern in Form einer langgestreckten Acht. Hierbei wird, wie bei all diesen Verfahren, ausgenutzt, dass der Schatten im Winter länger ist als im Sommer.

Ein weiteres Verfahren besteht darin, den Schattenstab der Sonnenuhr unterschiedlich dick zu machen und zum Ablesen an einer Skala eine bestimmte Kante des Schattens zu verwenden (siehe Abb. 8). Solche Stäbe bezeichnet man auch als Bernhardt'sche Walze (benannt nach dem Techniker Martin Bernhardt, der diese Walze patentieren ließ, ähnliche Ideen hatte aber auch schon der Engländer John Ryder Oliver Ende des 19. Jahrhunderts). Ebenfalls wegen der unterschiedlichen Sonnenhöhe zu den verschiedenen Jahreszeiten würde ein jeweils anderer Teil des Schattenstabs den Schatten auf die Skala werfen. Wegen der unterschiedlichen Dicke erreicht man so, dass diese Schattenkante entsprechend verschoben ist. Da die Zeitgleichung in den beiden Jahreshälften unterschiedlich ist, benötigt man einen Schattenstab für die erste und einen anderen Schattenstab für die zweite Jahreshälfte.

Im Verlauf der Zeit verschieben sich die beiden Beiträge zur Zeitgleichung gegeneinander. Während sich der Neigungswinkel zur Ekliptik nur sehr wenig und über einen sehr großen Zeitraum von über 40 000 Jahren verändert, wandert das Perihel im Verlauf von rund 21 000 Jahren einmal durch das Jahr. Dadurch verändert sich die Zeitgleichung langsam und somit auch die Form des Analemmas. Derzeit ist das Analemma leicht asymmetrisch, weil die Wintersonnenwende und das Perihel nicht auf denselben Tag fallen (das war z.B. im Jahr 1246 der Fall). Außerdem sind die beiden „Hanteln“ der Acht unterschiedlich groß. Wenn das Perihel im Frühlings- bzw. Herbstpunkt steht, sind die beiden Teile gleich groß.



Abbildung 8: Sonnenuhr mit Bernhardt'scher Winterwalze. (aus [5])

### 6.3 Der kürzeste Tag

Die Wintersonnenwende entspricht gemeinhin dem kürzesten Tag. Derzeit liegt dieses Ereignis um den 21. Dezember. Zu diesem Zeitpunkt befindet sich die Sonne über dem südlichen Wendekreis (dem Wendekreis des Steinbocks). In Freiburg liegen zwischen Sonnenaufgang und Sonnenuntergang an diesem Tag 8 Stunden, 22 Minuten und 13 Sekunden. Doch kürzester Tag heißt nicht, dass es an diesem Tag am frühesten dunkel wird oder am spätesten hell. Der früheste Sonnenuntergang liegt um den 11. Dezember (in Freiburg um 16:35), der späteste Sonnenaufgang ist um den 2. Januar (in Freiburg gegen 8:18). Alle angegebenen Uhrzeiten beziehen sich natürlich auf die mittlere Sonnenzeit (genauer MEZ) und nicht die wahre Sonnenzeit. Der Grund für den Unterschied ist die Zeitgleichung.

Eine ähnliche Aussage gilt für die Sommersonnenwende. Der längste Tag mit 16 Stunden 2 Minuten und 47 Sekunden in Freiburg ist der 21. Juni. Doch der frühste Sonnenaufgang liegt um den 16./17. Juni (gegen 5:18), der späteste Sonnenuntergang um den 25./26. Juni (gegen 21:32). Genaue Zeiten für die Tageslänge, Sonnenauf- und Untergang sowie weitere interessante Daten für jeden beliebigen Ort der Erde findet man auf der Webseite [4].

## 7 Anhang

### 7.1 Äquivalenz des zweiten Kepler'schen Gesetzes und der Drehimpulserhaltung

Sei  $\mathbf{r}$  der Verbindungsvektor von einem Brennpunkt der Ellipse - dem Brennpunkt, in dem sich die Sonne befindet - zu einem Punkt auf der Ellipse (dem Ort eines Planeten) und  $\Delta\mathbf{r}$  die in der Zeit  $\Delta t$  zurückgelegte Strecke in tangentialer Richtung. Das zweite Kepler'sche Gesetz besagt, dass die in dem Zeitraum  $\Delta t$  (für  $\Delta t \rightarrow 0$ ) überstrichene Fläche  $\Delta F$ , d.h.

$$\Delta F = \frac{1}{2} \mathbf{r} \times \Delta \mathbf{r}, \quad (22)$$

entlang der Bahnkurve immer gleich ist. Damit ist aber auch der Drehimpuls

$$\mathbf{L} = m \mathbf{r} \times \frac{d\mathbf{r}}{dt} \quad (23)$$

konstant. Das Argument gilt natürlich auch in der umgekehrten Richtung: Aus der Drehimpulserhaltung folgt das zweite Kepler'sche Gesetz.

## Literatur

- [1] aus Wikipedia „Analemma“; S. Wetzel, CC BY-SA 4.0.
- [2] Till Niermann - Eigenes Werk, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=21792620>
- [3] Ravensburger Puzzle; Antique World Map, Hondius.
- [4] <https://www.timeanddate.com/sun/germany/freiburg>
- [5] Wikipedia „Sonnenuhr“; Präzisions-Sonnenuhr mit Winterwalze am Carl Zeiss Planetarium, Stuttgart. <https://de.wikipedia.org/wiki/Sonnenuhr>
- [6] Wikipedia „Zeitgleichung“; <https://de.wikipedia.org/wiki/Zeitgleichung> (aufgerufen am 14.5.2023).
- [7] Wikipedia „Zeitzone“. <https://de.wikipedia.org/wiki/Zeitzone>

# Zeitsysteme

Thomas Filk, Universität Freiburg

Seit 1967 ist die Sekunde durch die Frequenz  $\Delta\nu_{\text{Cs}}$  des Hyperfeinstrukturübergangs in 133-Cäsium im Grundzustand definiert. Es gilt:

$$1 \text{ s} = 9\,192\,631\,770 T = 9\,192\,631\,770 \frac{1}{\Delta\nu_{\text{Cs}}}. \quad (1)$$

Die Definition der Sekunde über den Hyperfeinstrukturübergang in 133-Cäsium löste 1967 die Ephemeridensekunde ab, die über die Position der Erde relativ zur Sonne definiert war.

Man könnte meinen, dass mit der Definition der SI-Sekunde und der Konstruktion von Uhren, die diese Sekunde genau genug messen können, das Problem der Zeitmessung gelöst sei. Im Detail ergeben sich jedoch Probleme, die zu verschiedenen Zeitsystemen geführt haben. Zum einen muss man zwischen theoretischen Zeitsystemen, die auf idealen Uhren basieren, und den Realisierungen solcher Zeitsysteme durch tatsächlich existierende und mit Fehlern behaftete Uhren unterscheiden. In diesem Zusammenhang ist zu definieren, wie man aus den fehlerbehafteten Werten realer Messinstrumente eine Zeitskala definiert kann. Außerdem muss man einen Referenzpunkt wählen, da nach der speziellen und allgemeinen Relativitätstheorie Uhren an verschiedenen Orten (im Gravitationsfeld) oder in verschiedenen Bewegungszuständen unterschiedliche Zeitsysteme definieren. Und schließlich möchte man natürlich, dass die Zeitsysteme trotz ihrer derzeitigen Definition über atomare Eigenschaften noch etwas mit dem Tag-und-Nacht-Rhythmus der Erde zu tun haben.

Auch heute noch sind verschiedene Zeitsysteme in Gebrauch. Die folgende Liste bzw. Behandlung dieser Zeitsysteme kann daher nur einen groben Überblick geben.

## 1 Das Julianische Datum

Die Julianische Tageszählung (englisch *Julian day*) zählt, unabhängig von einem speziellen Kalender oder der Definition eines Jahres, die Tage seit einem Anfangsdatum, das als der 1. Januar des Jahres 4713 v. Chr. nach dem zurückextrapolierten Julianischen Kalender gewählt wird. Genauer beginnt diese Zählung mittags um 12 Uhr (Sonnenhöchstand am heutigen Nullmeridian durch Greenwich). Wählt man nicht den zurückgerechneten Julianischen Kalender sondern den zurückgerechneten Gregorianischen Kalender, so beginnt diese Zeitrechnung mit dem 24. November des Jahres 4714 v. Chr. Und da es weder im Julianischen noch im Gregorianischen Kalender das Jahr 0 gibt, entspricht das Jahr 4713 dem Jahr –4712 im astronomischen Kalender, der das Jahr 0 hinzunimmt.

Die Wahl dieses Datums hat historische Gründe. Zunächst werden hier drei Zyklen zu einer Epoche zusammengefasst: Der Menton-Zyklus oder Mondzyklus von 19 Jahren, der Sonnenzyklus von 28 Jahren<sup>1</sup> und der sogenannte Indiktionszyklus von 15 Jahren (ein Zyklus, der im Altertum und Mittelalter oft verwendet wurde und ursprünglich mit Neuberechnungen der Steuern einherging). Das Produkt dieser drei Zyklenlängen sind 7980 Jahre; dies bezeichnet man als eine Epoche. Das Jahr 4713 v. Chr. wurde im Mittelalter errechnet als ein Jahr, in dem alle drei genannten Zyklen nach damaliger Rechnung begannen. Es hat den Vorteil, dass es vor jeder historischen Datierung liegt, d.h.

<sup>1</sup>Im Julianischen Kalender haben 4 Jahre - die Periode dieser Kalenderzählung - eine Dauer von 1461 Tagen, dabei verschieben sich die Wochentage um 5 Tage - der Rest bei einer Teilung von 1461 durch 7. Alle  $4 \times 7 = 28$  Jahre fallen also die Wochentage wieder auf dasselbe Datum. Im Gregorianischen Kalender entspricht eine Periode 400 Jahren mit 146 097 Tagen. Dies ist glatt durch 7 teilbar, d.h. alle 400 Jahre fallen im Gregorianischen Kalender dieselben Daten wieder auf dieselben Wochentage.

es gibt keine datierten historischen Ereignisse vor diesem Jahr (außer astronomische Ereignisse, die man heute zurückrechnen kann).

Dem Julianische Tag, der am 1. Januar 2000 mittags 12 Uhr (Universal time im Gregorianischen Kalender) begann, entspricht nach der Julianischen Tageszählung der Tag 2 451 545. Soviel Tage sind seit dem 1. Januar 4713 v. Chr. nach dem zurückgerechneten Julianischen Kalender vergangen. Diese Zählung ist so einfach, dass sie nicht nur in der Astronomie gerne verwendet wird, sondern auch in abgewandelter Form in vielen Computersystemen.

Will man neben dem Tag auch die Uhrzeit berücksichtigen, werden Nachkommastellen angegeben. Der Tag hat 86 400 Sekunden, sodass der 1. Januar 2000, 22 Uhr 25 Minuten und 10 Sekunden dem Julianischen Datum 2 451 545,43414 entspricht. Sowohl die Julianische Tageszählung (im Englischen *Julian Day*) als auch das Julianische Datum (*Julian Date*) werden mit JD abgekürzt. Man beachte, dass diese Tageszählung abgesehen von der Angabe des Anfangsdatums nicht davon abhängt, ob ansonsten der Julianische oder Gregorianische Kalender (oder welcher andere Kalender auch immer) verwendet wird.

Eine häufig verwendete Abwandlung ist das sogenannte Modifizierte Julianische Datum, abgekürzt mit MJD. Dies wird beispielsweise auch von dem BIPM (Bureau International des Poids et Mesures) in seinen *Circular T*, den monatlichen Berichten zur Bestimmung der Internationale Atomzeit TAI, verwendet. Das MJD unterscheidet sich vom JD in zwei Punkten:

1. Der Tagesanfang wird auf 0 Uhr Mitternacht (statt auf 12 Uhr mittags) gelegt.
2. Der Beginn ist 0 Uhr, 17. November 1858. Das bedeutet, vom Julianischen Tag werden 2 400 000,5 Tage abgezogen:  $MJD = JD - 2\,400\,000,5$ .

Damit hat der 1. Januar 2000, 12 Uhr mittags, den Wert:  $MJD = 51\,544,5$ .

## 2 Die Ephemeridenzeit - ET

Die Sekunde als der 1/86 400-ste Teil eines Tages ist erst seit dem Mittelalter in Gebrauch und bis 1956 war die Sekunde als der 1/86 400-ste Teil eines mittleren Sonnentages definiert. Man wusste zu diesem Zeitpunkt allerdings, dass auch der mittlere Sonnentag nicht wirklich vollkommen gleichmäßig verläuft bzw. dass der mittlere Sonnentag davon abhängt, über welchen Zeitraum genau gemittelt wird. Daher verwendete man schon seit längerem die Stellung der Erde relativ zur Sonne zur Definition einer Sekunde. So war in der Zeit zwischen 1956 und 1968 die Sekunde definiert als der 1/31 556 925,9747-ste Teil des tropischen Jahres 1900 (genauer: des theoretisch hochgerechneten tropischen Jahres 1900, bestimmt aus den Bahndaten der Erde zum Zeitpunkt 31.12.1899, mittags 12 Uhr). Diese Definition der Sekunde bezeichnet man auch als die Ephemeridensekunde.

Schon in babylonischer Zeit (um 1000 v. Chr.) gab es Tabellen zu den Positionen von Sonne, Mond und Planeten, und der sogenannte Almagest von Ptolemäus mit seinen genauen Beschreibungen der Planetenbahnen gilt als Höhepunkt antiker Astronomie. Im Mittelalter wurden weitere Tabellen erstellt (bekannt sind die Toledaner Tafeln aus dem 12. Jahrhundert, die Alfonsinischen Tafeln aus dem 13. Jahrhundert, die Ephemeridentafeln von Regiomontanus von 1474, oder auch die Rudolfinischen Tafeln von 1627 von Johannes Kepler).

Seit dem Mittelalter gab es Tabellen, in denen die Position der Sonne, des Mondes und der Planeten relativ zu einem Koordinatensystem (meist durch Angabe der Rektaszension und Deklination) und/oder relativ zueinander mit Daten und Tageszeiten korreliert wurden.

### 3 TT, TCG und TCB

TT (*Terrestrial Time* oder *Temps terrestre*), TCG (*Geocentric Coordinated Time* oder *Temps-coordonnée geocentrique*) und TCB (*Barycentric Coordinate Time* oder *Temps-coordonnée barycentrique*) sind idealisierte Zeitsysteme, manchmal spricht man auch von Platonischen Zeitsystemen, die auf idealen Uhren unabhängig von konkreten Realisierungen beruhen. Sie dienen meist der theoretischen Beschreibung von Objekten in unserem Sonnensystem.

TCB (baryzentrische koordinierte Zeit) ist die Zeit einer idealen, die SI-Sekunde anzeigen den Uhr, die sich in einem idealisierten hypothetischen System befindet, das sich parallel zum Schwerpunkt (dem Baryzentrum) des Sonnensystems bewegt, also dieselben Bewegungen wie dieser Schwerpunkt ausführt, aber keinem Gravitationspotenzial unterliegt. Es wird typischerweise für die Beschreibung von Bahnkurven von Planeten oder Kometen verwendet sowie von Raketen oder Satelliten, die den gravitativen Einfluss der Erde verlassen. Entsprechend ist TCG (geozentrische koordinierte Zeit) die Zeit einer idealen, die SI-Sekunde anzeigen den Uhr, die sich in einem idealisierten System befindet, das sich parallel zum Schwerpunkt des Erdmittelpunkts bewegt, aber keinem Gravitationspotenzial unterliegt. Sie dient zur Beschreibung von Bahnkurven im gravitativen Einfluss der Erde, beispielsweise den Bahnkurven von Satelliten, dem Mond oder Raketen.

TT (terrestrische Zeit) ist die Zeit einer idealen Uhr, die sich auf der Oberfläche der Erde befindet. Hierbei ist die Oberfläche definiert durch die gravitative Äquipotentialfläche (also konstantes Gravitationspotenzial) zur mittleren Meereshöhe, wobei der Effekt der Erddrehung (Potenzial zur Fliehkraft) einbezogen wird, nicht jedoch der Einfluss von Strömungen oder Gezeiten. Diese Fläche bezeichnet man auch als das Geoid der Erde. Da die mittlere Meereshöhe (z.B. im Zusammenhang mit dem Klimawandel) keine Konstante ist, hat man für das Potenzial den Wert  $U/c^2 = 6,969\,290\,134 \cdot 10^{-10}$  definiert (der Quotient aus einem Potenzial und  $c^2$  ist eine dimensionslose Konstante), was bei einer homogenen Kugel von der Masse der Erde einem Radius von etwas über 6 360 km entspricht.

TT und TCG unterscheiden sich also um einen multiplikativen Faktor  $(1 - U/c^2)$ , um den TT langsamer ist als TCG. Außerdem wurde definiert, dass die beiden Zeiten am 1. Januar 1977, um 0 Uhr 32,184 Sekunden, gleich waren. Die seltsame Sekundenzahl ergibt sich daraus, dass TT möglichst exakt an die vorher gebräuchliche Ephemeridenzeit angepasst werden sollte. Außerdem entspricht diese Zeit exakt 0 Uhr nach der Internationalen Atomzeit TAI (siehe nächsten Abschnitt). Früher verwendete man für die terrestrische Zeit TT die Bezeichnung TDT (terrestrial dynamical time); diese Bezeichnung findet man gelegentlich immer noch.

### 4 Die Internationale Atomzeit - TAI

Die Internationale Atomzeit TAI (*Temps atomique international*) ist eine Realisierung von TT. Hierbei handelt es sich um eine Zeit, die an existierenden Uhren abgelesen wird. Da eine einzelne Uhr immer ungenau ist und auch mal Störungen haben kann, handelt es sich bei TAI um einen gewichteten Mittelwert von derzeit über 400 hoch präzisen Uhren an über 80 Orten. In der Physikalisch Technischen Bundesanstalt (PTB) in Braunschweig stehen zwei solche Atomuhren - genannt CSF1 und CSF2, hierbei handelt es sich um Cäsium Fountain Clocks -, die zur TAI beitragen. In regelmäßigen Abständen übermitteln diese Uhren ihre Zeiten an das BIPM. Dort werden die Daten um verschiedene Faktoren korrigiert (dazu zählen Laufzeiten der Übertragung sowie relativistische Korrekturen aufgrund der Höhenunterschiede der verschiedenen Uhren), die Uhren mit den größten Abweichungen werden nicht berücksichtigt und von rund 300 Uhren wird ein gewichtetes Mittel gebildet. Die Gewichtung der einzelnen Uhren richtet sich nach ihrem geschätzten Fehler (der sich aus einer Theorie ihrer Funktionsweise bestimmt) sowie Ungenauigkeiten oder Schwankungen in der Vergangenheit.

Einmal im Monat wird im *BIPM Circular T* veröffentlicht, um wie viel die einzelnen Uhren von dem berechneten Mittelwert zu bestimmten Zeitpunkten (die dann rund einen Monat zurückliegen) abweichen.

TAI versucht also eine möglichst genaue Realisation einer idealen, die SI-Sekunde anzeigen den Uhr zu sein, die sich auf der Oberfläche des Geoids befindet. Insofern ist sie eine Realisation von TT, bis auf die oben erwähnten 32,184 Sekunden, um die sich die beiden Zeiten am 1. Januar 1977, 0 Uhr (TAI-Zeit), unterschieden. TT sollte dabei an die Ephemeridenzeit angepasst werden, an die TAI schon 1958 angepasst worden war. Die 32,184 Sekunden Unterschied erklären sich also daher, dass zwischen diesen beiden Zeitpunkten (1958 und 1977) schon ein solcher Unterschied zwischen der Atomzeit und der Ephemeridenzeit bestand.

## 5 Universal Times - UT0, UT1, UT2

Die Universal Times beziehen sich direkt auf die Lage der Erde und sind somit eine Fortführung der Ephemeridenzeit. Die Universal Time ist definiert über den Winkel, den der Nullmeridian auf der Erde relativ zu einem ausgezeichneten Punkt auf dem Himmelsäquator (der Projektion des Erdäquators vom Mittelpunkt der Erde aus auf die Himmelskugel) hat. Dieser Punkt auf dem Himmelsäquator ist heute definiert über ein Referenzsystem am Himmel, das über die Lage von festen Punkten am Himmel - meist Quasare, die auch im Radiowellenbereich nachweisbar sind, sodass eine VLBI (Very Long Baseline Interferometry) eine sehr genaue Richtungsbestimmung auch am Tag ermöglicht - definiert ist. Dieses Referenzsystem bezeichnet man auch als ICRF (*International Celestial Reference Frame*).

Die Ephemeridenzeit bezog sich auf einen gemittelten Tag, wie er durch das (vom 31.12.1899, 12 Uhr, hochgerechnete) Jahr 1900 definiert war. Diese Zeit nannte man auch GMT (*Greenwich Mean Time*). Sie entspricht genau dem, was man später UT0 nannte: Eine Zeitbestimmung über einen gemittelten Sonnentag, also die Zeit eines realen Sonnentags, die durch die Zeitgleichung - also die Einflüsse der elliptischen Bahn der Erde sowie ihrer Neigung gegenüber der Ekliptik - korrigiert wird und somit zur Zeit eines mittleren Sonnentags gehört. Der Winkel ERA (*Earth Rotation Angle*) des Nullmeridians relativ zu dem Referenzsystem am Himmel ist bis auf eine Konstante direkt proportional zur UT. Die Proportionalitätskonstante berücksichtigt, dass es sich hierbei um eine Messung der Erdrotation relativ zu einem Himmelssystem (siderischen Referenzsystem) handelt, und diese Zeit korrigiert werden muss, um einen Sonnentag zu erhalten. Dementsprechend entspricht dieser Proportionalitätsfaktor ziemlich genau dem Faktor  $(1 + 1/365,25)$ , um den der siderische Tag im Vergleich zum Sonnentag korrigiert werden muss. Genauer lautet die Beziehung:

$$\text{ERA} = 2\pi(0,7790572732640 + 1,00273781191135448 \cdot T_u) \text{ rad}, \quad (2)$$

wobei  $T_u = (\text{JD}(UT1) - 2\,451\,545,0)$  ist, und JD(UT1) dem Julianischen Datum nach der UT1-Uhrzeit entspricht. ERA ist die gemessene Variable (ein Winkel) und JD(UT1) - ein Datum mit einem Zeitpunkt - wird daraus berechnet. ERA, also der gemessene Winkel, wird dabei um zwei Faktoren korrigiert: (1) die Zeitgleichung, die den Unterschied zwischen mittlerem und wahrem Sonnentag angibt, und (2) Schwankungen in der Drehachse der Erde (sogenannte Polarbewegungen), durch die der Längengrad, beispielsweise eines Observatoriums, nicht eindeutig bestimmt ist. Weitere jahreszeitliche Schwankungen, die beispielsweise in dem Zeitsystem UT2 berücksichtigt werden, werden für UT1 nicht berücksichtigt.

## 6 Universal Coordinated Time - UTC

Wir haben nun zwei realisierte Zeitsysteme (d.h., Zeitsysteme, die durch Messungen an realen physikalischen Systemen bestimmt werden): Die TAI, die über Atomuhren bestimmt wird, die möglichst nahe an der SI-Sekunde arbeiten, und UT1, die über die Lage der Erde relativ zur Sonne bestimmt wird. UT1 hat den Vorteil, dass 12 Uhr mittags mit unserer Vorstellung von „Sonnenhöchststand“ zusammenfällt, wohingegen TAI auf einer möglichst genauen und gleichmäßigen Realisierung einer Sekunde beruht. Damit entsteht das Problem: An welches Zeitsystem sollen wir uns halten, wenn diese beiden Zeitsysteme auseinanderlaufen. UT1 hat den Vorteil, unseren Vorstellung von Tag und Nacht zu entsprechen, TAI hat den Vorteil, dass die Sekunden immer gleich lang sind.

Um ein Auseinanderlaufen der beiden Zeitsysteme auszugleichen, hat man sich (nach anfänglichen Problemen bei der Namensgebung wie auch bei der genauen Definition) 1972 auf das Zeitsystem der UTC (*Universal Coordinated Time*) geeinigt. UTC verläuft parallel zu TAI, d.h., UTC richtet sich bezüglich der genauen Zeitangabe nach der besten Realisierung der SI-Sekunde auf der Erdoberfläche, und das ist die TAI. Bevor allerdings UT1 und TAI um 0,9 Sekunden auseinanderlaufen (weil die Erddrehung gewissen Schwankungen unterworfen ist), wird für UTC eine Schaltsekunde entweder eingefügt oder weggelassen. In der Vergangenheit wurden nur Schaltsekunden eingefügt, da die Erdrotation etwas langsamer ist als die TAI-Zeit. Als UTC im Jahre 1972 eingeführt wurde bestand schon ein Unterschied von 10 Sekunden zwischen UT1 und TAI, sodass damals definiert wurde UTC = TAI - 10 s. Seitdem wurden insgesamt 27 weitere Schaltsekunden eingefügt, die letzte am 31. Dezember 2016. Sämtliche Schaltsekunden wurden in der Vergangenheit entweder am 31. Juni oder am 31. Dezember eingefügt. In den letzten Jahren hat die Geschwindigkeit der Eigendrehung der Erde wieder etwas zugenommen, sodass nicht ausgeschlossen wird, dass in der nahen Zukunft zum ersten Mal in der Geschichte der UTC eine Sekunde „herausgenommen“ wird.

Das Einfügen von Schaltsekunden führte dazu, dass im Zeitsystem der UTC gelegentlich eine Minute 61 Sekunden hat. Statt nach der 59.-sten Sekunde die 0.te Sekunde folgen zu lassen, zählt man eine 60.ste Sekunde hinzu und beginnt dann den neuen Monat mit der Sekunde 0. Da in der Vergangenheit noch nie eine Schaltsekunde entfernt wurde, gibt es gewisse Bedenken, ob alle Computersysteme mit einem solchen Schritt zurecht kommen. Da auch beim Einfügen von Schaltsekunden in der Vergangenheit immer wieder Probleme mit verschiedenen digitalen Systemen auftraten, wurde in jüngerer Zeit überlegt, die Definition von UTC nochmals zu ändern. Es wird angestrebt, bis 2035 eine neue Definition zu finden, bei der nur alle paar Jahrhunderte eine Korrektur notwendig wird. Beispielsweise könnte man die Differenz zwischen UT1 und TAI auf mehrere Minuten anwachsen lassen, bevor Korrekturen vorgenommen werden. Rein subjektiv werden wir als Menschen die langsame Verschiebung der Mittagsstunde ohnehin erst bemerken, wenn sich die Differenz auf die Größenordnung einer Stunde summiert hat.

Die UTC ist die Zeit, die wir beispielsweise über Funkuhren, das Radio oder Fernsehen empfangen. Diese Zeit ist bis auf 0,9 Sekunden an die Orientierung der Erde relativ zum Referenzsystem des Himmels angepasst, läuft parallel zur Atomzeit TAI, d.h. ist sehr regelmäßig, hat aber den Nachteil, dass eine Minute gelegentlich eine Sekunde mehr oder weniger hat. Wegen der Unregelmäßigkeit der Erddrehung können Schaltsekunden nicht langfristig vorhergesagt werden, sondern erst in einem Zeitraum von einem halben Jahr.

## 7 Lokalzeit - Local Time

Die Lokalzeit ist die Zeit, die von Radio- oder Fernsehstationen, Funkuhren bzw. öffentlichen Uhren an einem Ort angezeigt wird. Sie richtet sich nach der UTC, unterliegt also unter anderem der Einfügung

oder Löschung von Schaltsekunden, unterscheidet sich von der UTC aber in zweierlei Hinsicht:

1. Sie berücksichtigt die lokale Zeitzone: UTC ist die Fortsetzung der GMT (Greenwich Mean Time) und richtet sich nach dem Nullmeridian durch Greenwich, d.h. nach dem Längengrad, wo im Jahresmittel mittags um 12 Uhr die Sonne ihren Höchststand hat. Damit man an nahezu allen Orten um die Mittagszeit den Sonnenhöchststand hat, wurde die Erde in Zeitzonen eingeteilt, die einseits von Längengraden, andererseits aber auch von Ländergrenzen berandet sind. Die meisten Zeitzonen unterscheiden sich von der Zeitzone von Greenwich um eine volle Stunde. Alle 15 Längengrade beginnt also rein rechnerisch eine neue Zeitzone, wobei die meisten Länder (Ausnahmen sind die USA, Kanada, Russland und Australien) nur eine Zeitzone haben, was z.B. für China bedeutet, dass es sich rechnerisch über fast vier Zeitzonen erstreckt aber im gesamten Land dieselbe Zeitzone gilt. Es gibt aber auch einige Länder mit halbstündiger Zeitverschiebung (Iran  $+3\frac{1}{2}$ , Afghanistan  $+4\frac{1}{2}$ , Indien  $+5\frac{1}{2}$ , Burma  $+6\frac{1}{2}$ , Zentralaustralien  $+9\frac{1}{2}$ ) sowie viertelstündiger Zeitverschiebung (Nepal  $+5\frac{3}{4}$ , kleine Teile von Australien  $+8\frac{3}{4}$  sowie die Chatham Islands  $+10\frac{3}{4}$ ).

Zentraleuropa (mit Ausnahme von Portugal und den Inselgruppen Kanaren, Madeira, Island und Azoren) verwenden die Mitteleuropäische Zeit (MEZ), die sich von der UTC um +1 Stunde unterscheidet, d.h., wenn es in Greenwich 12 Uhr mittags ist, ist es in Mitteleuropa bereits 1 Uhr mittags.

2. In den Sommermonaten wechseln viele Länder auf die Sommerzeit. Dazu verschiebt man die Zeitzone um eine weitere Stunde nach Osten, d.h., relativ zum Sonnenstand ist es eine Stunde später. Damit verbunden ist etwas längere Dunkelheit am Morgen und etwas längere Helligkeit am Abend. In diesem Fall spricht man in Mitteleuropa von der MESZ - Mitteleuropäische Sommerzeit, die sich von der UTC um +2 Stunden unterscheidet.

Ungefähr entlang des 180-sten Längengrads (im Pazifik, durch die Fiji-Inseln) verläuft die Internationale Datumsgrenze. Überquert man diese Grenze von West nach Ost, muss man das Datum um einen Tag zurückstellen, überquert man sie von Ost nach West stellt man das Datum um einen Tag vor.

## 8 GPS Time

Abschließend soll noch kurz auf die GPS-Zeit des Global Positioning Systems (GPS) eingegangen werden. Das GPS besteht unter anderem aus 24 aktiven Satelliten (es befinden sich derzeit - Dezember 2022 - 32 Satelliten im Orbit, davon sind 31 einsatzbereit, sieben der Satelliten dienen als Backup), die ständig ein Signal aussenden, das die genaue Zeit sowie den genauen Ort dieser Satelliten angibt. Mindestens vier dieser Satelliten befinden sich jederzeit in „Sichtlinie“ von jedem beliebigen Punkt der Erde aus. Aus den Signalen kann ein geeigneter Empfänger seine genaue Position sowie die genaue Zeit bestimmen.

Die vom GPS-System verwendete GPS-Zeit richtet sich im Wesentlichen nach der TAI, d.h. es wird die auf das Geoid der Erde bezogene Eigenzeit verwendet, ohne Einschub oder Wegnahme von Schaltsekunden. Das bedeutet unter anderem, dass die Uhren in den GPS-Satelliten „falsch gehen“: Es handelt sich nicht um SI-Uhren, die die Eigenzeit in den Satelliten messen, sondern diese Eigenzeiten werden um die relativistischen Effekte aufgrund der Bewegung des Satelliten sowie des Gravitationsfelds der Erde korrigiert, sodass diese Uhren die Zeit angeben, die an einem ruhenden Ort auf dem Geoid der Erde gilt. Damit stimmen diese Uhren in ihrem Zeittakt mit der Atomzeit TAI überein.

Als Startpunkt der GPS-Zeit wurde 0 Uhr am 6. Januar 1980 festgelegt. Zu diesem Zeitpunkt unterschieden sich UTC und TAI bereits um 19 Schaltsekunden. Da die GPS-Zeit zu diesem Zeitpunkt an die UTC angepasst wurde, unterscheiden sich GPS-Zeit und TAI also dauerhaft um 19 Sekunden:  $\text{GPS} = \text{TAI} - 19\text{s}$ . Derzeit (Dezember 2022) unterscheiden sich UTC und GPS-Zeit um 18 Sekunden; diese 18 Sekunden wurden seit dem 6. Januar 1980 bei der UTC als Schaltsekunden eingefügt. Damit geht UTC relativ zur GPS-Zeit um 18 Sekunden nach, d.h. es gilt:  $\text{GPS} = \text{UTC} + 18\text{s}$ . Diese Zeitdifferenz ändert sich aber, falls bei der UTC Schaltsekunden eingefügt oder weggelassen werden.

Die Zählung bei der GPS-Zeit verwendet mehrere Einheiten: Epochen, Wochen, Tage und Sekunden.

1. Eine Epoche besteht aus 1024 Wochen. Die Wochenzahl wird als 10-Bit Zeichenfolge übertragen. Der Woche 1023 folgt die Woche 0. Dies bezeichnet man auch als Rollover. Eine Epoche dauert somit 7168 Tage oder etwas über 19,6 Jahre. Da diese Rollover (bisher fanden zwei solche Rollover statt - in der Nacht vom 21. auf den 22. August 1999 und in der Nacht vom 6. auf den 7. April 2019) zu Problemen bei manchen Anwendern geführt haben, will man in naher Zukunft die Zeichenfolge für die Wochen auf 13 Bit erweitern, sodass nur ungefähr alle 157 Jahre ein solcher Rollover stattfindet. Derzeit (am 11. Dezember 2022) befinden wir uns in der GPS-Woche 2240 (dies ist die Anzahl der Wochen, die seit dem 6. Januar 1980 vergangen sind), also in der 192. Woche der Epoche 2.
2. Die Wochen werden mit Beginn vom 6. Januar 1980 gezählt. Da es jedoch derzeit noch wegen der 10-Bit-Folge für die Angabe der Wochen zu Rollover kommt, gibt die GPS-Zeit nur die Wochenzahl der laufenden Epoche wieder (am 11. Dezember 2022 war das die Woche 192).
3. Für jede Woche wird der Tag angegeben, also ein Wert zwischen 0 und 6. Die neue Woche beginnt mit dem Sonntag, dem Tag 0.
4. Für jeden Tag werden die Sekunden angegeben. Der Tag beginnt um Mitternacht 0 Uhr.

Damit besteht eine volle Angabe der GPS-Zeit aus: Wochenzahl (gesamt), Epoche + Woche innerhalb der Epoche (ein Wert zwischen 0 und 1023), Tag (ein Wert zwischen 0 und 6) und Sekunden an diesem Tag (ein Wert zwischen 0 und 86399). Die Gesamtzahl der Wochen bzw. die Epoche wird allerdings nicht im GPS-Signal kodiert. Der Empfänger bzw. Anwender muss also wissen, in welcher Epoche man sich befindet.

## 9 Kuriositäten

### 9.1 Die Datumsgrenze

Es wurde oben erwähnt, dass man bei der Überquerung der Datumsgrenze von West nach Ost das Datum um einen Tag zurückstellen muss (also 24 Stunden subtrahieren muss), bei der Überquerung von Ost nach West muss das Datum entsprechend um einen Tag vorgestellt werden. Als Scherzfrage für Kinder in den unteren Klassen bietet sich nun folgendes Gedankenexperiment an: Angenommen, man könnte mit einem sehr schnellen Flugzeug immer von West nach Ost um die Erde reisen und dabei die Datumsgrenze in kurzer Zeit mehrfach von West nach Ost überqueren, dann würde man jedesmal das Datum um einen Tag zurücksetzen. Kann man auf diese Weise in die Vergangenheit reisen, also ist man beispielsweise nach fünfmaliger Überquerung der Datumsgrenze um fünf Tage zurückgereist? Eine ähnliche Frage kann man natürlich auch bezüglich der Reisen in die Zukunft stellen, da man bei Reisen um die Erde von Ost nach West jedesmal beim Überqueren der Datumsgrenze das Datum um einen Tag vorstellen muss.

Natürlich geht das nicht: Wenn man von West nach Ost reist muss man ja jedesmal, wenn man in eine neue Zeitzone gelangt, die Uhr um eine Stunde vorstellen. Hat man die Erde dann einmal umrundet, wurde die Uhr um insgesamt 24 Stunden vorgestellt. Beim Überqueren der Datumsgrenze gleicht man dies wieder aus, indem die Uhr um 24 Stunden zurückgesetzt bzw. das Datum um 1 Tag zurückgesetzt wird.

## 9.2 Verschiebung der Datumsgrenze zum Millenium

Die Kiribati Inseln (offiziell die Republik Kiribati) bilden eine Inselgruppe in der Mitte des Pazifiks, die sich entlang des Äquators vom ungefähr 170. Längengrad Ost (die Insel Banaba, westlich der Gilbert Islands) bis zum 150. Längengrad West (die Insel Caroline, heute Millennium Island, in der Inselgruppe der Line Islands) erstreckt. Die Datumsgrenze verlief bis Mitte der 90er Jahre des letzten Jahrhunderts durch diese Inselgruppe, sodass auf verschiedenen Inseln zum selben Zeitpunkt ein unterschiedliches Datum herrschte. Zum 1. Januar 1995 wurde die Datumsgrenze von der Republik Kiribati so verlegt, dass sie nun östlich um die Line Inseln herumläuft und damit zwei Zeitzonen in das alte Datum hineinragt (das bei Inseln in der Nähe, z.B. auch auf Hawaii, das auf demselben Längengrad liegt, gültig ist). Als offizieller Grund wurde angegeben, dass ein unterschiedliches Datum innerhalb eines Landes zu Problemen in landesübergreifenden Angelegenheiten führe. Inoffiziell wurde allerdings auch nie bestritten, dass ein wirtschaftlicher Grund dahinter steckte: Auf diese Weise wurden die Line Islands die ersten Gebiete, die zum Millenium ins neue Jahrtausend wechselten. Man erhoffte sich dadurch als touristische Attraktion das erste Land der Welt zu sein, in dem der Sonnenaufgang im neuen Jahrtausend beobachtet werden konnte.

# Zeitmessung

Thomas Filk, Universität Freiburg

Im Jahre 1967 wurde beschlossen, die Sekunde über den Hyperfeinstrukturübergang von  $^{133}\text{Cs}$  im Grundzustand zu definieren: Eine Sekunde entspricht der Zeitdauer von 9 192 631 770 Schwingungen zu diesem Übergang. Damit erhebt sich die Frage, wie man die Zeit überhaupt messen kann. Die Geschichte der Zeitmessung ist eines der spannendsten Kapitel in der Geschichte der Physik.

## 1 Antike Zeitmesser

Für große Zeiträume dienten in der Antike immer die natürlichen Einheiten Tag und Monat, die sich leicht aus der Bewegung der Sonne (bzw. der Drehung der Erde) und des Monds bestimmen ließen. Auch wenn der Zyklus eines Jahres etwas schwieriger festzustellen war, kannte man schon in den frühen Kulturen Verfahren, mit deren Hilfe man beispielsweise die Tag-und-Nacht-Gleiche oder die Sommersonnenwende und damit ein Kalendersystem festlegen konnte. Schon den Babylonierinnen war bekannt, dass die vier astronomischen Jahreszeiten - die Zeiten zwischen den Tag-und-Nacht-Gleichen (Frühling und Herbst) und den Sonnenwenden (Sommer und Winter) - nicht gleich lang waren. Heute wissen wir, dass dies an der elliptischen Umlaufbahn der Erde um die Sonne liegt: Beim sonnennächsten Punkt der Erdbahn (dem Perihel - derzeit zwischen dem 2. und 5. Januar) bewegt sich die Erde nach dem zweiten Kepler'schen Gesetz etwas schneller um die Sonne als beim sonnenfernsten Punkt, dem Aphel (derzeit um den 3.-6. Juli). Und Ptolemäus wusste ebenfalls, dass sich die Sonne gleichmäßig entlang der Ekliptik bewegt und somit nicht gleichmäßig entlang der Projektion der Sonnenbahn auf den Äquator bzw. die Längengrade [5]. Die Korrekturen zwischen wahrem und mittlerem Sonnentag - die Zeitgleichung - werden schon bei Ptolemäus beschrieben, auch wenn sie erst im 17. Jahrhundert mit den Kepler'schen Gesetzen erklärt werden konnten. Im Folgenden geht es jedoch eher um die Einteilung des Tages in Untereinheiten und die Messung von kurzen Zeiträumen.

### 1.1 Sonnenuhren

Ein wichtiges und natürliches Zeitmaß war seit jeher die Sonnenuhr. Schon in vorhistorischen Zeiten dürften die Schatten von Felsvorsprüngen, Bäumen oder Sträuchern und ähnlichen natürlichen Gegenständen beobachtet und dadurch eine Tageseinteilung vorgenommen worden sein. Man wird auch früh festgestellt haben, dass die Sonne zu verschiedenen Jahreszeiten eine unterschiedliche Höhe hat und somit die Schattendlänge an einer Sonnenuhr von der Jahreszeit abhängt.

An manchen Sonnenuhren wurde diese jahreszeitliche Abhängigkeit genutzt, um die angezeigte wahre Sonnenzeit über ein sogenanntes Analemma (eine Figur in Form einer verzerrten Acht, mit der die Zeitgleichung so auf die Uhr projiziert wird, dass man aus dem Schatten die mittlere Sonnenzeit ablesen kann) mit der mittleren Sonnenzeit in Bezug zu setzen (siehe Abb. 1).

Das Analemma ist gleichzeitig ein Bild der Sonnenstände am Himmel: Macht man täglich zur selben mittleren Uhrzeit ein Bild vom Sonnenstand, so ergibt sich im Verlauf eines Jahres die Form des Analemmas.

### 1.2 Wasseruhren und andere Zeitmesser des Altertums

Ein Hilfsmittel zur Reproduktion gleicher Zeiträume war auch die Klepsydra: ein mit Wasser gefülltes Gefäß mit einer kleinen Öffnung an der unteren Seitenwand oder im Boden, durch die das Wasser

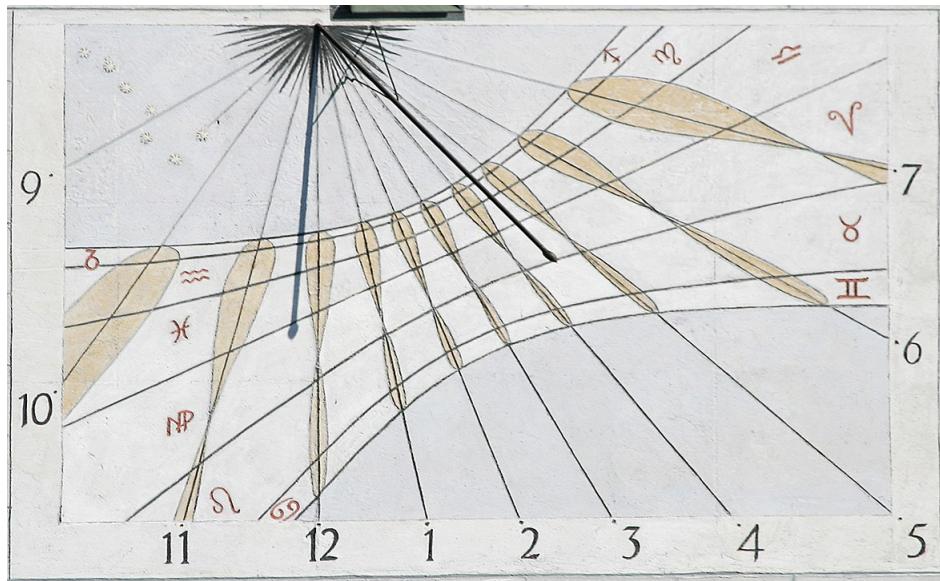


Abbildung 1: Sonnenuhr am Alten Rathaus in München. Anhand der Schattenlänge zu verschiedenen Jahreszeiten kann die wahre (angezeigte) Uhrzeit um die Zeitgleichung korrigiert werden, die durch die Analemmata an den Stundenlinien angegeben wird. Die Sternzeichen geben an, welcher Teil des Analemmata (linker oder rechter Rand) in einem bestimmten Monat zu verwenden ist. (aus [1])

auslaufen konnte, das dann in einem zweiten Gefäß aufgefangen wurde. Mit solchen Klepshydren wurden beispielsweise die Redezeiten bei politischen Versammlungen oder auch bei Gerichtsprozessen im alten Griechenland und Rom gemessen. Vermutlich geht hierauf der Begriff „seine Zeit ist abgelaufen“ zurück. Noch heute beruht die Sanduhr, gelegentlich zum Eierkochen oder auch bei Gesellschaftsspielen verwendet, auf diesem Prinzip.

Im frühen Mittelalter wurden die Klepshydren zu komplizierteren Wasseruhren abgewandelt. Dabei handelte es sich um teilweise sehr aufwendige Anlagen, durch die Wasser aus einem oberen Reservoir in ein tiefer gelegenes Reservoir floss und dabei komplexe Mechanismen in Gang setzte. Manche dieser Mechanismen glichen auch schon den späteren Hemmungen.

Insbesondere in Klöstern, wo teilweise auch zu Nachtstunden zum Gebet aufgerufen wurde, kamen später sogenannte Stundenkerzen - Kerzen genormter Dicke, bei denen Markierungen den groben Stundenverlauf anzeigen - hinzu. Einem ähnlichen Zweck dienten auch Ölkerzen, bei denen ein Docht in ein mit Öl gefülltes Gefäß ragte und am anderen Ende brannte. Die verflossene Zeit konnte an dem verbrauchten Öl abgelesen werden.

## 2 Uhren

Um 1300 erlaubte die Erfindung der sogenannten Hemmung die Konstruktion der ersten Räderuhren und damit von mechanischen Uhren, die auf oszillatorischen Prozessen beruhen. Wer den Mechanismus der Hemmung erfunden hat oder wann genau dieser Mechanismus erfunden wurde, ist nicht bekannt. Wir wissen heute lediglich, dass um 1300 die ersten Uhren entstanden sind, die auf diesem Prinzip beruhen.

## 2.1 Hemmungen

Eine Hemmung erlaubt es, eine mechanische Kraft (z.B. eine Gewichtskraft) in eine regelmäßige Bewegung umzuwandeln. Hängt beispielsweise ein Gewicht an einem langen Seil, aufgewickelt auf eine frei drehbare Rolle, so würde ungehemmt das Gewicht mit wachsender Geschwindigkeit absinken und die Rolle würde sich mit zunehmender Geschwindigkeit drehen. Eine Hemmung bewirkt, dass diese Drehung in eine meist abgehackte, konstant periodische Bewegung umgewandelt wird und somit für eine Zeitmessung verwendet werden kann. Die ältesten Hemmungen sind die sogenannten Spindelhemmungen mit einem Foliot - einem hin und her schwingenden Querbalken mit Gewichten, der über eine vertikale Achse mit zwei Plättchen in ein Zahnrad (das Kronrad) eingreift (siehe Abb. 2) - oder einer Unrast, einer rotierende Scheibe, die ansonsten nach einem ähnlichen Mechanismus wie das Foliot arbeitet.

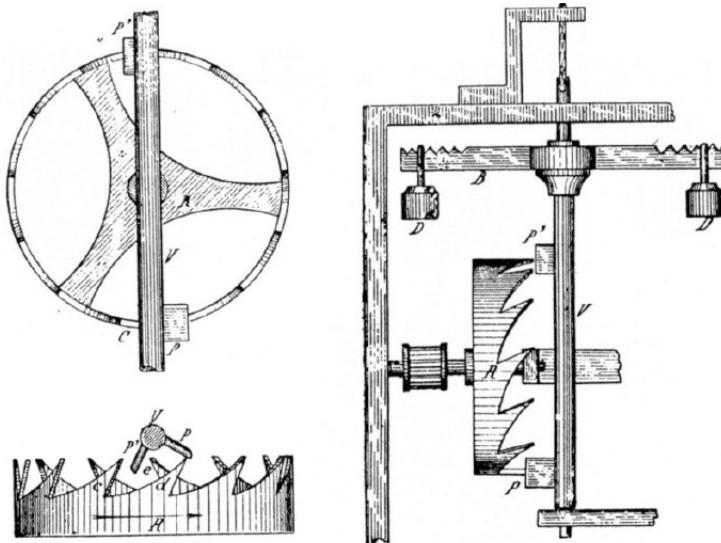


Abbildung 2: Eine Spindelhemmung mit Foliot. Das Kronrad  $R$  wird über ein Gewicht, das an einem über die Achse des Kronrads gewickelten Seil hängt (nicht dargestellt), angetrieben. Das Foliot  $B$  mit den Gewichten  $D$  besteht aus einem waagerechten Balken, der fest mit einer vertikalen Achse  $V$  verbunden ist. An dieser Achse befinden sich zwei Plättchen  $P$  und  $P'$ , die abwechselnd oben und unten in die Zähne ( $c$ ,  $d$  und  $e$ ) des Kronrads  $R$  einhaken. Dies führt zu einer Hin- und Herbewegung des Foliots und damit zu einer kontrollierten, langsamen, abgehackt gleichmäßigen Drehung des Kronrads. Über die Gewichte kann die Schwingungsfrequenz des Foliots und damit der Gang der Uhr reguliert werden. (aus [3])

Ein Video mit dem Mechanismus einer Spindelhemmung findet man unter [4]. Über Zahnrädermechanismen, die im Mittelalter schon lange bekannt waren, kann die Drehung des Kronrads auf andere Räder übertragen und dabei beliebig verlangsamt werden. Bis zum 17. Jahrhundert erreichten Uhren mit Spindelhemmungen eine Genauigkeit von rund 15 Minuten pro Tag.

## 2.2 Pendeluhrn

Im 17. Jahrhundert wurde die Spindelhemmung durch die mit einem Pendel gekoppelte Ankerhemmung ersetzt (siehe Abb. 3). Nachdem Galileo um 1630 festgestellt hatte, dass die Periode eines Pendels für kleine Auslenkungen nahezu unabhängig von dieser Auslenkung ist, erkannte man, dass sich das Pendel im Vergleich zum Foliot besser als Taktgeber eignet. Durch das Drehmoment, das

über ein Gewicht auf das Kronrad übertragen wird, gibt das Kronrad bei seinem Weiterdrehen über den Anker umgekehrt dem Pendel regelmäßig einen kleinen Stoß, sodass dieses mit einer kleinen aber konstanten Auslenkung schwingt und nicht durch Reibungseffekte zur Ruhe kommt.

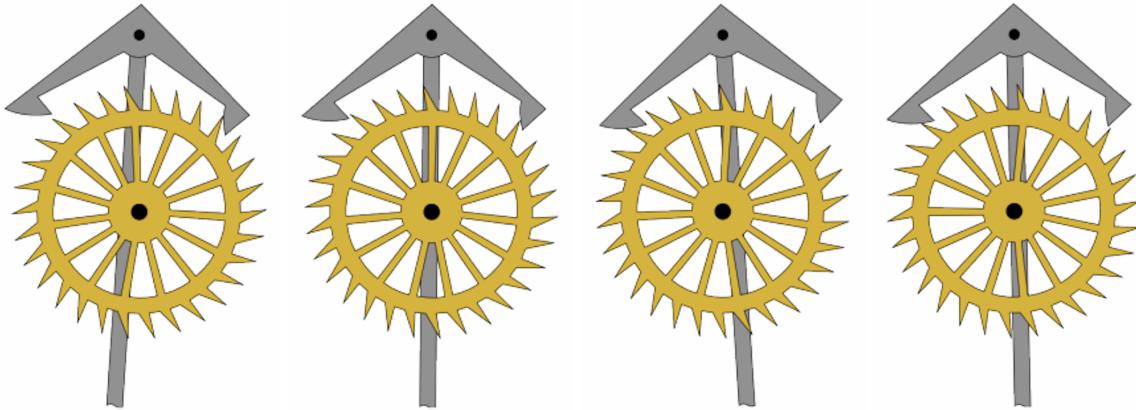


Abbildung 3: Eine Ankerhemmung. Der Anker ist fest mit einem Pendel verbunden, und wenn das Pendel schwingt, greift der Anker abwechselnd rechts und links in die Zacken des Kronrads. Das Kronrad dreht sich im Uhrzeigersinn und wird von einem Gewicht (nicht dargestellt) angetrieben. Bei jeder Vorwärtsbewegung überträgt es aufgrund der Zackenform und der Form des Ankerhakens einen kleinen Stoß auf den Anker und damit auf das Pendel. (aus [4])

Christiaan Huygens entwickelte um 1656 eine Uhr auf dem Prinzip der Pendeluhr. Nachdem Robert Hooke 1658 die Ankerhemmung erfunden hatte, die eine kleine Schwingungsauslenkung des Pendels ermöglichte und damit eine höhere Genauigkeit in der Pendelperiode, wurde die Pendeluhr zum Standard in der Uhrenkonstruktion. Hier zeigt sich das Grundprinzip der heutigen Uhren: Eine harmonische Schwingung dient einerseits als Standard für die Zeitmessung, andererseits wird sie durch den Mechanismus der Rückkopplung von einer äußeren Kraft (hier das Gewicht, welches das Kronrad antreibt) aufrecht erhalten.

### 3 Das Problem des Längengrads

Die Eroberung Konstantinopels im Jahre 1453 durch die osmanischen Truppen markiert einen entscheidenden Wendepunkt in der Geschichte des Abendlands. Oftmals wird in diesem Ereignis eine Ursache für den Wechsel vom Mittelalter zur Neuzeit gesehen. Diese Eroberung nach einer langen Belagerung löste zwei Bewegungen aus: (1) Nachdem der Landweg von Europa zu den attraktiven Handelsplätzen in Indien, Indonesien und China im Wesentlichen gesperrt bzw. unter islamischer Kontrolle war, was oft hohe Abgaben bzw. Zölle zur Folge hatte, suchte man nach anderen Wegen nach Indien. Dies führte nicht nur zur Umsegelung Afrikas, sondern auf der Suche nach einem direkten Weg nach Indien auch zur Entdeckung Amerikas und des Pazifischen Ozeans. (2) Die Flucht vieler Gelehrter aus Konstantinopel in den Westen, die in ihrem Gepäck oft wertvolle Bücher mit sich führten, löste die Renaissance aus - auf diese Weise wurden viele antike Schriften, teilweise in ihren arabischen Übersetzungen, im Westen erst bekannt.

Insbesondere die Erkundungsfahrten zur See brachten einige Probleme auf: Bis zu Beginn des 15. Jahrhunderts beschränkten sich Seefahrten hauptsächlich auf den Mittelmeerraum oder aber auf Fahrten entlang der europäischen Küsten nach Norden (Frankreich, England, Norwegen) oder entlang der afrikanischen Küste nach Süden (bei dieser Gelegenheit wurden die Kanarischen Inseln

sowie später die Inselgruppe um Madeira und die Azoren entdeckt). Im Verlauf des 15. Jahrhunderts erkundete man hauptsächlich den Seeweg um Afrika nach Indien, der dann von Vasco da Gama im Jahre 1498 auch gefunden wurde und eine Route von Europa nach Indien und China eröffnete, die nicht durch islamisch kontrolliertes Gebiet führte. Eine solche Umsegelung Afrikas war möglich, ohne über einen längeren Zeitraum hinweg die Küste als Anhaltspunkt für die Ortsbestimmung aus dem Blickfeld zu verlieren. Allerdings war diese Schiffsroute sehr lang und wegen der vielen Piraten in der Nähe des Horns von Afrika auch recht gefährlich.

Die andere Möglichkeit sah man in einer direkten Route von Europa nach Indien auf einem Weg in Richtung Westen. Diesen Weg wollte Columbus finden und entdeckte auf diese Weise 1492 das heutige Amerika (genauer entdeckte er eine Insel der Bahamas). Während eine Umfahrung Afrikas noch mit visuellem Küstenkontakt möglich war, musste man für die Westroute die vertrauten Küsten Europas und Afrikas verlassen. Nachdem Columbus die Bahamas entdeckt hatte und in der Folgezeit auch Länder auf dem amerikanischen Festland (sowohl in Nord- als auch Südamerika) entdeckt wurden, und nachdem Ferdinand Magellan um 1520 die erste Umsegelung Südamerikas und die Überquerung des Pazifischen Ozeans gelang, wurde das Problem der genauen Ortsbestimmung der neu entdeckten Länder und Inseln wichtig. Zum Einen wollten die Herrscher, die solche Erkundungsfahrten der Seefahrer finanzierten und die dabei entdeckten Länder für sich in Anspruch nahmen, wissen, wo genau sich diese Länder befanden, d.h. es kam die Frage auf, wie man diese Länder wiederfinden kann. Zum Anderen war es auch für Seefahrer wichtig zu wissen, wo genau man sich befand, um beispielsweise bei Unwetter oder in der Nacht zu vermeiden, auf Felsen oder Land aufzulaufen.

Eine Bestimmung des Breitengrads war auch auf See mit hinreichender Genauigkeit möglich, sofern das Wetter es zuließ. Entweder konnte man bei Tag den Sonnenhöchststand oder bei Nacht die Sterne beobachten und ausmessen. Der Höchststand der Sonne bei Tag oder beispielsweise die Höhe des Polarsterns bei Nacht erlaubten eine direkte Messung des Breitengrads. Problematischer war die Bestimmung des Längengrads: Hierzu musste man entweder eine direkte Messung vornehmen, d.h. man bestimmte die zurückgelegte Strecke aus der Geschwindigkeit des Schiffs und der Reisedauer - auf See war das nur bei ruhigem Wetter und ruhiger See (ohne Strömungen) möglich - oder man musste die genaue Uhrzeit an einem Referenzpunkt, dessen Längengrad bekannt war, kennen.

Eine dritte Möglichkeit, die vermutlich bei den ersten Pazifiküberquerungen verwendet wurde, bestand darin, unter einem konstanten Winkel zum Längengrad (im 16. Jahrhundert gab es schon recht gute Magnetkompassse sowie Kenntnisse zu den Abweichungen zwischen magnetischem und geographischen Nordpol) zu segeln und dann aus der Änderung des Breitengrads auf die Änderung im Längengrad zu schließen. Dieses Verfahren funktioniert nicht bei einer reinen Ost-West-Fahrt, also entlang eines konstanten Breitengrads. Solche Routen wurden gerne verwendet, da sich der Breitengrad leicht bestimmen ließ und diese Routen somit gut reproduzierbar waren.

Direkte Messungen waren schwierig, da der Fehler kumulativ ist (d.h., kleine Fehler in den Messungen - insbesondere systematische Fehler - addieren sich über einen längeren Zeitraum hinweg zu einem großen Fehler) und daher sehr anfällig für Ungenauigkeiten, beispielsweise bei nicht idealen Wetterverhältnissen oder Meeresströmungen, die es schwer bis unmöglich machten, die Geschwindigkeit des Schiffs zu bestimmen. Daher sah man die Lösung nur in einer genauen Zeitmessung: Wenn man an einem bestimmten Ort den genauen Zeitpunkt des Sonnenhöchststands misst und mit der gleichzeitigen lokalen Zeit an einem Referenzort (z.B. dem Heimathafen) vergleicht, kann man aus der Zeitdifferenz den Längengrad des momentanen Orts bestimmen.

Da sich die Erde in 24 Stunden einmal um ihre Achse dreht, ein Ort am Äquator in dieser Zeit somit 40 000 km „zurücklegt“, entspricht dies pro Minute einer Strecke von 27,7 km. Ein Längengrad am Äquator entspricht rund 111,11 km, und eine Längenminute rund 1 852 m, einer nautischen Meile. Für eine angestrebte Genauigkeit im Bereich von 5 bis 10 Kilometern bzw. eine Längengradmessung

mit einem Fehler zwischen 3 und 5 Bogenminuten musste die Zeit also auf rund 15 Sekunden bekannt sein. Das war mit den Uhren im späten Mittelalter oder der frühen Neuzeit kaum möglich. Selbst unter idealen Bedingungen auf festem Boden in abgeschlossenen Räumen erreichte man mit den Pendeluhrn im 17. Jahrhundert bestenfalls eine Genauigkeit von 15 Sekunden am Tag. Auf hoher See, wo das Schiff Stürmen ausgesetzt war oder auch extremen Temperatur- und Feuchtigkeitsschwankungen, eine Genauigkeit von 15 Sekunden über einen längeren Zeitraum (von z.B. 10–14 Tagen) zu erreichen, schien Anfang des 18. Jahrhunderts noch unmöglich. Nachdem bei einem großen Seeunglück der englischen Flotte im Jahre 1707 bei den Scilly Islands fast 2000 Seeleute ums Leben gekommen waren und vier Schiffe zerstört wurden, was auf eine fehlerhafte Ortsbestimmung des Kapitäns zurückging, setzte die britische Regierung im sogenannten *Longitude Act* 1714 eine Belohnung von letztendlich insgesamt 20 000 englischen Pfund (das entspricht heute rund 1,5–2 Millionen Euro) aus für denjenigen, der das Problem der Messung des Längengrads auf eine einfache und praktische Weise lösen konnte.

Zwei Ansätze wurden in diesem Zusammenhang verfolgt: (1) Die Erstellung sehr genauer Ephemeridentafeln (z.B. auch von den Jupitermonden) und (2) die Konstruktion einer Uhr, die über einen längeren Zeitraum auch unter den Bedingungen auf See eine Ortsbestimmung auf wenige Kilometer ermöglichte. Der erste Ansatz führte im 18. Jahrhundert zur Einrichtung vieler Sternwarten, unter anderem auch der Sternwarte von Greenwich. Auch die Arbeiten Newtons zur Himmelsmechanik können vor dem Hintergrund des Längengradproblems gesehen werden. Der zweite, letztendlich erfolgreiche Ansatz war die Konstruktion von Chronometern, also sehr genauen und gegen äußere Einflüsse weitgehend unempfindlichen Uhren, insbesondere der sogenannten H4 und H5 von John Harrison, der nach langem Kampf um 1775 einen Großteil des Preisgelds in Empfang nehmen durfte. Eine ausführliche und sehr lesbare Schilderung des Längengradproblems und seiner Lösung findet man in dem Buch von Dava Sobel [6].

## 4 Moderne Uhren

Da es hier nicht um eine ausführliche Darstellung der Geschichte der Zeitmessung gehen soll, seien nur die wichtigsten Entwicklungen im 19. und 20. Jahrhundert erwähnt.

Um 1850 entstanden die ersten elektrischen Uhren, bei denen die Anregung der Schwingung nicht mehr mechanisch über das langsam herabsinkende Gewicht gesteuert wurde, sondern durch einen elektrischen Strom, beispielsweise aus einer Batterie. Zwei Entwicklungen führten dabei zu den heutigen Atomuhren: (1) die Einführung des „Master-Slave-Prinzips“, bei dem Taktgeber - der Mechanismus, über den die Uhrzeit gemessen wird - und Takthalter - der Mechanismus, der für eine möglichst gleichmäßige Schwingung sorgt - getrennt wurden, was zu einer besseren Genauigkeit führte, und (2) die Ausnutzung von zunächst Kristallschwingungen (die Quarzuhr) und später atomaren Schwingungen (die Atomuhr).

### 4.1 Das Master-Slave-Prinzip

In den alten mechanischen Uhren, beispielsweise den Pendeluhrn der frühen Neuzeit, diente der periodische Prozess - das Schwingen des Pendels - sowohl zur Vorgabe einer möglichst gleichmäßigen Taktzeit, als auch zum Ablesen der Uhrzeit. Die Hemmung, beispielsweise die Ankerhemmung, übertrug dem Pendel bei jedem Takt eine minimale Energie von dem Kronrad, das durch ein Gewicht in Drehung versetzt wurde, sodass das Pendel nicht zur Ruhe kam. Gleichzeitig diente das sich drehende Kronrad aber auch als Ausgangspunkt für die rotierenden Zeiger der Uhr. Dieser Mechanismus war anfällig für Störungen und minimale Schwankungen.

Bei Master-Slave-Prinzip (zum ersten Mal umgesetzt in einer sogenannten Shortt-Uhr, benannt nach dem englischen Ingenieur William Hamilton Shortt) gibt es einen Takthalter, ein sogenanntes Primärpendel, das möglichst störungsfrei dieselbe Schwingungsperiode hält, und einen für das Ablesen der Uhrzeit bestimmten Taktgeber, in diesem Fall ein sogenanntes Sekundärpendel, über dessen Bewegung ein Zeigerwerk in Gang gesetzt wird, an dem die Uhrzeit abgelesen werden kann. Das Primärpendel befand sich bei der Shortt-Uhr in einem Vakuumzylinder, es erhielt in regelmäßigen (aber seltenen) Abständen einen wohldefinierten Impuls und sandte, ebenfalls in regelmäßigen nicht zu häufigen Abständen ein elektrisches Signal an das Sekundärpendel, das auf diese Weise gesteuert wurde und im selben Takt wie das Primärpendel schwang. Durch den Ablesemechanismus und andere äußere Einflüsse konnte das Sekundärpendel zwar in seiner Bewegung gestört werden, durch die regelmäßigen Pulse des Primärpendels wurde es aber immer wieder in Phase zum Primärpendel gebracht.

Dieses Prinzip - es wird nicht die Schwingung des Primäroszillators ausgelesen sondern die Schwingung eines Sekundäroszillators, der von dem Primäroszillator gesteuert wird - findet sich auch in späteren Präzisionsuhren, beispielsweise auch den heutigen Atomuhren. Die Shortt-Uhren waren die ersten mechanischen Uhren, mit denen um 1930 Schwankungen in der Erdrotation und damit Schwankungen in der mittleren Tageslänge nachgewiesen werden konnten.

## 4.2 Quarzuhren

1927 wurde die erste Quarzuhr vorgestellt. Nachdem um 1880 von den Brüdern Jacques und Pierre Curie die Piezoelektrizität entdeckt worden war - also dass die Anlegung einer Spannung an manche Kristalle zu einer leichten Verformung dieser Kristalle führt, und umgekehrt, dass eine mechanische Verformung der Kristalle eine Spannung erzeugt - gelang es in den 20er Jahren des 20. Jahrhunderts, diese im Hochfrequenzbereich zur Erzeugung von Quarzschwingungen und zum Auslesen dieser Schwingungen zu nutzen und damit die ersten Quarzuhren zu konstruieren. Den größten Einfluss auf die Genauigkeit hatte ihre Temperaturempfindlichkeit, die in den nächsten Jahren zunehmend unter Kontrolle gebracht werden konnte.

## 4.3 Atomuhren

Atomuhren nutzen die harmonischen Schwingungen im Zusammenhang mit Übergängen zwischen atomaren Zuständen. Bei den älteren Atomuhren, einschließlich den Cäsium-Uhren, handelt es sich dabei um Übergänge im Radiowellenbereich, meist Übergänge zur Hyperfeinstruktur. Sowohl bei Wasserstoff-Uhren als auch Cäsium-Uhren betrachtet man den Hyperfeinstrukturübergang zwischen zwei elektronischen Spinzuständen relativ zum Kernspin: Die Atome haben ein einzelnes Elektron in der äußeren Schale, dessen Spin parallel oder antiparallel zum Kernspin sein kann. Die Energie zwischen diesen beiden Zuständen unterscheidet sich nur minimal und entspricht langwelligen Strahlungen: Bei Wasserstoff handelt es sich um die in der Astronomie bekannte 21-cm-Strahlung - der Übergang entspricht einer Frequenz von 1 420 MHz bzw. einer Energie von  $5,87 \mu\text{eV}$  und somit hat die Strahlung im Vakuum eine Wellenlänge von 21 cm - und bei Cäsium handelt es sich um die zu Beginn dieses Kapitels angegebenen Werte.

Cäsium hat mehrere Vorteile: Einerseits gibt es von Cäsium nur ein stabiles Isotop (Cs-133), sodass der Einfluss unterschiedlicher Isotope in einer Probe vernachlässigt werden kann, Cäsium hat ein einzelnes Elektron in seiner äußeren Schale und sein Siedepunkt ist mit rund  $690^\circ\text{C}$  vergleichsweise niedrig. Schon bei rund  $100^\circ\text{C}$  gibt es thermisch verdampfende Atome, die sich für den Betrieb einer Atomuhr eignen.

In einem Resonator werden die Cäsiumatome zu Übergängen angeregt (Abb. 4), wobei der

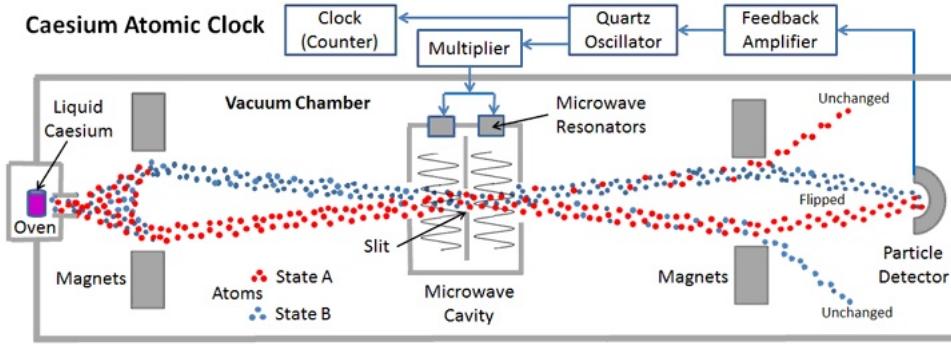


Abbildung 4: Prinzip einer Cäsium-Uhr: Aus einem Ofen strömen Cäsium-Atome nahezu gleichverteilt im Grundzustand und im angeregten Hyperfeinstrukturzustand. In einem Magnetfeld werden die Atome im Grundzustand von denen im angeregten Zustand getrennt (sie haben unterschiedliche magnetische Momente) und treten dann unter einem unterschiedlichen Winkel in einen Resonator. In diesem Resonator werden sie einem oszillierenden Feld ausgesetzt, dessen Frequenz über eine Rückkopplung auf die Eigenfrequenz des Übergangs eingestellt wird. Dadurch werden die Atome zu Übergängen angeregt. In einem Detektor hinter dem Resonator wird die Anzahl der Atome gemessen, die einen Übergang gemacht haben - je genauer die Resonatorfrequenz auf die Übergangsfrequenz eingestellt ist, umso mehr Atome werden zu einem Übergang angeregt. Der Detektor steuert über die Rückkopplung die Frequenz des Resonators, sodass möglichst viele Atome den Detektor erreichen. Abgelesen wird die Frequenz des Resonators. (aus [2])

Anteil der Atome, die einen Übergang machen, umso höher ist, je näher die Frequenz des Resonatorfeldes an der Übergangsfrequenz liegt. Die Frequenz des Resonatorfeldes wird von einem Quarzkristall gesteuert, der wiederum von einem Detektor hinter dem Resonator gesteuert wird. Dieser Detektor misst die Anzahl der Atome, die einen Übergang gemacht haben, und er steuert den Quarzkristall so, dass diese Anzahl maximal bleibt. An dem Quarzkristall wird dann auch die jeweilige Frequenz abgelesen und in eine Uhrzeit umgewandelt.

Moderne Atomuhren arbeiten bei optischen Frequenzen und erreichen damit nochmals eine um einen Faktor 1000 bis 10 000 höhere Genauigkeit. Sie befinden sich derzeit in der Testphase, es wird aber nicht ausgeschlossen, dass in einigen Jahren die Sekunde über solche Opticlocks neu festgelegt wird.

#### 4.4 Fountain-Clocks

Das bisher beschriebene Verfahren beruht auf Ideen von Isidor Isaac Rabi Mitte des letzten Jahrhunderts. Diese Ideen wurden von seinem „Schüler“ Norman Ramsey nochmals erweitert und zu größerer Präzision geführt. Eine wesentliche Verbesserung besteht darin, die Cs-Atome zweimal für einen kurzen Augenblick durch einen Resonator zu schicken, wobei die Präzession der Uhr umso größer ist, je mehr Zeit zwischen diesen beiden Augenblicken liegt. In sogenannten Fountain-Clocks („Springbrunnen-Uhren“) werden die Cäsiumatome zunächst abgekühlt (d.h. verlangsamt, dies geschieht mit Laserkühlung) und dann mit wenigen Metern pro Sekunde senkrecht nach oben gestoßen. Sie durchlaufen in einem freien Fall eine Parabelkurve, wobei sie zu Beginn und am Ende kurz durch einen Resonator fliegen. Auf diese Weise befinden sich die Cäsiumatome wesentlich länger in der Phase zwischen den Resonatoreinflüssen und die Genauigkeit der Frequenz kann besser reguliert werden.

## 5 Kuriositäten

### 5.1 Die tautochrone Kurve - Zykloide

Nachdem Galilei erkannt hatte, dass bei einem Pendel die Periode bei kleinen Auslenkungen nahezu unabhängig von der Amplitude des Pendels ist, hatte Christiaan Huygens 1656 die Idee, ein Pendel als Taktgeber einer Uhr zu verwenden. Ihm war jedoch bewusst, dass bei einem realen Pendel die Periode nicht vollkommen unabhängig von der Auslenkung ist und bei großen Auslenkungen doch deutliche Abhängigkeiten von der Amplitude auftreten. Huygens stellte sich daraufhin die Frage, wie man erreichen kann, dass die Periodendauer eines Pendels vollkommen unabhängig von der Auslenkung ist.

Das theoretische Problem besteht zunächst darin, die Bahnkurve zu bestimmen, die ein Massenpunkt durchlaufen muss, sodass die Periode der Schwingung unabhängig von der Auslenkung ist. Eine solche Bahnkurve bezeichnet man als tautochrone oder auch isochrone Bahnkurve. Eine Kreiscurve, wie sie bei einem Pendel konstanter Pendellänge durchlaufen wird, hat diese Eigenschaft nur bis auf Terme 4. Ordnung in der Auslenkung (z.B. im Auslenkungswinkel) und daher nur für kleine Auslenkungen. Das zweite, eher praktische Problem besteht in der Konstruktion eines Aufhängemechanismus, sodass die Masse an einem Fadenpendel tatsächlich die tautochrone Bahnkurve durchläuft.

Man kann analytisch mit Variationsrechnung die Form dieser Bahnkurve herleiten. Wir geben hier jedoch das Ergebnis an und beweisen, dass diese Bahnkurve die gewünschten Eigenschaften hat. Es handelt sich um eine sogenannte Zykloide, wie sie in Abb. 5 dargestellt ist. Eine Zykloide wird von einer Markierung am Rand eines Kreises beschrieben, wenn dieser Kreis an einer Geraden entlang rollt.

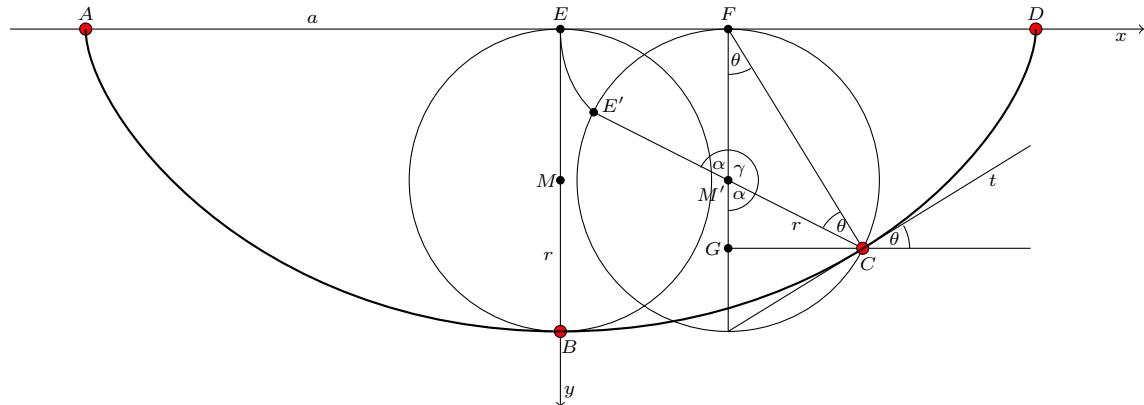


Abbildung 5: Die Zykloide. Wenn ein Kreis (Radius  $r$ ) auf einer geraden Linie  $a$  (gleichzeitig die  $x$ -Achse) abrollt, beschreibt ein Punkt  $B$  auf dem Rand des Kreises (rot eingezeichnet) den Bogen einer Zykloide. Es gilt immer  $\alpha = 2\theta$ , da sowohl  $\alpha + \gamma$  als auch  $\gamma + 2\theta$  einem Winkel von  $180^\circ$  entsprechen. Die Tangente  $t$  im Punkt  $C$  steht senkrecht auf der Verbindungslinie  $FC$  und schließt mit einer Waagerechten einen Winkel  $\theta$  ein. Der Nullpunkt des Koordinatensystems sei der Punkt  $E$ . Wenn der Kreis in diesem Punkt die Achse  $a$  berührt, sei  $B$  am Minimum der Kurve, also auf dem Kreis dem Punkt  $E$  gegenüber.

Wir wählen den Nullpunkt unseres Koordinatensystems als den Berührungsrand des Kreises mit der Geraden an der Stelle, wo die Markierung (Punkt  $B$ ) im Minimum ist. Rollt der Kreis um einen Winkel  $\alpha$  nach rechts, ist der Berührungsrand mit der Geraden der Punkt  $F$ . Die Distanz

zwischen  $E$  und  $F$  ist genau gleich der Bogenlänge zwischen dem ehemaligen Berührungspunkt (jetzt  $E'$ ) und dem Punkt  $F$ , also  $\overline{EF} = \alpha r$ , wobei der Winkel  $\alpha$  in Radianen ausgedrückt wird. Der Punkt  $B$  hat sich bis zum Punkt  $C$  weiterbewegt und dabei den Bogenabschnitt der Zykloide überstrichen.

Wir geben zunächst eine Parametrisierung der Zykloide an: Die  $x$ - und die  $y$ -Koordinate werden als Funktion des Winkels  $\alpha \in [-\pi, +\pi]$  beschrieben:

$$x(\alpha) = r(\alpha + \sin \alpha) \quad , \quad y(\alpha) = -r(1 + \cos \alpha). \quad (1)$$

Die  $x$ -Koordinate ergibt sich aus der Summe von zwei Beiträgen:  $\overline{EF} + \overline{GC} = r\alpha + r \sin \alpha$ , und die  $y$ -Koordinate ergibt sich aus:  $\overline{FM'} + \overline{M'G} = r + r \cos \alpha$  (in die negative  $y$ -Richtung).

Zum Beweis, dass die Zykloide tatsächlich die Lösung des Problems darstellt, ist zu zeigen, dass die Kraft proportional zur Bogenlänge der Auslenkung ist, dass es sich also um einen idealen harmonischen Oszillatator handelt. Dies unterscheidet die Zykloide von einer Parabel: Bei einer Parabel ist die Kraft proportional zur Auslenkung in  $x$ -Richtung (oder, was äquivalent ist, das Potenzial ist proportional zum Quadrat der Auslenkung in  $x$ -Richtung), bei einer Zykloide ist die Kraft proportional zur Bogenlänge. Die Richtung der Kraft ist tangential zur Kurve und somit ist

$$F = mg \sin \theta = mg \sin \frac{\alpha}{2}. \quad (2)$$

Die Bogenlänge entlang der Zykloide (also zwischen  $B$  und  $C$ ) erhalten wir aus

$$ds^2 = \left( \left( \frac{dx}{d\alpha} \right)^2 + \left( \frac{dy}{d\alpha} \right)^2 \right) (d\alpha)^2 \quad (3)$$

$$= (r^2(1 + \cos \alpha)^2 + r^2(\sin \alpha)^2) (d\alpha)^2 \quad (4)$$

$$= 2r^2(1 + \cos \alpha)(d\alpha)^2. \quad (5)$$

Mit der trigonometrischen Beziehung

$$1 + \cos \alpha = 1 + \cos^2 \frac{\alpha}{2} - \sin^2 \frac{\alpha}{2} = 2 \cos^2 \frac{\alpha}{2} \quad (6)$$

folgt

$$ds = 2r \cos \left( \frac{\alpha}{2} \right) d\alpha \quad (7)$$

und damit

$$s = 2r \int_0^\alpha \cos \left( \frac{\alpha'}{2} \right) d\alpha' = 4r \sin \frac{\alpha}{2} \Big|_0^\alpha = 4r \sin \frac{\alpha}{2}. \quad (8)$$

Für die Zykloide gilt also tatsächlich, dass die Bogenlänge proportional zur rücktreibenden Kraft ist und damit handelt es sich um einen idealen harmonischen Oszillatator mit einer von der Auslenkung unabhängigen Periode.

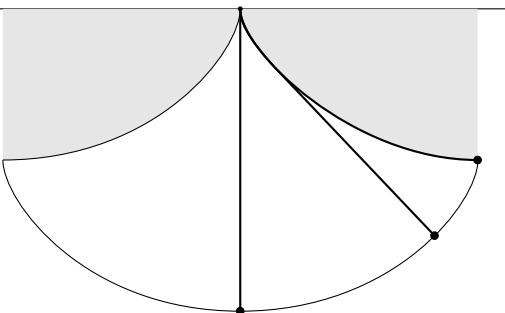


Abbildung 6: Die Huygens'sche Aufhängung. Die Zykloide löst gleichzeitig das zweite Problem bezüglich eines von der Auslenkung unabhängigen Pendels: Wenn das schwingende Pendel an seiner Aufhängung durch eine Zykloide begrenzt ist, so dass der obere Teil des Pendelseils an der Zykloide entlang läuft, folgt die Masse am Ende des Pendels einer Zykloide.

Das zweite Problem des Huygens'schen Pendels, die Aufhängung, wird ebenfalls durch die Zykloide gelöst. Abbildung 6 zeigt, wie man die Aufhängung eines Pendels durch eine Zykloide eingrenzen kann, sodass die Pendelmasse tatsächlich die Trajektorie einer Zykloide durchläuft. Dies bezeichnet man auch als die Huygens'sche Pendelaufhängung.

Huygens erhoffte sich durch einen solchen Mechanismus eine tautochrone Schwingung des Pendels auch unter extremen Bedingungen (z.B. auf einem dem Wind und Wellen ausgesetzten Schiff). Letztendlich war diese Idee nicht erfolgreich, statt dessen wurden bessere Hemmungen, die unter anderem kleinere Auslenkungen ermöglichen, entwickelt.

## Literatur

- [1] Till Niermann - Eigenes Werk, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=21792620>
- [2] aus: Electropaedia "Clock and Watch Movements", <https://mpoweruk.com/timekeepers.htm>
- [3] aus Wikipedia "Foliot". <https://de.wikipedia.org/wiki/Foliot#/media/Datei:Foliot.jpg>.
- [4] Spindel-Hemmung; <https://www.youtube.com/watch?v=UhFPb-ZZTyI>
- [5] O. Neugebauer; *A History of Ancient Mathematical Astronomy*, Studies in the History of Mathematics and Physical Sciences 1; Springer-Verlag Berlin Heidelberg GmbH, 1975.
- [6] Dava Sobel; *Longitude*, Fourth Estate, Second Printing 1995. Deutsch: *Längengrad*; Malik, National Geographic, 2013.

# Die Gezeiten - Ebbe und Flut

Thomas Filk, Universität Freiburg

Die Gezeiten - Ebbe und Flut - sind jedem bekannt, der mal an einer Ozeanküste war. Die Erscheinungen können jedoch sehr unterschiedlich sein: Meist erlebt man zweimal an einem Tag Flut und zweimal Ebbe, es gibt jedoch auch Küsten, an denen je nach Jahreszeit nur einmal am Tag Ebbe und Flut auftreten. Die Höhenunterschiede - der Tidenhub - können zwischen „kaum spürbar“ bis hin zu deutlich über 10 Metern schwanken. Der vermutlich höchste Tidenhub ist in der Bay of Fundy in Kanada. Dort wurden schon über 20 Meter gemessen.

Schon im Altertum war den Seefahrern bekannt, dass Ebbe und Flut irgendwie mit dem Stand von Mond und Sonne zu tun haben. Sowohl das deutsche Wort „Gezeiten“ als auch der Ausdruck „Tiden“ (niederdeutsch für „Zeiten“), der besonders in Norddeutschland üblich ist, deuten den engen Zusammenhang zur „Zeit“ an, der immer schon mit dem Stand der Gestirne in Verbindung gebracht wurde. Sonne und Mond sind für die Gezeiten verantwortlich, wobei - wie wir noch sehen werden - der Einfluss des Monds ungefähr doppelt so groß ist wie der Einfluss der Sonne.

Versucht man die Einzelheiten zu verstehen, erkennt man bald, dass die Gezeiten ein sehr komplexes Phänomen darstellen. Hier kann nur ein elementarer Einblick gegeben werden. Ausführlichere Informationen findet man z.B. in den Referenzen [1, 2, 3].

In Tabelle 1 sind die wichtigsten Größen zusammengefasst, die in diesem Kapitel benötigt werden.

Gravitationskonstante	$G = 6,67 \cdot 10^{-11} \frac{\text{m}^3}{\text{kg} \cdot \text{s}^2}$
Masse der Erde	$M_{\text{Erde}} = 5,97 \cdot 10^{24} \text{ kg}$
Masse des Monds	$M_{\text{Mond}} = 7,35 \cdot 10^{22} \text{ kg}$
Masse der Sonne	$M_{\odot} = 2 \cdot 10^{30} \text{ kg}$
Abstand Erde-Mond	$R_{EM} = 380\,000 \text{ km}$ (zwischen 363\,000 und 405\,500 km)
Abstand Erde-Sonne	$R_{ES} = 150\,000\,000 \text{ km}$
Erdradius	$R_{\text{Erde}} = 6\,375 \text{ km}$
Neigung der Erdachse zur Ekliptik	$\alpha = 23,44^\circ$

Tabelle 1: Die wichtigsten physikalischen Größen, die im Zusammenhang mit den Gezeiten auftreten. Es handelt sich um ungefähre bzw. gemittelte Angaben, die für eine grobe Abschätzung der Gezeitenkräfte ausreichen.

Anmerkung: Ich werde in diesem Kapitel oft von Fliehkräften sprechen, obwohl es sich dabei für viele nicht um wirkliche Kräfte handelt. Andererseits ist das Konzept der Kraft ohnehin ein Hilfskonstrukt, dessen „Wirklichkeit“, insbesondere im Zusammenhang mit der Gravitation, durchaus in Frage gestellt werden kann. Wer den Begriff Fliehkraft vermeiden möchte, kann dies immer durch „Richtungsänderung des Impulses“ ersetzen.

## 1 Gezeitenkräfte

Ausgangspunkt der Erklärungen sind immer die Gezeitenkräfte (engl. *tidal forces*) des Monds bzw. der Sonne. Gezeitenkräfte sind sogenannte „differenzielle Kräfte“, d.h., sie geben die Differenz eines

Kraftfelds bzw. die Differenz der Kräfte zwischen zwei Punkten an. Betrachten wir zunächst die gewöhnliche Schwerkraft.

Die Schwerkraft  $F$  eines Objekts der Masse  $M$  auf einen Gegenstand der Masse  $m$  im Abstand  $R$  ist

$$F = G \frac{Mm}{R^2}. \quad (1)$$

Bildet man  $F/m$  erhält man eine Beschleunigung. Dieser Wert ist unabhängig von der Masse  $m$  des „Probekörpers“:

$$a = G \frac{M}{R^2}. \quad (2)$$

Setzt man die Werte für die Gravitationskonstante  $G$ , die Masse des Monds  $M_{\text{Mond}}$  und den mittleren Abstand  $R_{EM}$  zwischen Erde und Mond ein, erhält man:

$$a_M = 6,67 \cdot 10^{-11} \frac{\text{m}^3}{\text{kg} \cdot \text{s}^2} \cdot \frac{7,35 \cdot 10^{22} \text{kg}}{3,8^2 \cdot 10^{16} \text{m}^2} \approx 3,4 \cdot 10^{-5} \frac{\text{m}}{\text{s}^2}, \quad (3)$$

wobei dieser Wert aufgrund der elliptischen Form der Mondbahn und dem damit verbundenen variierenden Abstand zwischen Erde und Mond zwischen  $2,98 \cdot 10^{-5} \frac{\text{m}}{\text{s}^2}$  und  $3,71 \cdot 10^{-5} \frac{\text{m}}{\text{s}^2}$  schwanken kann.

Entsprechend erhalten wir für den Einfluss Sonne:

$$a_{\odot} = 6,67 \cdot 10^{-11} \frac{\text{m}^3}{\text{kg} \cdot \text{s}^2} \cdot \frac{2 \cdot 10^{30} \text{kg}}{1,5^2 \cdot 10^{22} \text{m}^2} \approx 5,9 \cdot 10^{-3} \frac{\text{m}}{\text{s}^2}. \quad (4)$$

Der gravitative Einfluss der Sonne auf die Erde bzw. auf Gegenstände auf der Erde ist also über 170-mal größer als der Einfluss des Monds. Der führende, konstante Teil dieser Kraft wirkt jedoch auf alle Gegenstände auf der Erde gleichermaßen, d.h., wir spüren ihn nicht, da alle Gegenstände derselben Beschleunigung unterliegen und somit keine relativen Verschiebungen auftreten. Wir würden ihn spüren, wenn die Erde (durch was auch immer für einen überirdischen Mechanismus) in ihrem Zentrum an einem Punkt im Raum „festgehalten“ würde. Alle Gegenstände (insbesondere auch alle Wassermassen) würden in diesem Fall mit der obigen Beschleunigung zur Sonne hingezogen.

Für die Gezeiten sind jedoch die Gezeitenkräfte verantwortlich, d.h. die Unterschiede in der Schwerkraft des Monds (bzw. der Sonne) auf Gegenstände, die sich an verschiedenen Orten auf der Erde befinden. Der Unterschied zwischen der Gravitationsbeschleunigung des Monds auf den Schwerpunkt der Erde und einen Punkt an der Erdoberfläche, der dem Mond zugewandt ist, beträgt: ([Herleitung](#))

$$\Delta a = G \frac{M_{\text{Mond}}}{(R_{EM} - R_{\text{Erde}})^2} - G \frac{M_{\text{Mond}}}{R_{EM}^2} \approx G \frac{M_{\text{Mond}}}{R_{EM}^3} 2R_{\text{Erde}}. \quad (5)$$

Im letzten Schritt wurde nur der führende Term in  $R_{\text{Erde}}/R_{EM} \approx 1/60$  genommen, entsprechend kleiner sind die Korrekturen. Setzt man Zahlen für das Erde-Mond-System ein, erhält man für diese differenzielle Beschleunigung:

$$\Delta a_M \approx 1,14 \cdot 10^{-6} \frac{\text{m}}{\text{s}^2}, \quad (6)$$

wobei auch hier der Wert wieder zwischen  $0,94 \cdot 10^{-6} \frac{\text{m}}{\text{s}^2}$  und  $1,3 \cdot 10^{-6} \frac{\text{m}}{\text{s}^2}$  schwanken kann.

Wegen des wesentlich größeren Abstands zwischen Erde und Sonne und weil dieser Abstand kubisch eingeht, ist diese Beschleunigung nun für die Sonne kleiner:

$$\Delta a_{\odot} \approx 5 \cdot 10^{-7} \frac{\text{m}}{\text{s}^2}. \quad (7)$$

Während also die absolute Schwerkraft der Sonne auf die Erde rund 170 mal größer ist als die des Monds, ist die Gezeitenkraft an der Oberfläche der Erde rund 2,3 (schwankend zwischen 1,9 und

2,6) mal schwächer als die des Monds. Wie schon erwähnt, spüren wir die absolute Schwerkraft der Sonne und des Monds nicht, da sich die Erde auf ihrer Bahn „im freien Fall“ befindet. Die differenzielle Schwerkraft, also die Gezeitenkraft, ist jedoch wahrnehmbar, da der Schwerpunkt der Erde, und damit der Punkt im freien Fall, einer anderen Beschleunigung unterliegt als die Punkte an der Erdoberfläche. Die Punkte auf der dem Mond abgewandten Seite der Erde spüren eine entsprechend geringere Schwerebeschleunigung im Vergleich zum Mittelwert.

## 2 Der dem Mond abgewandte Gezeitenberg

Der dem Mond zugewandte Wasserberg der Gezeiten wird durch die höhere Gravitationskraft des Monds auf diese Wassermassen im Vergleich zum Erdschwerpunkt erklärt. Für den Wasserberg auf der dem Mond abgewandten Seite findet man zwei zunächst scheinbar verschiedene Erklärungen, nach denen einmal die höhere Fliehkraft an dieser Seite der Erde für den Wasserberg verantwortlich ist und einmal die schwächere Gravitationskraft. Beide Erklärungen sind richtig, allerdings muss man hier vorsichtig sein, keine Fehlvorstellungen zu generieren. Wir berechnen zunächst die Fliehkräfte an beliebigen Punkten der Erde und beschränken uns dabei auf das Erde-Mond-System, das für die Gezeiten den größten Einfluss hat. Die Effekte der Sonne lassen sich ebenso erklären und überlagern sich den Einflüssen des Monds.

### 2.1 Der Einfluss der Fliehkraft

Wir werden sehen, dass sich die Fliehkraft an jedem Punkt der Erde in zwei Anteile aufspalten lässt: Ein Anteil ist von einer Achse durch das Zentrum der Erde radial nach außen gerichtet, der zweite Anteil ist überall auf der Erde (und auch in ihrem Inneren) derselbe und bezieht sich auf die Bewegung des Erdzentrums um den Schwerpunkt des Erde-Mond-Systems. Bildet man die vektorielle Summe dieses zweiten Anteils der Fliehkraft und der Gravitationskräfte des Monds, bleiben gerade die Gezeitenkräfte mit einer Wirkung nach außen übrig. Diese erzeugen die Gezeiten.

#### 2.1.1 Der Schwerpunkt des Erde-Mond-Systems

Erde und Mond drehen sich um eine Achse, die durch den gemeinsamen Schwerpunkt  $D$  (Abb. 1) verläuft und senkrecht auf der Erde-Mond-Umlaufbahn steht. Der Schwerpunkt berechnet sich aus der Bedingung

$$r_1 M_1 = r_2 M_2 \quad \text{oder} \quad r_1 = \frac{M_2}{M_1 + M_2} R \quad \text{mit } R = r_1 + r_2. \quad (8)$$

Setzen wir für  $M_2$  die Masse des Monds, für  $M_1$  die Masse der Erde und für  $R = r_1 + r_2$  den Abstand Erde-Mond ein, erhalten wir für  $r_1$  - den Abstand vom Erdmittelpunkt  $Z$  zum Schwerpunkt  $D$  des Erde-Mond-Systems,  $r_1 = R_{ZD} \approx 4620\text{ km}$ . Das ist etwas weniger als 3/4-tel des Erdradius. Der Schwerpunkt des Erde-Mond-Systems liegt also innerhalb der Erde.

Für das System Erde-Sonne liegt dieser Schwerpunkt rund 450 km vom Zentrum der Sonne entfernt, also tief im Inneren der Sonne. Auch wenn sich die Situation für das Erde-Mond-System in dieser Hinsicht vollkommen vom Erde-Sonne-System unterscheidet, bleibt die Argumentation für die Gezeiten im Wesentlichen die Gleiche. Diese Argumentation hängt nicht von der genauen Lage des gemeinsamen Schwerpunkts ab.



Abbildung 1: Zwei Massen  $M_1$  und  $M_2$  drehen sich um einen gemeinsamen Schwerpunkt  $D$ . Die Abstände  $r_1$  und  $r_2$  ergeben sich aus dem Hebelgesetz.  $Z$  ist der Mittelpunkt der einen Masse (Erde).

### 2.1.2 Radiale Fliehkräfte

Wir stellen uns nun das Erde-Mond-System als einen starren Körper vor, bei dem sich Erde und Mond um eine feste Achse durch den gemeinsamen Schwerpunkt  $D$  drehen und sich dabei immer dieselbe Seite zuwenden. Für den Mond ist das richtig und wir werden in Abschnitt 1 auch eine Begründung dafür finden, für die Erde gilt dies jedoch nicht: Sie dreht sich zusätzlich noch um eine Achse durch ihren Mittelpunkt  $Z$ . Drehungen der Erde um eine Achse durch ihren Mittelpunkt haben aber (unter den hier angenommenen idealisierten Bedingungen einer kugelförmigen Erde) keinen Einfluss auf die Gezeiten, da ihr Effekt - die Fliehkraft zu dieser Drehung - in radialer Richtung von der Drehachse durch  $Z$  nach außen zeigt und im selben Abstand von der Drehachse auch denselben Wert hat. Diese Kräfte führen zu einer Abplattung der Erde, die dadurch am Äquator etwas dicker ist als entlang von Großkreisen durch die Pole.

### 2.1.3 Die Fliehkräfte auf der Erde

Der Schwerpunkt des Erde-Mond-Systems sei also ein fester Punkt der Erde, den wir mit  $D$  bezeichnen; er markiert eine Drehachse durch diesen Punkt (siehe Abbildung 2). Allgemein ist die Fliehkraft auf einen Gegenstand der Masse  $m$  an einem Punkt  $C$  durch

$$\vec{F} = m\omega^2 \vec{R}_{DC} \quad (9)$$

gegeben, wobei  $\vec{R}_{DC}$  der Verbindungsvektor von der Drehachse  $D$  zum Punkt  $C$  ist und  $\omega$  die Winkelgeschwindigkeit der Drehung bezeichnet (sie entspricht einer Umlaufzeit des Erde-Mond-Systems um den gemeinsamen Schwerpunkt). Auch hier bietet es sich an, die Beschleunigung  $\vec{a} = \omega^2 \vec{R}_{DC}$  aufgrund dieser Kraft zu betrachten. Da die Winkelgeschwindigkeit für alle punktförmigen Objekte auf der Erde dieselbe ist, spielt nur der Abstandsvektor  $\vec{R}_{DC}$  vom Drehzentrum  $D$  zum Punkt  $C$  eine Rolle. Auf der dem Mond zugewandten Seite der Erde (Punkt  $B$ ) ist dieser Abstand sehr klein,  $R_{DB} \approx 1755$  km, im Vergleich zur abgewandten Seite (Punkt  $A$ ),  $R_{DA} \approx 11000$  km. Im Zentrum der Erde heben sich die Fliehkraft und die Anziehungskraft des Mondes gerade auf. Oft heißt es nun, dass sich auf der dem Mond zugewandten Seite die größere Gravitationskraft des Mondes und die kleinere Fliehkraft addieren, während auf der abgewandten Seite die Fliehkraft größer sei, sodass eine Nettokraft übrig bliebe, selbst wenn man die kleinere Gravitationskraft abzieht. Diese Erklärung für die beiden Wasserberge ist so nicht ganz richtig, da die für die Gezeiten relevanten Fliehkräfte an allen Punkten der Erde gleich sind, wie die folgende Überlegung zeigt.

Dazu berechnen wir die Fliehbeschleunigung auf einen beliebigen Punkt  $C$  (er muss nicht an der Erdoberfläche liegen). Diese Fliehbeschleunigung ist durch

$$\vec{a} = \omega^2 \vec{R}_{DC} = \omega^2 (\vec{R}_{ZC} - \vec{R}_{ZD}) \quad (10)$$

gegeben (siehe Abb. 2).

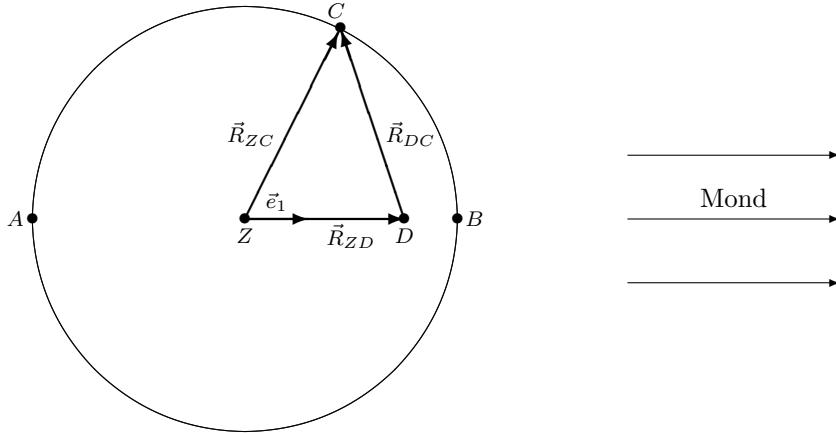


Abbildung 2: Die Fliehkraft bzw. -beschleunigung auf einen allgemeinen Punkt  $C$ . Der Schwerpunkt des Erde-Mond-Systems und damit der Mittelpunkt der Erde-Mond-Umlaufbahn ist der Drehpunkt  $D$ .  $Z$  bezeichnet den Mittelpunkt der Erde. Die Ansicht ist „von oben“, d.h., bei  $Z$  und  $D$  handelt es sich eigentlich um Drehachsen.

Wie man sieht, kann man diese Fliehbeschleunigung in zwei Anteile aufteilen: Ein Anteil zeigt vom Erdmittelpunkt  $Z$  radial nach außen (Richtung  $\vec{R}_{ZC}$ ) - dieser Anteil addiert sich zu der täglichen Drehung der Erde um ihre Achse und trägt nicht zur Gezeitenwirkung bei.<sup>1</sup> Der zweite Anteil ist unabhängig vom Punkt  $C$ , also für alle Punkte der Erde derselbe. Er ist immer parallel zur Verbindungslinie vom gemeinsamen Schwerpunkt  $D$  in die dem Mond abgewandte Richtung, d.h. in die Richtung der Achse durch den Erdmittelpunkt  $Z$ . Dieser zweite Anteil ist gleich der Gravitationskraft auf den Mittelpunkt  $Z$  der Erde. Für die Gravitationsbeschleunigung bedeutet das:

$$\vec{a}_Z = G \frac{M_{\text{Mond}}}{R_{EM}^2} \vec{e}_1 - \omega^2 \vec{R}_{ZD} = 0 \quad \text{oder} \quad G \frac{M_{\text{Mond}}}{R_{EM}^2} \vec{e}_1 = \omega^2 \vec{R}_{ZD}. \quad (11)$$

Hierbei ist  $\vec{e}_1$  ein Einheitsvektor von der zentralen Drehachse durch den Erdmittelpunkt  $Z$  in Richtung des Monds.

Bilden wir nun die Summe der beiden Beschleunigungen und nutzen dabei Gl. 11, erhalten wir für den Punkt  $C$ :

$$\vec{a}_C = 2G \frac{M_{\text{Mond}}}{R_{EM}^3} (\vec{e}_1 \cdot \vec{R}_{ZC}) \vec{e}_1. \quad (12)$$

Anmerkungen:

1. Wir haben es bei den obigen Betrachtungen mit drei verschiedenen Drehachsen zu tun: (1) die Drehachse der Erde durch ihren Mittelpunkt  $Z$  - sie ist um etwas über 23 Grad zur Ekliptik geneigt; (2) die Drehachse des Erde-Mond-Systems durch den Schwerpunkt  $D$ , wegen der Neigung der Mondumlaufbahn relativ zur Ekliptik von rund 5 Grad schwankt diese Neigung relativ zur Drehachse der Erde zwischen 18 und 28 Grad; (3) eine Achse parallel zur Drehachse Erde-Mond durch das Zentrum  $Z$  der Erde, um diese Achse dreht sich die Erde einmal monatlich bei einem Umlauf des Erde-Mond-Systems relativ zum Fixsternhimmel.

<sup>1</sup>Hier muss man eigentlich etwas vorsichtiger sein: Die tägliche Drehung der Erde erfolgt um ihre Rotationsachse, die relativ zur Ekliptik um 23,4 Grad geneigt ist. Die Drehung der Erde, von der hier die Rede ist, erfolgt einmal im Monat um eine Achse durch das Erdzentrum, die parallel zur Drehachse des Erde-Mond-Systems durch ihren gemeinsamen Schwerpunkt  $D$  ist. Diese beiden Achsen sind nicht identisch, auch wenn sie beide durch den Erdschwerpunkt verlaufen. Beide Drehungen haben keinen Einfluss auf die Gezeiten.

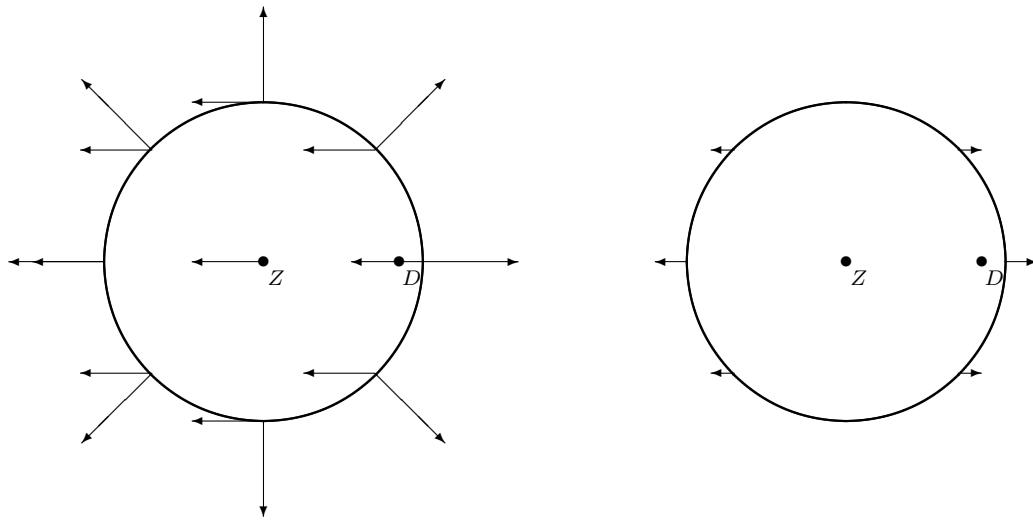


Abbildung 3: (links) Die Fliehkräfte bzw. -beschleunigungen lassen sich an jedem Punkt in zwei Anteile aufspalten: ein Anteil, der radial nach außen zeigt und proportional zum Abstand vom Erdmittelpunkt ist - dieser Anteil trägt nicht zu den Gezeiten bei. Ein zweiter Anteil, der an jedem Punkt der Erde derselbe ist und gleich der Fliehkraft auf das Zentrum  $Z$  der Erde. (rechts) Lässt man den radialen Anteil der Fliehkräfte weg - er trägt nicht zu den Gezeiten bei - und addiert man zu dem konstanten vom Mond weggerichteten Teil der Fliehkraft die Gravitationskraft des Monds, heben sich die Fliehkraft und der zentrale Teil der Gravitationskraft weg. Es bleiben nur die Gezeitenanteile der Gravitation. Diese sind für Ebbe und Flut auf der Erde verantwortlich.

2. Wir hatten schon mehrfach erwähnt, dass die Fliehkräfte zu Drehungen um Achsen durch den Erdmittelpunkt nicht zur Gezeitenwirkung beitragen, da diese Fliehkräfte radial von der Drehachse weg nach außen wirken und ihr Betrag nur vom Abstand von der Drehachse abhängt. Diese Kräfte sind also symmetrisch zur Drehachse. Durch die tägliche Drehung der Erde um ihre Achse wirkt am Äquator eine Beschleunigung von  $a = R_{\text{Erde}}\omega^2$ , wobei  $\omega$  einer Umdrehung am Tag entspricht, also  $\omega = (2\pi)/(24 \cdot 60 \cdot 60) \text{ s}^{-1}$ . Diese Beschleunigung beträgt rund  $a = 0,0337 \frac{\text{m}}{\text{s}^2}$ , ist also um ein Vielfaches größer als die Gezeitenkräfte. Diese Beschleunigung trägt zu einer Abplattung der Erde bei: Der Umfang der Erde am Äquator ist größer als entlang der Pole.

## 2.2 Die Gezeitenkräfte

Eine zweite Erklärung der Gezeiten betont einen anderen Gesichtspunkt, ist aber letztendlich äquivalent zu der Erklärung im letzten Abschnitt.

Wir stellen uns statt der Erde drei Objekte im Abstand von einem punktförmig angenommenen Massezentrum (z.B. dem Mond) vor. Der Einfachheit wegen sei dieses anziehende Massezentrum weit von diesen drei Objekten entfernt. Das mittlere der drei Objekte entspreche der Erde; zwei weitere Objekte - eines dem Mond zugewandt, das andere dem Mond abgewandt - entspreche Wassermassen auf der dem Mond zugewandten bzw. abgewandten Seite der Erde (siehe Abb. 4).

Auf alle drei Objekte wirkt in erster Näherung die Schwerkraft des Massezentrums. Bei dem vorderen Objekt (in Richtung des Massezentrums) kommt die Gezeitenkraft hinzu, da es näher am Massezentrum liegt, bei dem hinteren Objekt ist die Gesamtkraft um die Gezeitenkraft geringer, da es weiter vom Massezentrum entfernt ist. Würden diese drei Objekte im freien Fall auf das Massezentrum

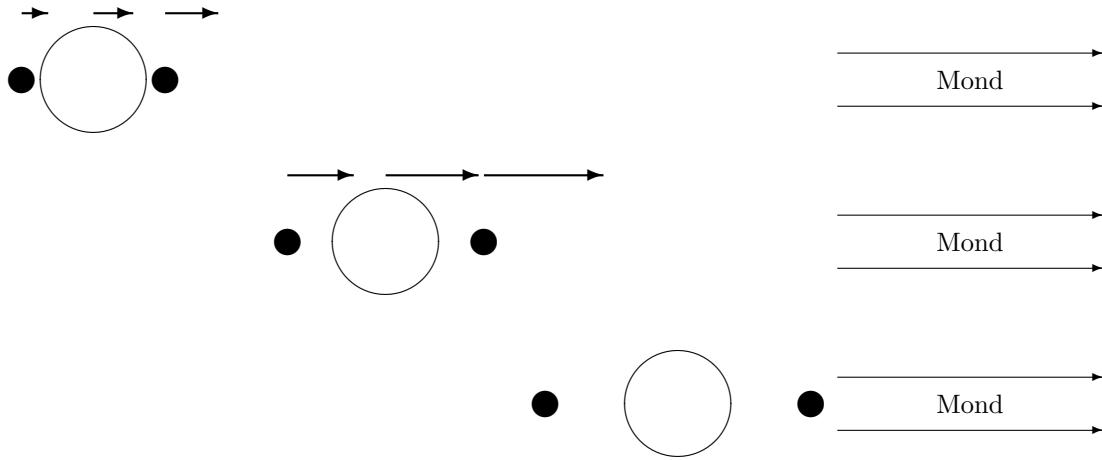


Abbildung 4: Die Gezeitenkräfte ziehen einen Gegenstand auseinander, sofern er nicht durch andere Kräfte zusammengehalten wird.

zufallen, würde sich der Abstand zwischen ihnen vergrößern: Das vordere Objekt fällt schneller, das hintere langsamer als das mittlere Objekt (die Erde als Ganzes).

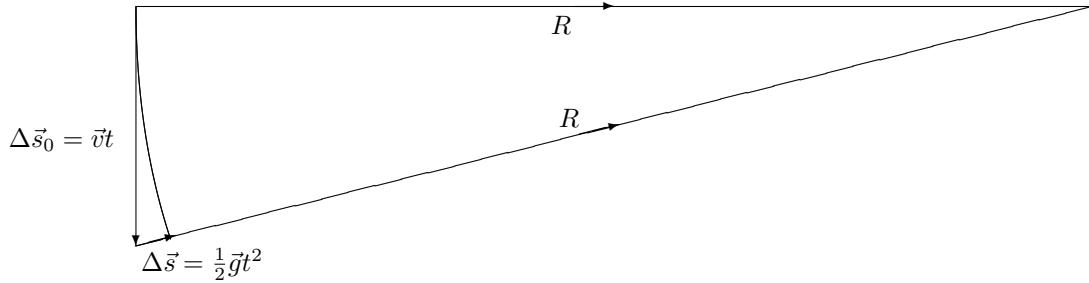


Abbildung 5: Eine Zentripetalkraft zieht einen Gegenstand von einer Geraden gerade so ab, dass die Beschleunigung diesen Gegenstand auf einer Kreisbahn hält.

Zu diesem freien Fall kommt nun eine Kreisbewegung hinzu, die gerade so ist, dass die Beschleunigung des mittleren Objekts dieses auf der Kreisbahn hält. Die Bedingung dafür ist, dass die in der (infinitesimalen) Zeitdauer  $t$  zurückgelegte tangentiale Strecke  $\Delta\vec{s}_0 = \vec{v}t$  abhängig der Strecke aufgrund der Beschleunigung ( $\Delta\vec{s} = \frac{1}{2}\vec{g}t^2$ ) zum Zentrum der Zentripetalkraft gerade wieder dem Abstand  $R$  von diesem Zentrum entspricht (siehe Abb. 5). Nach dem Satz von Pythagoras gilt somit:

$$R^2 + (vt)^2 = \left(R + \frac{1}{2}gt^2\right)^2 = R^2 + Rgt^2 + \frac{1}{4}g^2t^4. \quad (13)$$

Vernachlässigen wir den Term  $t^4$ , da  $t$  infinitesimal klein sein soll, folgt die Bedingung:

$$(vt)^2 = Rgt^2 \quad \text{oder} \quad g = \frac{v^2}{R}. \quad (14)$$

Für einen vollen Umlauf benötige der Körper die Zeit  $T$ , in der die Strecke  $U = 2\pi R$  (der Umfang der Kreisbahn) zurückgelegt wird. Es ist also  $vT = 2\pi R$  oder  $v = \omega R$  mit der Umlauf(winkel)frequenz  $\omega = 2\pi/T$ . Damit ergibt sich schließlich die Bedingung, die schon mehrfach verwendet wurde:

$$g = \omega^2 R. \quad (15)$$

Das dem Massezentrum zugewandte Objekt bewegt sich auf einer kleineren Kreisbahn, d.h., bei ihm ist die Anziehung durch die Masse etwas größer als die Kreisbeschleunigung. Bei dem abgewandten Objekt ist es umgekehrt: Seine Kreisbahn hat einen größeren Radius, bei ihm ist somit die Anziehung durch das Massezentrum etwas geringer als es seiner Kreisbeschleunigung entspricht. Es würde also nach Außen getrieben, wenn es nicht durch andere Kräfte an die mittlere Masse gebunden wäre.

### 3 Spring- und Nipptide

Die Einflüsse von Sonne und Mond überlagern sich, sodass die Gezeitenkräfte besonders intensiv sind, wenn Sonne, Erde und Mond auf einer Linie liegen. Dabei ist es zunächst nicht wichtig, ob sich Sonne und Mond von der Erde aus gegenüberliegen (also Vollmond ist), oder ob Sonne und Mond von der Erde aus auf einer Seite sind (also bei Neumond). In beiden Fällen erhält man eine sogenannte Springflut oder Springtide. Diese tritt rund zweimal in einem Monat auf.

Andererseits ist der Einfluss auf die Gezeiten besonders schwach, wenn Sonne und Mond von der Erde aus betrachtet unter einem Winkel von 90 Grad erscheinen, d.h. bei zunehmendem oder abnehmendem Halbmond. Man spricht in diesem Fall von einer Nipptide.

Die Intensität einer Springflut kann durch verschiedene Faktoren noch verstärkt werden. Zum einen handelt es sich um rein geometrische Faktoren im Sonnen- und Mondstand: Zum Beispiel, wenn der Mond sich gerade in seiner Periapsis, also dem erdnächsten Punkt seiner elliptischen Umlaufbahn um die Erde (bzw. den gemeinsamen Schwerpunkt des Erde-Mond-Systems) befindet. Weitere wichtige Faktor sind allerdings Wetterbedingungen, z.B. wenn zeitgleich zur Springflut auch ein landeinwärtsiger Sturm weht.

### 4 Die Neigung der Erdachse

Da der Wasserstand gerade bei Hafeneinfahrten für die Schiffahrt von Bedeutung ist, wurde dieser teilweise schon seit Jahrhunderten gemessen. Diese Aufzeichnungen sind nicht nur für den Klimawandel (Anhebung des Meeresspiegels) von Bedeutung, sondern verdeutlichen auch die Komplexität der Gezeiten. In einer harmonischen Analyse (also einer Frequenzanalyse oder Fourier-Zerlegung) der Wasserstände kann man viele hundert Anteile erkennen. Neben dem idealisierten Einfluss von Sonne und Mond, die nicht exakt dieselbe Frequenz haben - ein Sonntag dauert 24 Stunden, ein Mondtag ist jedoch um rund 50 Minuten länger -, spielt auch die Elliptizität der Mond- und Sonnenbahnen eine wichtige Rolle. Außerdem kommen ortsabhängige Strömungsverhältnisse hinzu.

Ein wichtiger Faktor ist aber auch die Neigung der Erdachse um 23,5 Grad relativ zur Ekliptik sowie die Neigung der Mondumlaufbahn relativ zur Ekliptik von etwas über 5 Grad. Dieser Einfluss kann an manchen Orten der Erde und zu manchen Jahreszeiten dazu führen, dass nur eine Ebbe und eine Flut am Tag auftreten.

Wie man in Abb. 6 erkennt, gibt es Orte auf der Erdoberfläche, die zu bestimmten Zeiten, wenn die Erdachse zum Mond gerichtet ist (und natürlich auch, wenn sie von ihm weggerichtet ist), mitten im Flutberg befinden, wohingegen sie etwas über 12 Stunden später (ein Mond-Tag hat 24 h plus 50 min) vergleichsweise weit entfernt von dem gegenüberliegenden Flutberg sind. Das kann den Effekt haben, dass man an solchen Orten nur einmal am Tag eine Flut und nur einmal am Tag eine Ebbe wahrnimmt.

Allerdings ist dies nur ein geometrischer Effekt auf die Gezeiten. In Wirklichkeit spielen sehr viele Faktoren eine weitaus dominantere Rolle, wann, wo und wie intensiv Gezeiten an einem Ort auftreten. Insbesondere spielen die Tiefenverhältnisse des Meeres und auch die geographischen Verhältnisse der Küste eine sehr wichtige Rolle oder auch ob es sich um eine Ost- oder Westküste

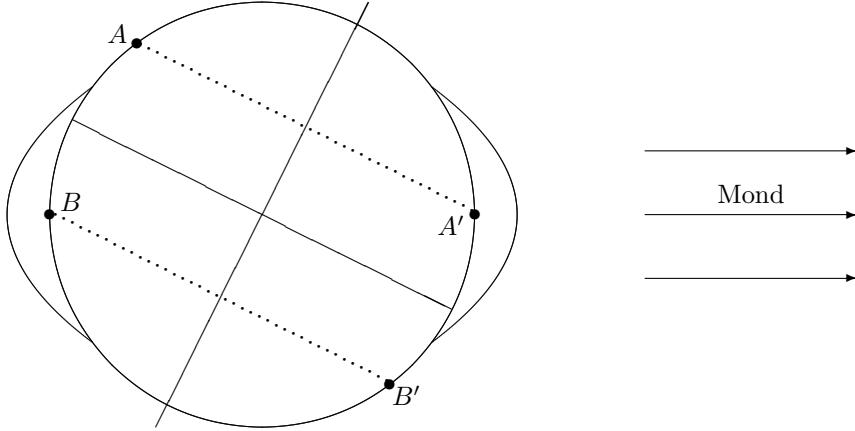


Abbildung 6: Durch die Neigung der Erdachse zur Ekliptik ist einer der Flutwulste oberhalb des Äquators, der gegenüberliegende unterhalb. An bestimmten Orten, z.B. A oder B, kann es vorkommen, dass nur einer der beiden Flutberge auftritt. Zwischen den jeweiligen Positionen, A bzw. B und A' bzw. B' liegen etwas über 12 Stunden.

handelt. Auch die Dynamik der Strömungen spielt hier eine wichtige Rolle. Tritt an einem Ort nur einmal innerhalb von 24 Stunden eine Flut oder Ebbe auf, spricht man von diurnalen Gezeiten, ansonsten von semi-diurnalen Gezeiten.

## 5 Herleitung einiger Gleichungen

### 5.1 Gezeitenbeschleunigung

Berechnet werden soll in führender Ordnung (Gl. 5)

$$\Delta a = G \frac{M_{\text{Mond}}}{(R_{EM} - R_{\text{Erde}})^2} - G \frac{M_{\text{Mond}}}{R_{EM}^2}. \quad (16)$$

Dazu wird aus dem ersten Term auf der rechten Seite der Abstand Erde-Mond ausgeklammert,

$$G \frac{M_{\text{Mond}}}{(R_{EM} - R_{\text{Erde}})^2} = G \frac{M_{\text{Mond}}}{R_{EM}^2} \frac{1}{(1 - \frac{R_{\text{Erde}}}{R_{EM}})^2}, \quad (17)$$

und der hintere Term in Potenzen von  $\frac{R_{\text{Erde}}}{R_{EM}}$  entwickelt. Die Entwicklung lautet allgemein:

$$\frac{1}{(1-x)^2} = 1 + 2x + O(x^2). \quad (18)$$

Damit folgt:

$$G \frac{M_{\text{Mond}}}{(R_{EM} - R_{\text{Erde}})^2} = G \frac{M_{\text{Mond}}}{R_{EM}^2} \left( 1 + 2 \frac{R_{\text{Erde}}}{R_{EM}} + \dots \right) \quad (19)$$

und wir erhalten für die Gezeitenbeschleunigung (die führenden Terme heben sich weg):

$$\Delta a = G \frac{M_{\text{Mond}}}{(R_{EM} - R_{\text{Erde}})^2} - G \frac{M_{\text{Mond}}}{R_{EM}^2} = 2G \frac{M_{\text{Mond}}}{R_{EM}^2} \left( \frac{R_{\text{Erde}}}{R_{EM}} + O \left( \left( \frac{R_{\text{Erde}}}{R_{EM}} \right)^2 \right) \right). \quad (20)$$

## Literatur

- [1] Hicks, S.D.; *Understanding Tides*, NOAA Report; 2006; <https://repository.oceanbestpractices.org/handle/11329/594> (aufgerufen am 13.5.2023).
- [2] Kowalik, Z., Luick, J.L.; *Modern Theory and Practice of Tide Analysis and Tidal Power*, Austides Consulting, Eden Hills, 2019; <https://austides.com/downloads/> (aufgerufen am 13.5.2023).
- [3] Parker, B.B.; *Tidal Analysis and Prediction*; NOAA Special Publication NOS CO-OPS 3; 2007; <https://repository.oceanbestpractices.org/handle/11329/632> (aufgerufen am 13.5.2023).

# Gezeiten und Tageslänge

Thomas Filk, Universität Freiburg

In diesem Kapitel wird das Phänomen der Gezeiten (typischerweise zwei Flutberge und zwei Ebbe-senken am Tag, vornehmlich in Richtung des Mondes) und ihre Entstehung vorausgesetzt. Es wird gezeigt, wie dieses Phänomen zu einer Verlangsamung der Erddrehung und damit zu einer Verlängerung der Tage führt, und dass dies zur Folge hat, dass sich der Mond langsam von der Erde entfernt.

## 1 Die Tage werden länger

Die Erde dreht sich einmal täglich um ihre Achse. Das ist im Vergleich zu einem Erde-Mond-Umlauf sehr schnell. Da es sich um eine Drehung um eine Achse durch das Zentrum der Erde handelt, hat diese keinen Einfluss auf die Entstehung der Gezeiten. Allerdings wirken sich die Gezeiten auf diese Drehung aus: Der unebene Meeresgrund sowie die kontinentalen Küsten bewirken, dass es eine Reibung zwischen Erde und Wassermassen gibt, die sogenannte Gezeitenreibung. Anders ausgedrückt, die Wassermassen der Gezeiten treffen auf die Ostküsten der Kontinente mit einer etwas größeren Wucht als auf die Westküsten. Dadurch wird die Erde in ihrer relativen Drehung zu den Wasserbergen der Gezeiten abgebremst.

Ein ähnlicher Effekt hat vermutlich auch dazu geführt, dass der Mond, der ursprünglich sicherlich eine Eigendrehung relativ zur Erde hatte, abgebremst wurde und mittlerweile der Erde immer dieselbe Seite zuwendet. Hier waren es zwar keine Wassermassen, doch durch den Gezeiteneffekt von der Erde auf den Mond wurden dort die Landmassen leicht angehoben. Einen ähnlichen Effekt gibt es auch bei den Landmassen der Erde, er macht aber nur rund 30 Zentimeter Höhenunterschied aus. Der Mond hat dadurch im Verlauf der Zeit an Rotationsenergie verloren und dreht sich heute nur noch einmal im Monat um seine Achse, wobei er der Erde immer dieselbe Seite zeigt.

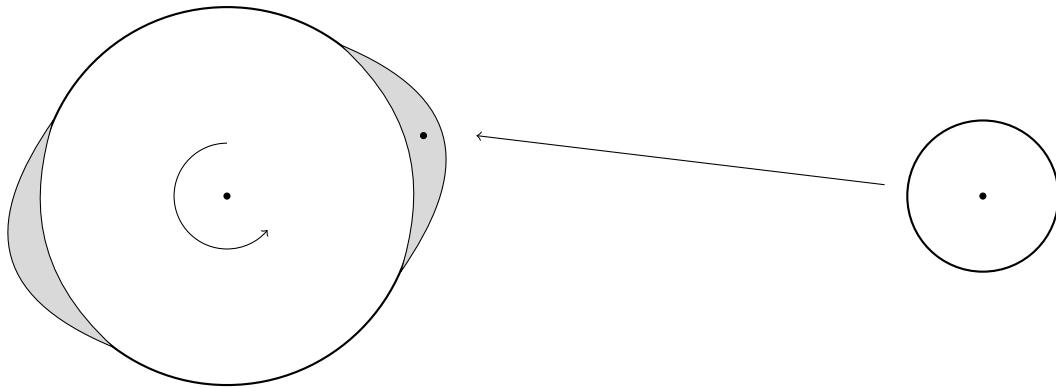


Abbildung 1: Durch die Eigendrehung der Erde relativ zu den Flutbergen kommt es zu einer „Reibung“, bei der die Erde abgebremst wird, wodurch die Tage länger werden. Andererseits haben die voreilenden Flutberge eine beschleunigende Wirkung auf den Mond, der dadurch an Energie gewinnt und sich von der Erde entfernt. Der Drehimpuls - vorher in der Eigenrotation der Erde, nachher in der Bewegung des Monds - bleibt erhalten.

Durch die raschere Drehung der Erde werden die Wasserberge der Flut etwas vorangetrieben, sodass diese nicht mehr auf der Verbindungslinie zwischen Erde und Mond liegen, sondern etwas davor

(siehe Abb. 1). Auf der einen Seite bewirkt nun die Anziehungskraft des Monds auf diese Wasserberge das Abbremsen der Erdumdrehung, auf der anderen Seite - im Gegenzug - wird der Mond durch die Anziehungskraft der Wasserberge etwas beschleunigt. Das hat einerseits den Effekt, dass die Tage auf der Erde länger werden (die Erde dreht sich langsamer - das macht in hundert Jahren rund 2 Millisekunden pro Tag aus, d.h. die Dauer eines Tages hat in den letzten 150-200 Millionen Jahren um rund eine Stunde zugenommen), andererseits nimmt der Abstand des Monds von der Erde zu. Im Jahr sind das derzeit rund 3,8 Zentimeter.

Der Drehimpuls der Erde nimmt zwar durch die Bremswirkung der Gezeiten langsam ab, aber dadurch nimmt der Bahndrehimpuls des Monds (bzw. genauer des Erde-Mond-Systems) langsam zu. Insgesamt bleibt der Gesamtdrehimpuls des Erde-Mond-Systems erhalten.

## 2 Ein paar „ $\pi \times$ Daumen“-Rechnungen

### 2.1 Die Zunahme der Tageslänge

Der Drehimpuls des Erde-Mond-Systems steckt in erster Linie in der Mondbahn (Drehung des Monds um den gemeinsamen Schwerpunkt) und in zweiter Linie in der Eigendrehung der Erde. Der Bahndrehimpuls der Erde und die Eigendrehung des Monds können vernachlässigt werden. Tabelle 1 gibt einige Größenordnungen an. Die Bahndrehimpulse von Erde und Mond berechnen sich nach

$$L_{\text{Bahn}} = mrv = mr^2\omega = mr^2\frac{2\pi}{T}, \quad (1)$$

( $m$  die Masse des Objekts,  $r$  der Abstand Schwerpunkt-Mittelpunkt des Objekts,  $v$  die Geschwindigkeit des Objekts,  $\omega$  die zugehörige Winkelgeschwindigkeit).<sup>1</sup> Die Umlaufperiode  $T$  ist die siderische Periode der Mondbahn, das sind ungefähr 27,3 Tage. Die Eigendrehimpulse wurden nach

$$L_{\text{Eigen}} = \frac{2}{5}mR^2\omega, \quad (2)$$

berechnet ( $m$  Masse,  $R$  Radius des Objekts,  $\omega$  bzw.  $T$  die Eigenfrequenz bzw. Periode der Drehung; für den Mond ist  $T = 27,3$  Tage, für die Erde ist  $T = 1$  Tag oder 86 400 Sekunden). Hier wird die Annahme gemacht, dass die Masse der Objekte konstant verteilt ist, es sich also um eine starre Vollkugel handelt. Dies ist insbesondere für die Erde mit ihrem flüssigen Erdinneren nicht wirklich gegeben, stellt aber eine gute Näherung dar.

Wenn nun die Eigendrehung der Erde aufgrund der Gezeitenreibung abnimmt, fließt dieser Drehimpuls praktisch ausnahmslos in den Bahndrehimpuls des Monds. Die Summe der Änderungen verschwindet und wir können annehmen, dass

$$\frac{\Delta L_{\text{Erde/Eigen}}}{\Delta t} = -\frac{\Delta L_{\text{Mond/Bahn}}}{\Delta t}. \quad (3)$$

---

<sup>1</sup> Das Verhältnis der Drehimpulse von Erd- zu Mondbahn lässt sich einfacher abschätzen: Nach Gl. 8 ist  $m_{\text{Erde}}R_{\text{Erde-D}} = m_{\text{Mond}}R_{\text{Mond-D}}$ , wobei  $R_{X-D}$  der Abstand des Mittelpunkts des Körpers  $X$  zum gemeinsamen Schwerpunkt  $D$  ist. Die Winkelgeschwindigkeit ist für die beiden Körper im Erde-Mond-System dieselbe, also erhalten wir die Gleichung

$$m_{\text{Erde}}R_{\text{Erde}}\omega = m_{\text{Mond}}R_{\text{Mond}}\omega.$$

Diese ist gleichbedeutend zu

$$R_{\text{Erde}}L_{\text{Erde/Bahn}} = R_{\text{Mond}}L_{\text{Mond/Bahn}}$$

oder auch

$$m_{\text{Mond}}L_{\text{Erde/Bahn}} = m_{\text{Erde}}L_{\text{Mond/Bahn}}.$$

Das Verhältnis der Bahndrehimpulse ist also gleich dem umgekehrten Verhältnis der Massen. Dieses ist immer gleich (es beträgt ungefähr 1 : 81,2), daher ist auch das Verhältnis der Bahndrehimpulse unabhängig vom Abstand zwischen Erde und Mond.

System	Drehimpuls	prozentualer Anteil
Mondbahn um Schwerpunkt	$2,83 \cdot 10^{34} \text{ kg} \cdot \text{m}^2/\text{s}$	79%
Erdbahn um Schwerpunkt	$3,4 \cdot 10^{32} \text{ kg} \cdot \text{m}^2/\text{s}$	1%
Eigendrehung Erde	$7,06 \cdot 10^{33} \text{ kg} \cdot \text{m}^2/\text{s}$	19,9%
Eigendrehung Mond	$2,4 \cdot 10^{29} \text{ kg} \cdot \text{m}^2/\text{s}$	< 0,001%

Tabelle 1: Größenordnung der Drehimpulse im Erde-Mond-System. Die Zahlen sind wiederum nur ungefähre Angaben und wurden aus den Parametern in Tab. 1 berechnet. Unsicherheiten liegen in der Annahme einer Kreisbahn für den Mond sowie in der Formel für die Eigendrehimpulse, die eine konstante Masseverteilung voraussetzt. Insbesondere ist der Eigendrehimpuls der Erde etwas kleiner (ungefähr  $5,85 \cdot 10^{33} \text{ kg} \cdot \text{m}^2/\text{s}$ ). Damit erhöht sich der Anteil des Bahndrehimpulses des Monds im Vergleich zum Gesamtdrehimpuls auf rund 82%, entsprechend verringert sich der Anteil des Eigendrehimpulses der Erde auf rund 17%.

Im Grenzfall  $\Delta t \rightarrow 0$  wird dieser Quotient zum Drehmoment.

Wir kennen das Drehmoment der Gezeitenreibung nicht. Aber wir kennen ziemlich genau den Abstand Erde-Mond und wissen, dass der Abstand des Monds im Laufe eines Jahres um rund 3,8 cm zunimmt. Bevor wir diesen Wert in Gl. 1 einsetzen müssen wir berücksichtigen, dass eine Änderung des Abstands (bzw. des Bahnradius) auch eine Änderung der Umlaufzeit zur Folge hat. Aus der Bedingung „Gravitationskraft = Fliehkraft“ folgt:

$$G \frac{m_{\text{Erde}} m_{\text{Mond}}}{R_{\text{EM}}^2} = m_{\text{Mond}} R_{\text{EM}} \left( \frac{2\pi}{T} \right)^2 \quad \Rightarrow \quad T = 2\pi \left( \frac{R_{\text{EM}}^3}{G m_{\text{Erde}}} \right)^{1/2}. \quad (4)$$

Dies ist das berühmte dritte Kepler'sche Gesetz: Die Quadrate der Umlaufzeiten verhalten sich wie die Kuben der Halbachsen. Setzen wir diese Beziehung in Gl. 1 ein, erhalten wir den Drehimpuls der Mondumlaufbahn als Funktion des Abstands (sowie weiterer Konstanten):

$$L_{\text{Mond/Bahn}} = m_{\text{Mond}} \sqrt{G m_{\text{Erde}} R_{\text{EM}}}. \quad (5)$$

Die Änderung des Drehimpulses als Funktion des Abstands ist somit: ([Herleitung](#))

$$\Delta L_{\text{Mond/Bahn}} = \frac{m_{\text{Mond}}}{2} \sqrt{\frac{G m_{\text{Erde}}}{R_{\text{EM}}}} \Delta R. \quad (6)$$

Für  $\Delta R = 3,8 \text{ cm}$  ist dies die Drehimpulsänderung in der Mondbahn in einem Jahr. Um denselben Wert ändert sich der Eigendrehimpuls der Erde in einem Jahr. Da sich in Gl. 2 nur  $\omega$  ändern kann, folgt:

$$\Delta L_{\text{Erde/Eigen}} = \frac{2}{5} m_{\text{Erde}} R_{\text{Erde}}^2 \Delta \omega = -\frac{2}{5} m_{\text{Erde}} R_{\text{Erde}}^2 \frac{2\pi}{T^2} \Delta T. \quad (7)$$

(Hier wurde  $\Delta \omega = \frac{d\omega}{dT} \Delta T$  ausgenutzt.) Das Minuszeichen deutet an, dass der Drehimpuls kleiner wird (also  $\Delta L$  negativ ist) wenn die Periode länger wird (also  $\Delta T$  positiv ist). Insgesamt erhalten wir somit:

$$\Delta T = \frac{5}{8\pi} T^2 \frac{m_{\text{Mond}}}{m_{\text{Erde}} R_{\text{Erde}}^2} \sqrt{\frac{G m_{\text{Erde}}}{R_{\text{EM}}}} \Delta R. \quad (8)$$

Mit den Gleichungen für den Bahndrehimpuls des Monds (Gl. 5) und dem Eigendrehimpuls der Erde (Gl. 2) können wir diese Beziehung auch umschreiben:

$$\frac{\Delta T}{T} = \frac{1}{2} \frac{L_{\text{Mond/Bahn}}}{L_{\text{Erde/Eigen}}} \frac{\Delta R}{R_{\text{EM}}}. \quad (9)$$

Diese Beziehung besitzt eine „rasche“ Herleitung: Wir wissen, dass

$$\Delta L_{\text{Erde/Eigen}} + \Delta L_{\text{Mond/Bahn}} = 0. \quad (10)$$

Die beiden Relationen

$$L_{\text{Erde/Eigen}} \propto \frac{1}{T} \quad \text{und} \quad L_{\text{Mond/Bahn}} \propto \sqrt{R_{\text{EM}}} \quad (11)$$

führen auf (man bilde jeweils die logarithmischen Ableitungen, sodass die Proportionalitätsfaktoren wegfallen)

$$\Delta L_{\text{Erde/Eigen}} = -\frac{L_{\text{Erde/Eigen}}}{T} \Delta T \quad \text{und} \quad \Delta L_{\text{Mond/Bahn}} = \frac{1}{2} \frac{L_{\text{Mond/Bahn}}}{R_{\text{EM}}} \Delta R_{\text{EM}} \quad (12)$$

Eingesetzt in Gl. 10 erhalten wir Gl. 9. Nach Tabelle 1 ist  $\frac{L_{\text{Mond/Bahn}}}{L_{\text{Erde/Eigen}}} \approx 4$ , außerdem ist  $\frac{\Delta R_{\text{EM}}}{R_{\text{EM}}} \approx 10^{-10}$  und  $T = 86\,400 \text{ s}$ . Damit folgt

$$\Delta T = 2 \cdot 10^{-10} \cdot 0,864 \cdot 10^5 \text{ s} \approx 1,7 \cdot 10^{-5} \text{ s}. \quad (13)$$

Dies ist die Änderung in der Tageslänge in einem Jahr. In 100 Jahren sind das etwas weniger als 0,002 s.

## 2.2 Wie alt ist der Mond?

Die Frage nach dem Alter des Monds ist für viele Bereiche von Bedeutung. Die Vermutung ist, dass der Mond von rund 4 Milliarden Jahren, also kurz nach der Entstehung der Erde, durch einen riesigen Meteoriteneinschlag in die Erde entstanden ist. Wenn wir ganz naiv die 3,8 cm pro Jahr, um die die Entfernung des Monds von der Erde zunimmt, extrapoliieren, kommen wir bei 4 Milliarden Jahren auf eine Entfernung von rund  $1,52 \cdot 10^8 \text{ m}$  oder 152 000 km. Diese Entfernung ist deutlich zu groß. Vermutlich hatte der Mond kurz nach seiner Entstehung eine Entfernung von der Erde, die knapp über der Grenze lag, bei welcher der Mond aufgrund der Gezeitenkräfte auseinandergerissen worden wäre (diese Grenze bezeichnet man auch als Roche-Grenze), das sind rund 15 000-20 000 km.

Es ist offensichtlich, weshalb die naive Extrapolation der 3,8 cm pro Jahr für die Zunahme der Erde-Mond-Entfernung nicht korrekt ist. Mehrere Gründe spielen hier eine wichtige Rolle. Insbesondere muss berücksichtigt werden, dass der Abstand zwischen Erde und Mond früher kleiner war.

1. Damit folgt, dass die Gezeitenkräfte, die sich wie  $1/R_{\text{EM}}^3$  verhalten, wesentlich größer waren. Bei einem großzügig angenommenen Faktor 10 zwischen dem ursprünglichen Abstand und dem heutigen Abstand macht dies für die Gezeitenkräfte einen Faktor 1000 aus. Dieser Einfluss auf Ebbe und Flut lässt sich kaum abschätzen.
2. Ein kleinerer Abstand zwischen Erde und Mond bedeutet auch, dass sich der Mond wesentlich schneller um die Erde gedreht hat. Nimmt man auch hier einen Faktor 10 an, folgt aus dem dritten Kepler'schen Gesetz, dass die Mondzyklen um mehr als einen Faktor 30 kürzer waren als heute, der Mond sich also in weniger als einem (heutigen) Tag um die Erde gedreht hat. Damit war auch die Zeitdauer zwischen Ebbe und Flut deutlich kürzer.
3. Schließlich hat sich die Erde früher wesentlich schneller gedreht als heute und damit folgten Ebbe und Flut rascher aufeinander. Der Reibungseffekt pro Jahr war wegen der häufigeren Fluten (unabhängig von ihrer Stärke) höher.

Berücksichtigt man all diese Effekte, so hat sich der Mond insbesondere zu Beginn deutlich schneller von der Erde entfernt, als es unsere naive Extrapolation angibt. Eine Rechnung, die all diese Effekte berücksichtigt, kommt auf ein Alter des Monds von rund 1,3 Milliarden Jahren (siehe [?]).

Dies wiederum ist eine zu kurze Zeitdauer. Dieses Ergebnis wird gelegentlich von Creationisten (Anhängern einer auf der Darstellung der Bibel basierenden Schöpfungsgeschichte, nach der Gott die Welt von etwas über 6000 Jahren erschaffen hat) angeführt um zu argumentieren, dass die sogenannten wissenschaftlichen Berechnungen für das Alter der Erde (und des Monds), die auf 4 Milliarden Jahre deuten, falsch sein müssen. Sie sehen hierin einen Widerspruch zu den beobachteten Tatsachen.

Wäre das Ergebnis von 1,3 Milliarden Jahren für das Alter des Monds korrekt, wäre das tatsächlich ein Problem für die übliche wissenschaftliche Theorie über das Alter der Erde. Der wesentliche Unsicherheitsfaktor für die obige Berechnung ist jedoch der Einfluss der um einen Faktor 10–1000 (als Größenordnung) stärkeren Gezeitenkräfte auf Ebbe und Flut und die damit verbundene Gezeitenreibung. Insbesondere wird vermutet, dass die Gezeitenreibung in früheren Jahrtausenden einen deutlich geringeren Einfluss hatte, als eine einfache Extrapolation von heutigen Verhältnissen vermuten ließe: Es gab vor rund 200 Millionen Jahren nur einen großen Kontinent (Pangaea), und vermutlich gab es in der Geschichte der Erde vor ein bzw. zwei Milliarden Jahren mehrere Phasen, in denen es nur einen Kontinent gab. Dadurch war die Gezeitenreibung deutlich geringer (es gab weniger Küsten) und der Effekt des Drehimpulsaustauschs zwischen Erde und Mond war wesentlich kleiner. Hier gibt es noch viele Unsicherheiten, das Argument der Creationisten ist jedoch sicherlich zu einfach und nicht haltbar.

### 2.3 Wie geht es weiter?

Die Erde wird sich in Zukunft immer langsamer drehen und der Mond dabei immer mehr entfernen. Die Monate werden dadurch ebenfalls länger. Ist der derzeitige Eigendrehimpuls (siehe Tab. 1) der Erde „aufgebraucht“ und ganz auf den Mond übergegangen, wird der Mond einen Bahndrehimpuls von rund  $3,54 \cdot 10^{34} \text{ km}\cdot\text{m}/\text{s}^2$  haben. Nach Gl. 5 bedeutet dies einen Abstand Erde-Mond von etwas über  $R_{EM} = 425\,000 \text{ km}$  haben und der Monat wird rund 33 (heutige) Sonnentage dauern. Die Rotation der Erde ist dann an den Mond gebunden, so wie jetzt schon die Rotation des Monds gebunden ist, d.h. der Mond zeigt der Erde immer dieselbe Seite. Daher dauert ein Monat dann nur einen Sonnentag.

### 3 Herleitung einiger Gleichungen

#### 3.1 Differenzieller Drehimpuls

Der direkte Weg, aus

$$L_{\text{Mond/Bahn}}(R_{\text{EM}}) = m_{\text{Mond}} \sqrt{Gm_{\text{Erde}} R_{\text{EM}}} \quad (14)$$

das Differenzial zu  $L$  abzuleiten ist die Taylor-Entwicklung bzw. die erste Ableitung:

$$\frac{dL_{\text{Mond/Bahn}}(R_{\text{EM}})}{dR_{\text{EM}}} = \frac{1}{2} m_{\text{Mond}} \sqrt{\frac{Gm_{\text{Erde}}}{R_{\text{EM}}}} \quad (15)$$

und somit folgt:

$$\Delta L_{\text{Mond/Bahn}}(R_{\text{EM}}) = \frac{1}{2} m_{\text{Mond}} \sqrt{\frac{Gm_{\text{Erde}}}{R_{\text{EM}}}} \Delta R. \quad (16)$$

Man kann die Ableitung auch umgehen:

$$\Delta L_{\text{Mond/Bahn}} \equiv L_{\text{Mond/Bahn}}(R_{\text{EM}} + \Delta R) - L_{\text{Mond/Bahn}}(R_{\text{EM}}) \quad (17)$$

$$= m_{\text{Mond}} \sqrt{Gm_{\text{Erde}}(R_{\text{EM}} + \Delta R)} - m_{\text{Mond}} \sqrt{Gm_{\text{Erde}} R_{\text{EM}}}. \quad (18)$$

Wir klammern  $R_{\text{EM}}$  aus:

$$\Delta L_{\text{Mond/Bahn}} = m_{\text{Mond}} \sqrt{Gm_{\text{Erde}} R_{\text{EM}} \left(1 + \frac{\Delta R}{R_{\text{EM}}}\right)} - m_{\text{Mond}} \sqrt{Gm_{\text{Erde}} R_{\text{EM}}} \quad (19)$$

$$= m_{\text{Mond}} \sqrt{Gm_{\text{Erde}} R_{\text{EM}}} \left(\sqrt{1 + \frac{\Delta R}{R_{\text{EM}}}} - 1\right) \quad (20)$$

Wir nutzen nun die Entwicklung der Wurzelfunktion:

$$\sqrt{1 + x} \approx 1 + \frac{1}{2}x + O(x^2) \quad (21)$$

und erhalten:

$$\Delta L_{\text{Mond/Bahn}} = \frac{1}{2} m_{\text{Mond}} \sqrt{Gm_{\text{Erde}} R_{\text{EM}}} \frac{\Delta R}{R_{\text{EM}}} = \frac{1}{2} m_{\text{Mond}} \sqrt{\frac{Gm_{\text{Erde}}}{R_{\text{EM}}}} \Delta R. \quad (22)$$

# Der Nachthimmel

Thomas Filk, Universität Freiburg

Die funkelnden Sterne am Nachthimmel haben sicherlich schon unsere prähistorischen Vorfahren fasziniert und vermutlich haben auch sie schon Bilder, Gestalten oder seltsame Wesen in den Verteilungen der Sterne gesehen. Jedenfalls lassen sich viele Namen in mesopotamische, babylonische, persische, altägyptische und griechische Zeiten zurückverfolgen. Es gibt sogar Vermutungen, dass Zeichnungen in den südfranzösischen Höhlen von Lascaux vor mindestens 15 000 Jahren schon Darstellungen von Sternbildern enthalten. Ptolemäus erwähnt 48 Sternbilder, die von der nördlichen Halbkugel aus gut sichtbar sind. Im Laufe der Zeit kamen unzählige Sternbilder hinzu, insbesondere im 17. und 18. Jahrhundert die Sternbilder in der Nähe des Südpols. Viele Sternbilder sind aber auch wieder in Vergessenheit geraten. Nachdem das Durcheinander bei den Sternbildern im 19. Jahrhundert zu groß wurde, entschloss sich die Internationale Astronomische Vereinigung (IAU) im Jahre 1922, insgesamt 88 Sternbilder als verbindlich zu definieren. Die genauen Grenzen und Namen dieser Sternbilder wurden in den Folgejahren festgelegt.

Streng genommen sollte man zwischen Sternbildern und Asterismen unterscheiden: Ein Sternbild bezeichnet eine Fläche am Nachthimmel, die durch ihre Grenzen (Abfolgen von sphärischen Kreisbögen zu konstanter Rektazension und Deklination) definiert sind. Ein Asterismus bezeichnet eine Konfiguration einzelner Sterne, die ein eingängiges Muster zeigen. Beispielsweise ist der Große Wagen, bestehend aus sieben Sternen, ein Asterismus, der zum Sternbild Ursa Major (Große Bärin) gehört, das wesentlich mehr Sterne und auch eine größere Fläche umfasst. Tierkreiszeichen sind spezielle Sternbilder, die in der Ebene der Ekliptik liegen.

Bevor wir auf die Sternbilder, Asterismen und Tierkreiszeichen genauer eingehen, müssen wir Himmelskoordinaten einführen, mit deren Hilfe wir den Ort von Sternen, Planeten oder anderen Himmelsobjekten bezeichnen können.

## 1 Himmelskoordinaten

Es gibt viele Himmelskoordinaten, die in unterschiedlicher Form in Gebrauch sind. In diesem Kapitel beschränken wir uns jedoch auf vier sphärische Koordinatensysteme: lokale Himmelskoordinaten, Äquatorialkoordinaten, Ekiptikkoordinaten und galaktische Himmelskoordinaten, wobei im späteren Verlauf nur die Äquatorialkoordinaten von Bedeutung sind.

### 1.1 Lokale Koordinaten

Lokale Koordinaten haben als Ursprung den Standort des Beobachters, die Koordinatenachsen sind in der Tangentialebene dieses Standorts nach der Nord-Süd bzw. Ost-West-Richtung ausgerichtet. Die dritte Achse steht senkrecht auf der Tangentialebene, zeigt also vom Beobachtenden aus senkrecht nach oben zum sogenannten Zenit.

Eine strenge Definition dieses Koordinatensystems stößt auf kleine Schwierigkeiten, z.B. bei der Definition der Tangentialebene bzw. der Senkrechten dazu. Wählt man für die Form des Erdkörpers das Geoid (eine Äquipotentialfläche zur Schwerkraft auf der Höhe des Meeresspiegels), ein dieses Geoid annäherndes Ellipsoid oder eine Kugel? Das Geoid hat den Vorteil, dass man durch eine sogenannte Lotlinie - ein Gewicht an einem Faden, das senkrecht herunterhängt - die Senkrechten zur Tangentialebene genau bestimmen kann. Es handelt sich beim Geoid allerdings im Detail um

eine komplizierte Fläche, und insbesondere müssen bei der praktische Bestimmung noch Einflüsse von Mond und Sonne (Gezeitenkräfte) herausgerechnet werden. Die Kugelform ist geometrisch am einfachsten. Die Senkrechte ist definiert durch eine gedachte Linie durch den Erdmittelpunkt und den Beobachterstandort.

Dieses Koordinatensystem hat den Nachteil, dass Ortsbezeichnungen am Himmel sowohl vom lokalen Beobachterstandort, der Uhrzeit und der Jahreszeit abhängen und sich somit nur für sehr grobe Ortsbezeichnungen am Himmel eignet (im Sinne von „Im Juli sieht man das Sternbild Schütze am späten Abend in Südrichtung“). Aus diesem Grund gehen wir auch auf die oben genannten Probleme mit der genauen Definition nicht weiter ein.

## 1.2 Äquatoriale Koordinaten

Das für praktische astronomische Zwecke meist verwendete Koordinatensystem bezieht sich auf den Himmelsäquator bzw. den Himmelsnordpol. Der Himmelsnordpol ist definiert durch die gedachte Verlängerung der Rotationsachse der Erde, die einen Punkt am Himmel auszeichnet. Der zugehörige Himmelsäquator steht senkrecht auf der Rotationsachse und kann als eine Projektion des Erdäquators vom Erdmittelpunkt aus auf den Himmel gedacht werden. Winzige Schwankungen der Erdachse (sogenannte Nutationsbewegungen, die am geographischen Nordpol zu Fluktuationen im Bereich von mehreren Metern führen können und im Bereich von Tagen bis wenigen Jahren liegen) werden dabei unberücksichtigt gelassen oder herausgerechnet.

Allerdings führt die Präzession der Erde (die Drehung der Rotationsachse um eine Achse senkrecht zur Ekliptik mit einer Umlaufperiode von knapp 26 000 Jahren) zu einer langsamen, stetigen Verschiebung dieses Koordinatensystems. Für genauere Ortsangaben muss also angegeben werden, auf welchen Zeitpunkt man das Koordinatensystem bezieht. Heute oft gebräuchlich ist der 1. Januar des Jahres 2000, 12 Uhr mittags, was dann oft mit J2000.0 (oder kurz J2000) bezeichnet wird: J steht für Julianisches Datum, 2000 für das Jahr und 0 für den 1. Januar 12 Uhr mittags, was in der Julianischen Datumsangabe dem Tageswechsel entspricht. Man beachte, dass J2000 sowohl für einen Zeitpunkt steht, als auch für ein ausgezeichnetes äquatoriales Koordinatensystem. Dieses Koordinatensystem bezeichnet man auch schon mal mit ICRF (International Celestial Reference Frame). Hierbei handelt es sich um ein Koordinatensystem, das durch nahezu 300 außergalaktische Objekte (meiste Quasare oder ähnliche Radioquellen) festgelegt wurde und das mit dem Koordinatensystem J2000 übereinstimmt.

Früher wählte man auch die sogenannten Bessel'schen Epochen.<sup>1</sup> Sie gehen von einem exakten tropischen Jahr aus, definiert durch den Stand der Sonne relativ zum Frühlingspunkt. Die Grenzen der Sternbilder beziehen sich beispielsweise auf die Epoche B1875.0, also das Bessel'sche Jahr 1875 (1. Januar). Wegen der Schaltjahre fallen diese Zeitpunkte aber schon mal auf unterschiedliche Tage, sodass man zu der Julianischen Zählweise überging.

Die Festlegung des Himmelsnordpols und des Himmelsäquators legt das Koordinatensystem am Himmel noch nicht fest. Es muss noch ein ausgezeichneter Punkt auf dem Himmelsäquator als Bezugspunkt für eine Winkelkoordinate entlang des Äquators gewählt werden. Dieser Bezugspunkt sollte unabhängig von der momentanen (tageszeitabhängigen) Lage der Erde sein (also z.B. nicht die Projektion des Nullmeridians auf die Himmelskugel).

Als Bezugspunkt auf dem Himmelsäquator dient der sogenannten *Frühlingspunkt*. Der Frühlingspunkt ist ein Punkt auf der Himmelssphäre, der sich aus dem Schnittpunkt von zwei Großkreisen am Himmel bestimmt (siehe Abb. 1, links): Der eine Großkreis ist der Himmelsäquator - die

---

<sup>1</sup>Umgangssprachlich oder auch geologisch versteht man unter einer Epoche meist einen längeren Zeitraum. In der Astrophysik versteht man unter einer Epoche jedoch den Zeitpunkt eines bestimmten Ereignisses, im vorliegenden Fall den genauen Zeitpunkt (1. Januar 2000), bei dem die Erde eine bestimmte Lage relativ zur Sonne hatte.

Projektion des Erdäquators vom Erdmittelpunkt aus betrachtet auf die Himmelskugel. Der zweite Großkreis ist die sogenannte Himmelsekliptik - die Projektion der Erdumlaufbahn um die Sonne auf die Himmelskugel vom Sonnenmittelpunkt aus betrachtet. Umgekehrt kann man die Himmelsekliptik auch definieren als die Projektion der Sonne - vom Erdmittelpunkt aus betrachtet (für praktische Zwecke reicht auch die Projektion der Sonne an den Himmel vom augenblicklichen Beobachtungsstandort) - an die Himmelssphäre. Auch wenn sich die Sonne im Laufe eines Tages für einen Beobachter einmal um die Erde zu bewegen scheint, bewegt sich ihre Projektion an den Sternenhimmel (der allerdings nicht gleichzeitig mit der Sonne zu sehen ist) nur um rund einen Grad pro Tag, weil sich der Sternenhimmel aufgrund der Eigendrehung der Erde im Wesentlichen mit der Sonne bewegt. Diese scheinbare Bewegung der Sonne überstreicht im Laufe eines Jahres einen Großkreis: die Himmelsekliptik. Wegen der Neigung der Erdachse relativ zur Ekliptik (derzeit rund 23,4 Grad) sind die beiden genannten Großkreise verschieden. Sie schneiden sich in zwei Punkten: dem Frühlingspunkt und dem Herbilstpunkt. Die Zeitpunkte, in denen sich die Sonne von der Erde aus betrachtet in diesen Punkten befindet, bezeichnet man auch als Äquinoktien oder Tag-und-Nacht-Gleichen.

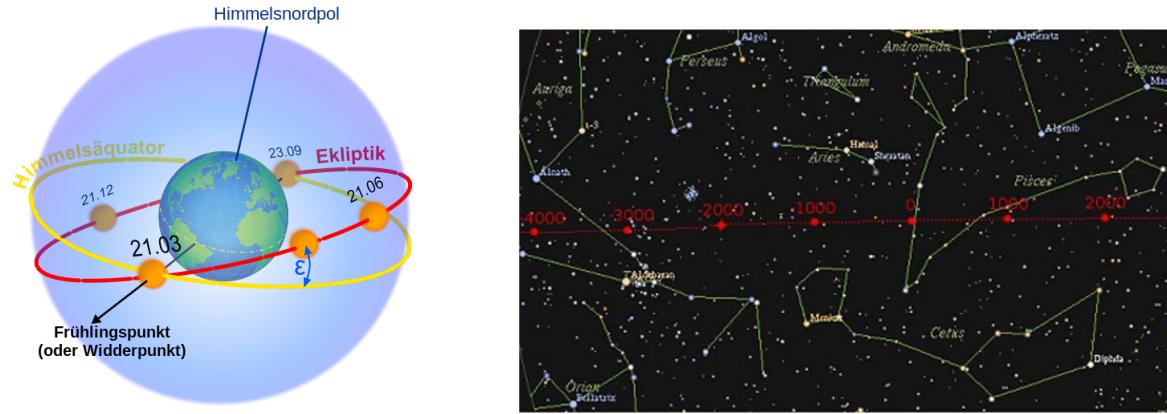


Abbildung 1: (links) Der Frühlingspunkt als der Schnittpunkt zwischen dem Himmelsäquator und der Ekliptik. (rechts) Der Frühlingspunkt in den letzten Jahrtausenden. Während der Frühlingspunkt heute (im Jahr 2023) im Sternbild Fische (Pisces) steht, stand er vor rund 2500 Jahren im Sternbild Widder (Aries). Und in einigen Jahrhunderten wird er im Sternbild Wassermann stehen. (links [4], rechts [2])

Die beiden Großkreise - Himmelsekliptik und Himmelsäquator - schneiden sich in zwei Punkten. Der Frühlingspunkt ist definiert als der Schnittpunkt, bei dem sich die Sonne von Süden nach Norden durch die Äquatorialebene bewegt, beim Herbilstpunkt bewegt sie sich von Nord nach Süd durch die Äquoralebene. Derzeit liegt der Frühlingspunkt im Sternbild Fische. In einigen hundert Jahren wird er in das Sternbild Wassermann eintreten, und in babylonischer Zeit, als die Tierkreiszeichen benannt wurden, befand sich der Frühlingspunkt im Sternbild Widder. Aus diesem Grunde heißt er auch heute noch gelegentlich Widderpunkt (siehe Abb 1, rechts).

Durch die Angabe der genannten Bezugsgrößen - Himmelsnordpol, Himmelsäquator und Frühlingspunkt auf dem Himmelsäquator - können wir ein Koordinatensystem für den Himmel, also ein sphärisches Koordinatensystem konstruieren. Dazu definiert man Längengrade - halbe Großkreisbögen vom Himmelsnordpol zum Himmelssüdpol (auch als Stundenlinien bezeichnet) - und Breitengrade - Kreisbögen, deren Punkte unter einem konstanten Winkel relativ zum Himmelsnordpol stehen. Die Breitengrade werden durch die sogenannte Deklination gekennzeichnet: Dies ist der Winkel  $\delta$  von der Äquatorialebene zu einem Himmelsobjekt entlang eines Längengrads. Die Deklination wird als Winkel zwischen  $+90^\circ$  und  $-90^\circ$  (Nord- bzw. Südrichtung) angegeben. Genaue Bezeichnungen

verwenden meist Bogenminuten, Bogensekunden und dann Nachkommastellen. Der Längengrad wird ausgehend vom Frühlingspunkt meist in Stundenwinkeln angegeben: der volle Himmelsäquator wird entgegen dem Uhrzeigersinn in 24 Stundenwinkel eingeteilt, entsprechend sind genauere Bezeichnungen in Minuten und Sekunden. Diesen Stundenwinkel bezeichnet man als Rektazension. Beispielsweise hat Arkturus, einer der hellsten Sterne des Nordhimmels im Sternbild Bootes (Bärenhüter), die Rektazension  $14^{\text{h}} 15^{\text{m}} 39,67207^{\text{s}}$  und die Deklination  $+19^{\circ} 10' 56,6730''$ , bezogen auf die Epoche J2000.

### 1.3 Ekliptische Koordinaten

Statt des Himmelsäquators und dem dazu senkrecht stehenden Himmelsnordpol kann man auch den Großkreis zur Ekliptik der Sonne und einen dazu senkrecht stehenden Punkt am Himmel als Koordinatensystem wählen. Statt der Rektazension und der Deklination verwendet man hier den ekliptischen Längengrad und den ekliptischen Breitengrad. Beide werden in Winkelmaßen angegeben, wobei der ekliptische Längengrad eine Winkeleinteilung der Ekliptik in  $360^{\circ}$  angibt und der ekliptische Breitengrad im Bereich zwischen  $+90^{\circ}$  (ekliptischer Nordpol) bis  $-90^{\circ}$  (ekliptischer Südpol) liegt. Auf der Himmelsekliptik wird ebenfalls der Frühlingspunkt als Nullpunkt für die Längengrade ausgezeichnet.

Dieses Koordinatensystem wird seltener verwendet (wegen der Bewegung vieler Planeten in der Nähe der Ekliptik findet es zur Beschreibung solcher Himmelsobjekte schon mal Verwendung), sodass wir hier nicht weiter darauf eingehen.

### 1.4 Galaktische Koordinaten

Schließlich verwendet man in der Astronomie auch gelegentlich das galaktische Koordinatensystem. Als ausgezeichneter Referenzkreis dient ein Großkreis am Himmel, der mehr oder weniger durch die Ebene der Milchstraße, also die Sterne in der Ebene unserer Galaxie, ausgezeichnet ist. Der Ursprung des dreidimensionalen Koordinatensystems ist die Sonne, das Zentrum der Galaxie (im Sternbild Schützen) dient als Referenzpunkt auf der galaktischen Ebene. Der galaktische Nordpol steht senkrecht auf der galaktischen Ebene und liegt im mittleren oberen Teil des Sternbilds Haar der Berenike (Coma Berenices), nördlich des Sternbilds Jungfrau und rund  $30^{\circ}$  südlich (also vom Nordpol weggerichtet) des Großen Wagens.

Von der Erde aus betrachtet handelt es sich zwar um ein sphärisches Koordinatensystem, es dient aber auch bei bekannten Entfernung zu Himmelsobjekten als dreidimensionales Koordinatensystem zur Lagebezeichnung von Objekten innerhalb unserer Galaxie.

## 2 Sternbilder

Wie schon erwähnt wurden von der IAU (International Astronomical Union) 88 Sternbilder mit ihren Grenzen als verbindlich festgelegt (siehe Abb. 2). Die Grenzen bestehen aus Kreisbögen zu konstanter Deklination bzw. konstanter Rektazension - bilden somit in einer äquatorialen Zylinderprojektion senkrechte und waagerechte Linien - und umranden jeweils eine Fläche.<sup>2</sup> In der Summe überdecken diese Flächen die gesamte Himmelskugel.

In manchen Sternbildern sind mit bloßem Auge kaum markante Sterne zu sehen. Dass man diesen Flächen trotzdem ein eigenes Sternbild zugeordnet hat liegt daran, dass man mit der Erfindung des Fernrohrs unzählige Objekte entdeckte, die mit bloßem Auge nicht erkennbar sind. Um den Ort

<sup>2</sup>Das Sternbild Schlange (Serpens) besteht als einziges Sternbild aus zwei Flächen, die man als Schlangenkopf (Serpens caput) und Schlangenschwanz (Serpens cauda) bezeichnet. Die beiden Teile befinden sich rechts und links vom Sternbild Schlangenträger (Ophiuchus) und sind in den Sommermonaten am Abend in südwestlicher Richtung gut zu sehen.

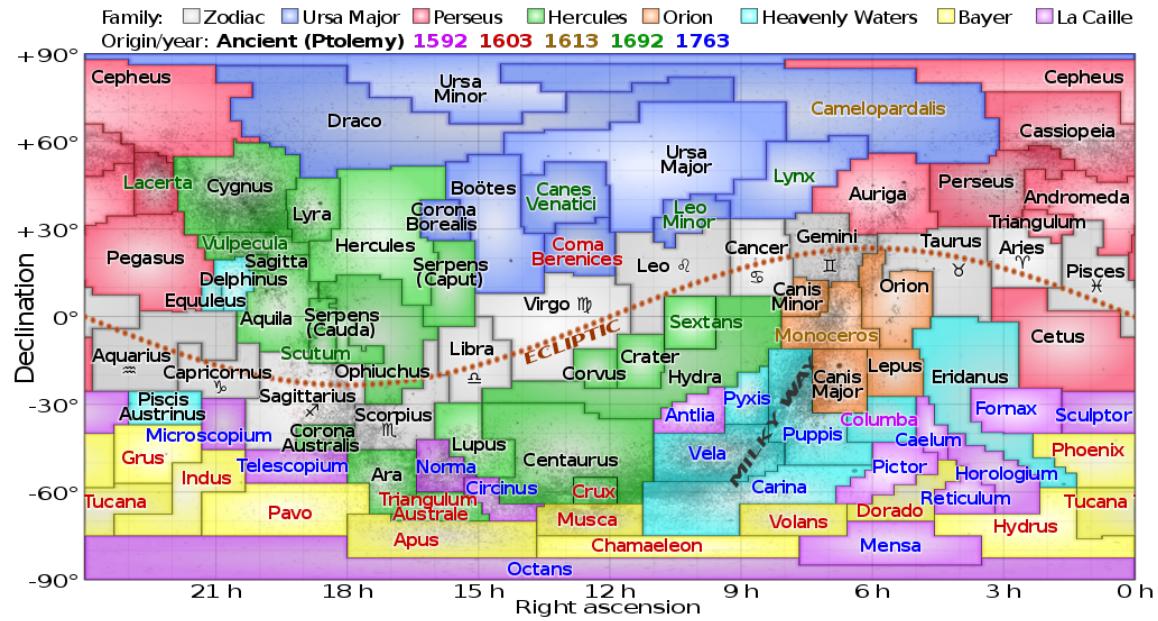


Abbildung 2: Die 88 Sternbilder in der Äquatorialdarstellung. Die Hintergrundfarben beziehen sich auf verschiedene Gruppen von Sternbildern, insbesondere handelt es sich bei den grau hinterlegten Sternbildern um die klassischen Tierkreiszeichen. In schwarzer Schrift sind die klassischen Sternbilder des Ptolemäus bezeichnet, andere Farben kennzeichnen Sternbilder, die erst in späterer Zeit hinzugekommen sind. (aus [3])

dieser Objekte am Himmel grob angeben zu können, hat man den gesamten Himmel mit Sternzeichen überdeckt. Auf der Webseite des *Strasbourg astronomical Data Center* [1] findet man die exakten Grenzen (bezogen auf das Besseljahr 1875) aller Sternbilder. Sehr wenige Sternbilder (z.B. der Sextant - Sextans - südlich vom Sternbild Löwen) bilden ein einfaches Quadrat. Eines der komplexesten Sternbilder ist das Sternbild Drachen (Draco) mit rund 50 Grenzlinien, das sich zwischen dem kleinen und großen Bären halb um den Polarstern windet.

Jedes Sternbild besitzt einen lateinischen Namen und eine Abkürzung mit drei Buchstaben: z.B. UMa für Ursa Major - Große Bärin (im Deutschen meist als Großer Bär bezeichnet) oder Sgr für Sagittarius - Schütze. Diese Nomenklatur geht auf Henry Norris Russell zurück (nach dem auch das Hertzsprung-Russell-Diagramm benannt ist), der diese Bezeichnungen 1922 vorschlug. Innerhalb eines Sternbilds werden die hellsten Sterne meist durch einen griechischen Buchstaben sowie den Genitiv des lateinischen Sternbildnamens bezeichnet, wobei sich in der Regel die Reihenfolge  $\alpha$ ,  $\beta$ , etc. nach der Helligkeit der Sterne richtet. Allerdings gilt diese Regel nicht streng: Im Sternbild Ursa Major (Großer Bär) scheint sich die Reihenfolge eher an der Reihenfolge in dem Asterismus Großer Wagen zu orientieren als an der Helligkeit. So ist  $\alpha$  Ursae Majoris der zweithellste Stern in diesem Sternbild, wohingegen  $\epsilon$  Ursae Majoris der hellste ist. Wenn das griechische Alphabet durchgelaufen ist, verwendet man oft lateinische Buchstaben und anschließend Zahlen.

Bei Doppel- oder Dreifachsternsystemen bezeichnet man die Sterne dieses Systems nach dem Namen mit einem angehängten Großbuchstaben. So ist  $\alpha$  Centauri ein Dreifachsternsystem, wobei  $\alpha$  Centauri A und  $\alpha$  Centauri B vergleichsweise helle Sterne sind, wohingegen es sich bei  $\alpha$  Centauri C (auch Proxima Centauri genannt, da es sich mit 4,26 Lichtjahren um den uns nächsten Stern nach der Sonne handelt) um einen roten Zwerg handelt, der nur in einem guten Fernrohr sichtbar ist. Oft verwendet man auch die Abkürzungen, beispielsweise  $\alpha$  Cen A.

Die hellsten Sterne haben eigene Namen, die meist aus der Antike stammen. Auf [5] findet man eine Liste der 100 hellsten Sterne mit ihren Namen, Helligkeiten, Magnituden, offiziellen Bezeichnungen sowie ihrer Rektazension und Deklination. Etwas umfangreichere Listen findet man auf [6].

### 3 Tierkreiszeichen

Tierkreiszeichen sind Sternzeichen, die auf der Ekliptik liegen, d.h. vor denen von der Erde aus betrachtet die Sonne im Laufe eines Jahres einmal steht (siehe Abb. 2). Streng genommen sollte man zwischen den sogenannten siderischen Tierkreiszeichen (die man besser als Sternbilder der Ekliptik bezeichnet) und den tropischen Tierkreiszeichen unterscheiden.

Die siderischen Tierkreiszeichen oder auch Ekliptiksternbilder sind die Sternzeichen, wie sie von der IAU definiert wurden und die auf der Ekliptik liegen. Sie unterteilen den Vollkreis der Ekliptik in nicht immer exakt gleich lange Anteile von ungefähr  $30^\circ$ . Allerdings gibt es hier schon eine Besonderheit: Das Sternbild des Schlangenträgers (Ophiuchus) überdeckt zwischen dem Sternbild Skorpion (Scorpio) und dem Sternbild Schützen (Sagittarius) mit seinem unteren Rand ebenfalls einen Teil der Ekliptik. Die beiden Anteile der Ekliptik zum Skorpion und zum Schlangenträger ergeben zusammen einen Abschnitt von ungefähr  $30^\circ$ , wobei der Anteil des Schlangenträgers deutlich größer als der des Skorpions ist. Der Schlangenträger zählt aber in der Tradition nicht zu den Tierkreiszeichen. In manchen Listen wird der Schlangenträger heute als 13. Tierkreiszeichen hinzugezählt, insbesondere wenn man von den siderischen Tierkreiszeichen bzw. Ekliptiksternbildern spricht.

Dem gegenüber gibt es die tropischen Tierkreiszeichen. Sie unterteilen die Ekliptik in zwölf exakt gleich große Anteile von  $30^\circ$  und tragen dieselben Bezeichnungen wie die siderischen Tierkreissternbilder. Allerdings hat sich aufgrund der Präzession der Erde der Frühjahrspunkt im Verlauf der Zeit verschoben. Vor etwas über 2500 Jahren, als die Bezeichnungen der Tierkreiszeichen in Babylonien entstanden sind, befand sich der Frühjahrspunkt im Sternbild Widder, weshalb man ihn auch heute noch manchmal als Widderpunkt bezeichnet. Auf diese Einteilung der Ekliptik in zwölf gleiche Teile bezieht sich auch heute noch die Astrologie: Die astrologischen Tierkreiszeichen beginnen am 21. März mit dem Zeichen Widder und schreiten in nahezu gleichen zeitlichen Abständen um jeweils  $30^\circ$  voran.<sup>3</sup> Heute liegt der Frühjahrspunkt im Sternbild Fische, und in einigen Jahrhunderten wird er ins Sternbild Wassermann gewandert sein. Das bedeutet, derzeit sind die tropischen und die siderischen Tierkreiszeichen gegeneinander um ein Zeichen verschoben: Am 21. März, wenn das tropische Tierkreiszeichen Widder beginnt, befindet sich die Sonne im Sternzeichen Fische.

Unter dem Zodiak, früher oft synonym zu Tierkreiszeichen verwendet, versteht man heute ein Band von rund  $\pm 10$  Grad um die Ekliptik, in der sich die meisten Planetenbahnen sowie die Bahn des Mondes bewegen. Auch der Zodiak ist in zwölf exakt  $30^\circ$  umfassende Bereiche unterteilt und richtet sich insofern nach den tropischen Tierkreiszeichen.

Gewöhnlich beginnt man die Reihe der Tierkreiszeichen mit dem Widder (Aries), in dem vor rund 2500 Jahren der Frühlingspunkt lag. In aufsteigender Rektazension, also entgegen dem Uhrzeigersinn, folgen die Tierkreiszeichen Stier (Taurus), Zwillinge (Gemini), Krebs (Cancer), Löwe (Leo), Jungfrau (Virgo), Waage (Libra), Skorpion (Scorpius), Schütze (Sagittarius), Steinbock (Capricornus), Wassermann (Aquarius) und Fische (Pisces), wobei man im Zusammenhang mit den Ekliptiksternbildern zwischen den Skorpion und den Schützen noch den Schlangenträger als 13. Tierkreissternbild einfügt.

---

<sup>3</sup>Die Wendepunkte und Äquinoktien definieren jeweils gleiche Abschnitte von  $90^\circ$ . Wegen der elliptischen Bahn und der damit verbundenen ungleichen Geschwindigkeit der Erde um die Sonne sind die zugehörigen räumlichen und zeitlichen Abschnitte jedoch nicht exakt gleich lang.

## 4 Asterismen

Asterismen sind einzelne Gruppen von Sternen, die ein markantes Muster zeigen. Bekannte Beispiele sind der große und kleine Wagen in den Sternbildern Ursa Major und Ursa Minor mit jeweils sieben Sternen, wobei man bei nicht optimalen Sichtverhältnissen vom kleinen Wagen oft nur drei Sterne sieht, oder auch das Himmels-W der Cassiopeia aus fünf Sternen. Ein weiteres Beispiel (bekannt unter anderem aus dem Film „Men in Black“) ist der Gürtel des Orion, bestehend aus drei hellen Sternen in einer Reihe.

Ein Asterismus kann auch aus Sternen zu verschiedenen Sternbildern bestehen. Beispiele hier sind das Frühlingsdreieck, bestehend aus Spica (Sternbild Jungfrau - Virgo), Arktur (Sternbild Bärenhüter - Bootes) und Regulus (Sternbild Löwe - Leo), oder auch das Sommerdreieck, bestehend aus Wega (Sternbild Leier - Lyra), Altair (Sternbild Adler - Aquila) und Deneb (Sternbild Schwan - Cygnus). Hierbei handelt es sich jeweils um eine Gruppe von drei besonders hellen Sternen, die man im Frühjahr bzw. im Sommer am Abendhimmel sehen kann. Ein weiteres Beispiel ist das Wintersechseck, bestehend aus den Sternen Capella (Sternbild Fuhrmann - Auriga), Aldebaran (Sternbild Stier - Taurus), Rigel (Sternbild Orion), Sirius (Sternbild Großer Hund - Canis Major), Prokyon (Sternbild Kleiner Hund - Canis Minor) und Pollux (Sternbild Zwillinge - Gemini). Diese Sterngruppen erstrecken sich teilweise über einen großen Teil des Nachthimmels.

Es gibt aber auch sehr kleine Asterismen, die man mit bloßem Auge kaum auflösen kann, wohl aber mit einem einfachen Fernglas: Ein Beispiel ist der „Kleiderbügel“ im Sternbild Fuchs (Vulpecula, zwischen den Sternbildern Adler und Schwan gelegen). Der Asterismus besteht aus 10 Sternen, von denen sechs nahezu auf einer Geraden liegen, plus vier Sterne, die in der Mitte über dieser Geraden einen Dreiviertelkreis bilden. Insgesamt erscheint diese Konfiguration, die offiziell den Namen Collinder 399 trägt, wie ein Kleiderbügel.

## Literatur

- [1] Strasbourg astronomical Data Center, Grenzen der Sternbilder:  
<https://vizier.cds.unistra.fr/vizier/VizieR/constellations.htm>.
- [2] Rocket Site „Wann beginnt das Zeitalter des Wassermanns?“ <https://damthoitrang.org/de/wann-beginnt-das-zeitalter-des-wassermanns/>
- [3] aus Wikipedia „Zodiac“, Quelle: <http://svs.gsfc.nasa.gov/vis/a000000/a003500/003572>,  
Autor: Cmglee, Timwi, NASA.
- [4] Wikipedia „Äquinoktium“, <https://commons.wikimedia.org/wiki/File:Ecliptic.svg>
- [5] Wikipedia „Liste der hellsten Sterne“: [https://de.wikipedia.org/wiki/Liste\\_der\\_hellsten\\_Sterne](https://de.wikipedia.org/wiki/Liste_der_hellsten_Sterne)
- [6] Listen heller Sterne: <http://stars.astro.illinois.edu/sow/bright.html> (Webseite von Jim Kahler, die 172 hellsten Sterne) und <http://www.atlasoftheuniverse.com/stars.html> (Atlas of the Universe; die 300 hellsten Sterne).

# Die Kosmische Entfernungsleiter

Thomas Filk, Universität Freiburg

Die Bestimmung astronomischer bzw. kosmischer Entfernung ist sowohl ein spannendes Thema der Geschichte der Physik als auch ein interessantes Forschungsfeld der modernen Physik. Während der Abstand von der Erde zum Mond schon im Altertum bekannt war (siehe Abschnitte 1 und 2) kennt man den Abstand von der Erde zur Sonne - die sogenannte Astronomische Einheit AU (astronomical unit) - mit einer gewissen Verlässlichkeit erst seit dem 18. Jahrhundert.

Tabelle 1 enthält einige Daten, die in diesem Abschnitt gelegentlich verwendet werden.

Bezeichnung	Symbol	Wert
Durchmesser Erde	$D_E$	$D_E = 12\,740 \text{ km}$
Durchmesser Mond	$D_M$	$D_M = 3\,474 \text{ km}$
Durchmesser Sonne	$D_S$	$D_S = 1\,400\,000 \text{ km}$
Abstand Erde-Mond	$r_M$	$r_M = 380\,000 \text{ km}$
Abstand Erde-Sonne	$r_S$	$r_S = 150\,000\,000 \text{ km}$
Winkel Sonne-Mond bei Halbmond	$\alpha$	$\alpha = 89^\circ 51'$

Tabelle 1: Einige Daten des Sonne-Erde-Mond-Systems.

## 1 Aristarchos von Samos

Aristarchos von Samos lebte um 310 v.Chr. bis 230 v.Chr, also kurz nach Aristoteles (384-322 v.Chr.) und ungefähr zeitgleich mit Archimedes (284-212 v.Chr.). Er stellte sich die Sonne als ein riesiges Himmelsfeuer vor, das unter anderem den Mond anstrahlt. Er deutete eine Mondfinsternis korrekt als ein Ereignis, bei dem sich die Erde zwischen Sonne und Mond schiebt und dadurch auf dem Mond der Erdschatten sichtbar wird. Außerdem deutete er den Halbmond als das Ereignis, bei dem die Verbindungslinie Sonne-Mond senkrecht zur Verbindungslinie Erde-Mond steht. Aus diesen beiden Interpretationen sowie den zugehörigen Messungen konnte er *relative* Größen bestimmen: das Verhältnis der Abstände Erde-Mond zu Erde-Sonne, das Verhältnis Monddurchmesser zu Sonnen-durchmesser, das Verhältnis Monddurchmesser zu Erddurchmesser und schließlich konnte er daraus das Verhältnis der Größe der Sonne zur Größe der Erde bestimmen. Seine Schlussfolgerung war, dass die Sonne wesentlich größer sein muss als die Erde, und damit sollte seiner Meinung nach die Sonne im Zentrum des Universums stehen. Dadurch wurde er einer der ersten Vertreter eines heliozentrischen Weltbilds.

In vereinfachten Rechnungen nimmt man gelegentlich an, dass der Erdschatten beim Mond denselben Durchmesser hat wie die Erde. Das setzt voraus, dass die Sonne „unendlich“ weit entfernt ist, sodass der Erdschatten durch nahezu paralleles Sonnenlicht entsteht. Diese Annahme konnte Aristarchos aber nicht machen: Erstens wollte er ja erst zeigen, dass die Sonne sehr weit von der Erde entfernt ist, und zweitens war der relative Abstand, den er aus seinen Messungen erhielt, um einen Faktor 20 kleiner als in Wirklichkeit. Außerdem steht eine solche Annahme in einem (etwas versteckten) Widerspruch zu einer in die Rechnungen eingehenden Beobachtung und führt zu einem deutlichen Fehler, wie gegen Ende dieses Abschnitts gezeigt wird.

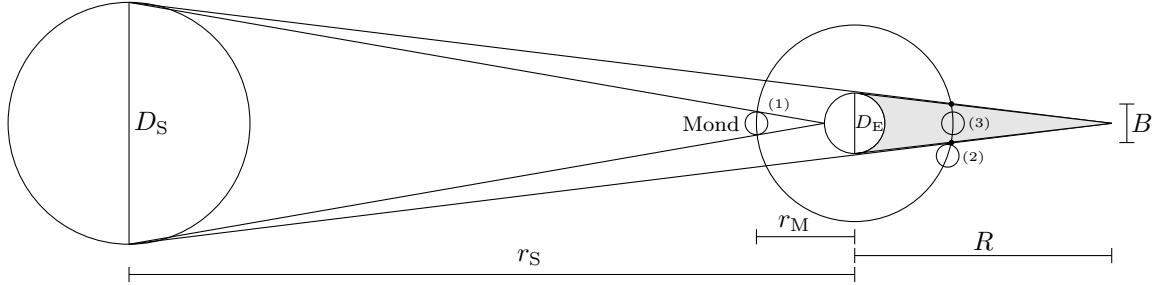


Abbildung 1: Nicht maßstabsgerechte Verhältnisse zwischen Sonne, Mond und Erde. Es sind drei Mondphasen dargestellt: (1) Der Mond bei einer Sonnenfinsternis, er erscheint genauso groß wie die Sonne; (2) der Mond tritt in den Erdschatten und (3) der Mond befindet sich im Erdschatten.  $R$  bezeichnet die Länge des Erdschattens und  $B$  seine Breite an der Stelle des Monddurchgangs.

Statt dessen nahm Aristarchos an, dass sich der Erdschatten hinter der Erde wie ein Kegel verjüngt und daher der Mond bei einer Mondfinsternis einen kleineren Schattendurchmesser durchläuft, als es dem Erddurchmesser entspricht (siehe Abb. 1). Um diese Verjüngung des Erdschattens berücksichtigen zu können, musste er erst bestimmen, wie weit die Sonne von der Erde relativ zum Mond entfernt ist.

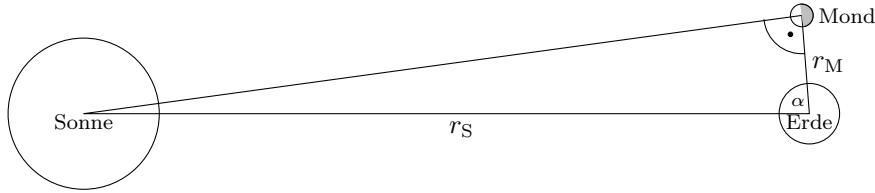


Abbildung 2: Der Mond bei Halbmond. Der Winkel, unter dem vom Mond aus betrachtet die Zentren von Sonne und Erde erscheinen, beträgt  $90^\circ$ . Der Winkel, unter dem von der Erde aus betrachtet die Zentren von Sonne und Mond erscheinen ist  $\alpha$ . Der Kosinus von  $\alpha$  entspricht dem Verhältnis von den Abständen Erde-Mond zu Erde-Sonne.

Für die Bestimmung der relativen Entfernung der Sonne ging Aristarchos von folgender Überlegung aus: Bei Halbmond ist der Winkel zwischen der Verbindungslinie Sonne-Mond und der Verbindungslinie Erde-Mond ein rechter Winkel (siehe Abb. 2). Wenn man nun hier auf der Erde den Winkel  $\alpha$  zwischen den Verbindungslinien Erde-Sonne und Erde-Mond bestimmt, erhält man das Verhältnis der Abstände von Erde-Sonne zu Erde-Mond. Heute würden wir dafür schreiben:

$$\frac{r_M}{r_S} = \cos \alpha. \quad (1)$$

Allerdings erhielt Aristarchos für diesen Winkel mit  $\alpha = 87^\circ$  einen viel zu kleinen Wert. Der tatsächliche Wert ist  $89^\circ 51'$ . Vermutlich war sich Aristarchos darüber im Klaren, dass der Wert für  $\alpha$  kleiner als  $90^\circ$  sein muss und sein Wert ist eher als eine untere Grenze zu verstehen. Mit dem Wert von Aristarchos ist der Abstand Sonne-Erde rund 19 mal größer als der Abstand Erde-Mond. In Wirklichkeit ist das Verhältnis knapp 400. Da die Sonne von der Erde aus betrachtet ähnlich groß erscheint wie der Mond (bei einer Sonnenfinsternis wird die Sonne vom Mond gerade eben bedeckt), ist die Sonne um dasselbe Verhältnis größer als der Mond - für Aristarchos rund 19 mal größer. Das bedeutet:

$$\frac{D_S}{D_M} = \frac{r_S}{r_M}. \quad (2)$$

Aus Abbildung 1 erhalten wir die folgenden geometrischen Beziehungen (in beiden Fällen aus dem Strahlensatz):

$$\frac{D_S}{D_E} = \frac{R + r_S}{R} = 1 + \frac{r_S}{R} \quad \text{und} \quad \frac{B}{D_E} = \frac{R - r_M}{R} = 1 - \frac{r_M}{R}. \quad (3)$$

Eine weitere Beziehung erhielt Aristarchos aus der Messung von zwei Zeitdauern:  $t_1$  sei die Zeitdauer zwischen dem Moment, in dem der Mond in den Schatten der Erde eintritt, bis zu dem Moment, wo er sich ganz im Erdschatten befindet, und  $t_2$  sei die Zeitdauer, in der sich der Mond vollständig im Erdschatten befindet. Das Verhältnis dieser beiden Zeiten ist

$$\frac{t_1}{t_2} = \frac{D_M}{B - D_M} \quad \text{oder} \quad B = \frac{t_1 + t_2}{t_1} D_M. \quad (4)$$

Aus den beiden Beziehungen in Gl. 3 können wir die Länge  $R$  des Erdschattens eliminieren, die Breite  $B$  des Erdschattens bei der Mondbahn können wir durch Gl. 4 ersetzen und schließlich können wir noch den Sonnendurchmesser  $D_S$  mit Hilfe von Gl. 2 durch den Monddurchmesser  $D_M$  und das bekannte Verhältnis  $\frac{r_S}{r_M}$  ersetzen. Nach einer etwas längeren Rechnung erhalten wir dann:

$$\frac{D_E}{D_M} = \frac{\left(1 + \frac{t_1 + t_2}{t_1}\right)}{1 + \frac{r_M}{r_S}}. \quad (5)$$

Auf der rechten Seite stehen nur Größen, die Aristarchos gemessen hatte. Aus seinen Beobachtungen schloss er, dass die Erde ungefähr 3 mal größer ist als der Mond (sein Wert war 2,85, der wirkliche Wert ist 3,67). Zusammen mit seinem Faktor 19 zwischen der Größe des Monds und der Größe der Sonne erhielt er somit, dass die Sonne rund 6,7 mal größer sein muss als die Erde. Der wirkliche Faktor ist knapp 110.

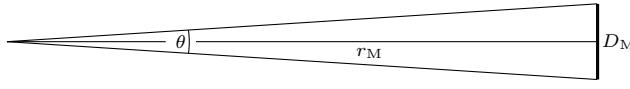


Abbildung 3: Aus dem Öffnungswinkel, unter dem der Mond von der Erde aus gesehen wird (ungefähr  $30'$ ) kann man auf das Verhältnis von Monddurchmesser  $D_M$  zum Abstand Erde-Mond  $r_M$  schließen.

Darüber hinaus konnte Aristarchos aus der scheinbaren Größe des Monds (ungefähr  $0,5^\circ$  Öffnungswinkel) das Verhältnis von der Größe des Monds  $D_M$  zu seinem Abstand  $r_M$  von der Erde bestimmen (siehe Abb. 3). Die Beziehung ist:

$$\tan \frac{\theta}{2} = \frac{D_M}{2r_M}. \quad (6)$$

Hätten wir für die Breite  $B$  des Erdschattens beim Mond einfach den Erddurchmesser angenommen, hätten wir direkt aus Gl. 4 die Beziehung

$$\frac{D_E}{D_M} = \frac{t_1 + t_2}{t_1} \quad (7)$$

erhalten. Doch dieses Ergebnis folgt nicht, wenn wir in Gl. 5 den Grenzfall  $r_S \rightarrow \infty$  nehmen. Immerhin ist in Wirklichkeit das Verhältnis  $r_M/r_S \approx 1/400$  und somit in guter Näherung vernachlässigbar. Der zusätzliche Term „ $1+$ “ im Zähler von Gl. 5 röhrt ebenfalls daher, dass wir einen kegelartigen Schatten angenommen haben, der an der Stelle des Monds schon deutlich kleiner ist als der Erddurchmesser.

Wenn  $r_S$  im Vergleich zu  $r_M$  groß wird, muss wegen Gl. 2, die wir bei der Herleitung von Gl. 5 verwendet haben, auch der Sonnendurchmesser im selben Verhältnis zunehmen. Das bedeutet aber, dass die Schattenlänge  $R$  nahezu konstant bleibt (die Korrektur hierzu wird durch den Nenner von Gl. 5 beschrieben) und damit auch das Verhältnis, um das der Erdschatten beim Mond kleiner ist. Für das Erde-Sonne-Mond-System ist  $R$  ungefähr das Vierfache von  $r_M$ , und damit ist der Erdschatten beim Mondabstand schon um ungefähr den Faktor 3:4 kleiner als der Erddurchmesser.

## 2 Eratosthenes von Kyrene

Wie wir gesehen haben, hat Aristarchos nur Verhältnisse von Größen bestimmt, keine absoluten Werte. Dazu muss mindestens eine der Größen bekannt sein. Dieses Problem löste Eratosthenes von Kyrene, der kurz nach Aristarchos lebte. Seine genauen Lebensdaten sind nicht bekannt, aber bis auf ein oder zwei Jahre lebte er von 276 v.Chr. bis 194 v.Chr. Bekannt ist Eratosthenes unter anderem durch das gleichnamige „Sieb“, mit dem man eine Tabelle der Primzahlen erhält.

Eratosthenes wusste, dass es in der Nähe von Assuan (damals Syene) einen tiefen Brunnen gab, bei dem die Sonne nur einmal im Jahr - am 21. Juni zur Mittagszeit (d.h. bei Sonnenhöchststand) - den Boden beleuchtete. Er deutete dies als die Tatsache, dass an diesem Tag die Sonne senkrecht über Assuan steht. Assuan liegt am 24. Breitengrad und somit nur wenig über dem Sommerwendekreis der Sonne (bei  $23,5^\circ$ ). Weiterhin nahm Eratosthenes an, dass die Stadt Alexandria auf demselben Längengrad wie Assuan liegt, was nicht ganz richtig ist: Alexandria liegt rund  $3^\circ$  weiter westlich. Dieser Winkel ist jedoch klein genug, um in die folgenden Überlegungen nur unwesentlich einzugehen. Schließlich wusste Eratosthenes noch, dass der Schatten eines Obelisken in Alexandria zur Mittagszeit des 21. Juni unter einem Winkel von  $1/50$ .tel eines Vollkreises (also  $\theta = 7,2^\circ$ ) fiel. Aus der Annahme, dass die Sonnenstrahlen parallel einfallen, konnte er daraus schließen, dass der Erdumfang das 50-fache des Abstands von Alexandria nach Syene beträgt.

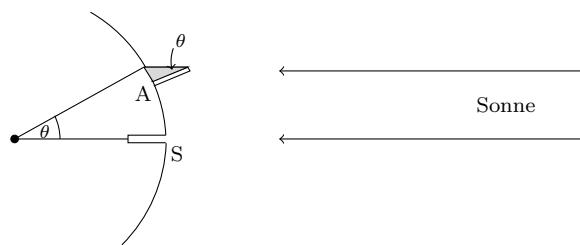


Abbildung 4: Zur Bestimmung des Erdumfangs nach Eratosthenes. Am 21. Juni steht die Sonne mittags senkrecht über einem Brunnen in Syene (S). Zum gleichen Zeitpunkt wirft ein Obelisk in Alexandria (A) einen Schatten unter dem Winkel  $\theta$ . Unter demselben Winkel erscheinen Syene und Alexandria vom Erdmittelpunkt aus betrachtet.

Wie er diesen Abstand bestimmt hat, ist nicht genau bekannt. Es gilt aber als wahrscheinlich, dass er königliche Schrittzähler eingesetzt hat, diesen Abstand zu messen. Das Ergebnis waren 5000 Stadien. Es ist auch nicht bekannt, welche Einheit für ein Stadion Eratosthenes verwendet hat (es waren damals mehrere Einheiten gebräuchlich), vermutlich aber handelte es sich um eine Einheit nahe bei 160 m.<sup>1</sup> In diesem Fall hätte er die Entfernung Alexandria-Syene nach heutigen Einheiten zu 800 km bestimmt. Da dies  $1/50$ . des Erdumfangs entsprechen soll, bestimmte er den Erdumfang zu 40 000 km. Wie genau diese Werte mit seinen Werten übereinstimmen, ist nicht bekannt, aber immerhin handelte es sich um ein wissenschaftliches Verfahren, wohingegen viele andere Werte für die Größe der Erde auf reinen Vermutungen beruhten.

<sup>1</sup>Das mesopotamische Stadion betrug 148,5 m, das griechische 177,6 m. In manchen griechischen Schriften findet man auch 157,5 m.

### 3 Die Größe der Erde

Bis in die frühe Neuzeit wurden die antiken Ergebnisse zur Bestimmung der Größe der Erde oder des Abstands Erde-Sonne kaum übertroffen. Die Zahlen von Eratosthenes blieben obskur, solange nicht geklärt war, was genau unter einem Stadion zu verstehen war. Doch die allgemeine Idee zur Bestimmung der Größe der Erde war korrekt.

Zu Beginn des 17. Jahrhunderts gab es zwei Entdeckungen, die einen wesentlichen Fortschritt brachten: (1) die Erfindung des Fernrohrs um 1610 und (2) die Kepler'schen Gesetze, insbesondere das dritte Kepler'sche Gesetz, das eine Beziehung zwischen den großen Halbachsen und den Umlaufzeiten eines Planeten um die Sonne herstellte (veröffentlicht 1619).

Mit Hilfe des Fernrohrs konnte die Erdgröße wesentlich genauer bestimmt werden. Hier ist insbesondere Jean Picard (1620–1682) zu erwähnen, der um 1670 eine sehr genaue Vermessung des Erdradius vornahm. Voraussetzung dafür waren (1) die Festlegung von zwei möglichst weit voneinander entfernten Orten auf demselben Längengrad, (2) eine möglichst genaue Bestimmung der Entfernung zwischen diesen beiden Orten und (3) eine möglichst genaue Messung der Breitengrade dieser Orte. Picard verwendete zur Winkelmessung zwischen zwei Orten Quadranten, die mit kleinen Fernrohren mit Fadenkreuzen ausgestattet waren. Dadurch waren sehr präzise Winkelmessungen möglich. Diese verwendete er einerseits, um den Breitengrad beispielsweise durch Sternbeobachtungen genau zu bestimmen, andererseits konnte er über Triangulationen<sup>2</sup> die Entfernungen zwischen zwei Punkten ausmessen, sofern die Länge einer Basislinie bekannt war. Durch weitere Anwendung dieses Verfahrens kann man auch größere Abstände bestimmen. Letztendlich genügt die Kenntnis einer Basislinie, um über genaue Winkelmessungen große Entfernungen ausmessen zu können. Ob sich zwei Orte genau auf demselben Längengrad befinden, konnte man dadurch feststellen, dass man von einem erhöhten Punkt (Gebäude, Hügel oder Bergspitze) exakt zur Mittagszeit (definiert durch den lokalen Sonnenhöchststand, der die Richtung „Süden“ anzeigt) jeweils einen Punkt im Süden und einen im Norden anvisierte.

Picard erreichte eine Genauigkeit von ungefähr 0,1% für seine Bestimmung des Erdradius. Nachdem im 18. Jahrhundert mehrere solche Messungen an verschiedenen Orten der Erde durchgeführt worden waren, erkannte man, dass die Erde keine exakte Kugelform hat. Gleiche Breitengradunterschiede hatten in Polnähe einen größeren Abstand als in Äquatornähe, was auf die Form eines abgeplatteten Ellipsoids schließen ließ.

### 4 Der Abstand Erde-Sonne

Die Idee der Parallaxenmessung wird in Abb. 5 verdeutlicht. Beobachtet man ein Objekt  $S$  von zwei verschiedenen Punkten  $a$  und  $b$  aus, deren Abstand  $D$  bekannt ist, kann man aus dem Winkel  $\alpha$ , um den dieses Objekt verschoben zu sein scheint, die Entfernung zu  $S$  bestimmen. Es wird dabei die Formel 6 verwendet. Vor einem „unendlich weit“ entfernten Hintergrund kann man den Winkel  $\alpha$  sehr leicht bestimmen, indem man den Winkel zwischen den scheinbaren Orten  $A$  und  $B$  des Objekts vor dem Hintergrund ausmisst. Nach einem ähnlichen Verfahren bestimmen wir intuitiv mit unseren Augen den Abstand von Gegenständen, wobei  $D$  dem Abstand der Augen entspricht.

Eines der Hauptargumente in der Antike gegen das heliozentrische Weltbild des Aristarchos war das Fehlen jeglicher Parallaxen. Das bezog sich einerseits auf das Fehlen von Parallaxen der

<sup>2</sup>Von zwei Punkten aus, deren Abstand bekannt ist, wird ein dritter Punkt anvisiert und es werden die beiden Winkel zwischen dem jeweils anderen Punkt und dem dritten Punkt vermessen. Daraus lassen sich die Entfernungen der beiden Punkte zu dem dritten Punkt bestimmen. Zur Kontrolle kann man von diesem dritten Punkt aus den Winkel zu den beiden Ausgangspunkten bestimmen. Jean Picard entdeckte auf diese Weise, dass Licht an unterschiedlichen Luftsichten - z.B. unterschiedlich bezüglich ihrer Temperatur oder ihres Drucks - gebrochen wurde.

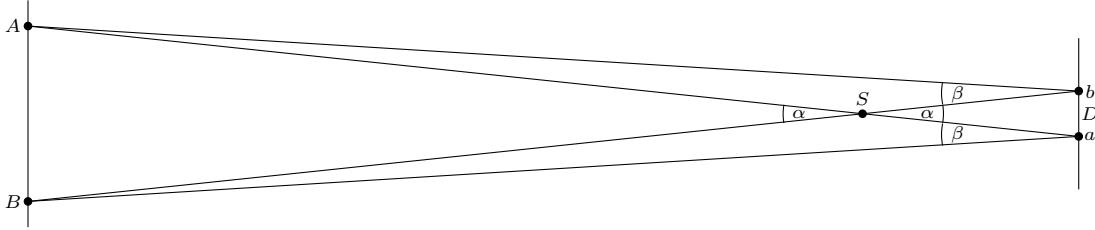


Abbildung 5: Die Parallaxenmessung. Beobachtet man ein Objekt  $S$  von zwei verschiedenen Orten  $a$  und  $b$  aus, so erscheint dieser Gegenstand um den Winkel  $\alpha$  verschoben. Die beiden Richtungen  $A$  und  $B$  erscheinen von  $a$  bzw.  $b$  aus unter dem Winkel  $\beta$ . Vor einem „unendlich weit“ entfernten Hintergrund gilt  $\alpha = \beta$ . Kennt man den Abstand  $D$  zwischen den Punkten  $a$  und  $b$ , kann man aus dem Winkel  $\alpha$  den Abstand zu  $S$  bestimmen. Selbst wenn  $A$  und  $B$  nicht unendlich weit entfernt sind, wie bei der Parallaxe der Venus vor der Sonne, kann man aus der Kenntnis der relativen Abstände und dem gemessenen Winkel  $\beta$  den Winkel  $\alpha$  bestimmen.

Planeten - wie beispielsweise des Mars, dem erdnächsten Planeten, den wir gegen einen Nachthimmel beobachten können -, wobei hier der Erdradius den Maßstab für den Abstand setzt unter dem eine Parallaxe beobachtet wird. Andererseits bezog sich das aber auch auf die fehlende Parallaxe der Sterne, wobei hier der Radius der Erdumlaufbahn um die Sonne diesen Maßstab setzt. Doch schon Archimedes hatte in einem Kommentar zu Aristarchos in seinem *Sandrechner* die zu großen Abstände zwischen den Himmelskörpern für die nicht beobachteten Parallaxen verantwortlich gemacht.<sup>3</sup>

Grundsätzlich war die Idee von Aristarchos zur Bestimmung des Verhältnisses  $r_M/r_S$  richtig, es setzt aber voraus, dass man den Winkel zwischen Mond und Sonne bei Halbmond genau messen kann. Die Fehler beruhen sowohl auf der Bestimmung des Zeitpunkts, wann genau Halbmond ist, als auch auf der exakten Messung des Winkels, der sehr nahe bei  $90^\circ$  liegt. Kleine Messfehler führen hier zu großen Unsicherheiten, sodass ähnliche Messungen in den Folgezeiten keine wesentlich besseren Ergebnisse erbrachten. Erst um 1635 wurde die Messung von Godefroy Wendelin nach dem Verfahren von Aristarchos mit einem Fernrohr wiederholt, was zu etwas besseren Werten führte: Er bestimmte den Abstand Erde-Sonne zu ungefähr 90 000 000 km.

Ähnlich wie schon bei Aristarchos waren die relativen Abstände in unserem Sonnensystem spätestens seit den Kepler'schen Gesetzen sehr gut bekannt. Das dritte Kepler'sche Gesetz stellt eine Beziehung zwischen den Umlaufzeiten der Planeten und ihren großen Halbachsen her, und die Umlaufzeiten ließen sich sehr genau bestimmen. Allerdings bezieht sich das dritte Kepler'sche Gesetz auf denselben Zentralkörper (der Faktor zwischen dem Quadrat der Umlaufzeit und der dritten Potenz der Halbachse einer Bahn hängt von der Masse des Zentralkörpers ab), sodass man aus den Umlaufzeiten und Halbachsen beispielsweise der Mondbahn nicht auf die entsprechenden Größen bei den Planetenbahnen schließen konnte. Immerhin wusste man nun, dass man lediglich eine absolute Entfernung im Planetensystem bestimmen musste, um alle anderen Entfernungen zwischen den Planeten und der Sonne bzw. den Planeten untereinander ebenfalls zu kennen.

Im 17. Jahrhundert wurden mit Hilfe des Fernrohrs Versuche unternommen, aus einer Parallaxenmessung des Mars die Entfernung Mars-Erde bei deren Minimum zu bestimmen. Damit hätte man auch den Abstand Sonne-Erde gekannt. Jean Richter und Giovanni Domenico Cassini erhielten

<sup>3</sup>Er hatte im *Sandrechner* explizit angenommen, dass sich der Abstand der Sterne zum Abstand Erde-Sonne ähnlich verhält wie der Abstand Erde-Sonne zum Durchmesser der Erde. Damit brachte er zum Ausdruck, dass der Abstand Erde-Sonne zum Durchmesser der Erde zu groß für eine Parallaxenmessung der Sonne von der Erde aus ist, und entsprechend der Abstand der Sterne zu groß, um im Verlauf eines Jahres von verschiedenen Punkten der Erdbahn aus beobachtet zu werden.

auf diese Weise einen Wert für die Astronomische Einheit, der immerhin schon weniger als 10% vom tatsächlichen Wert abwich. Auf einen ähnlichen Wert kam auch John Flamsteed. Im Gegensatz zu Richter und Cassini, die die Mars-Parallaxe von Orten auf zwei verschiedenen Breitengraden vornahmen (Paris und Cayenne in Französisch Guayana, knapp über dem Äquator), nahm Flamsteed die Messung alleine vor, indem er die Parallaxe zwischen Abend und Morgen beobachtete und somit die Drehung der Erde ausnutzte (da sich die Erde in dieser Zeit um eine unbekannte Distanz weiter um die Sonne bewegte, wobei diese Distanz wiederum nur über den Abstand Erde-Sonne bestimmt werden konnte, wurde die Rechnung etwas komplizierter).

1639 fand ein Venus-Transit vor der Sonne statt, den Jeremiah Horrocks nutze, um eine Parallaxenmessung der Venus vor der Sonne zu messen. Er erhielt einen ähnlichen Wert wie Wendelin mit einem Fehler von rund 40%.

Edmund Halley (basierend auf Arbeiten von James Gregory) hatte 1716 in einem Artikel gezeigt, dass eine Beobachtung einer Venus-Parallaxe vor der Sonne zu genaueren Ergebnissen führen könnte, wenn man nicht die Parallaxenwinkel direkt ausmisst, sondern die Zeiten misst, für die die Venus vor der Sonnenscheibe sichtbar ist. Aus diesen Zeitdauern kann der Winkel  $\alpha$  der Parallaxe bestimmt werden. Die Tatsache, dass die Sonne nicht unendlich weit entfernt ist und insofern die beobachtete Parallaxe vor der Sonne nicht gleich dem Parallaxenwinkel ist, macht das Problem nicht wesentlich komplizierter, da man aus den Kepler'schen Gesetzen das Verhältnis der Abstände kannte und daraus die richtige Parallaxe berechnen konnte (Abb. 6).

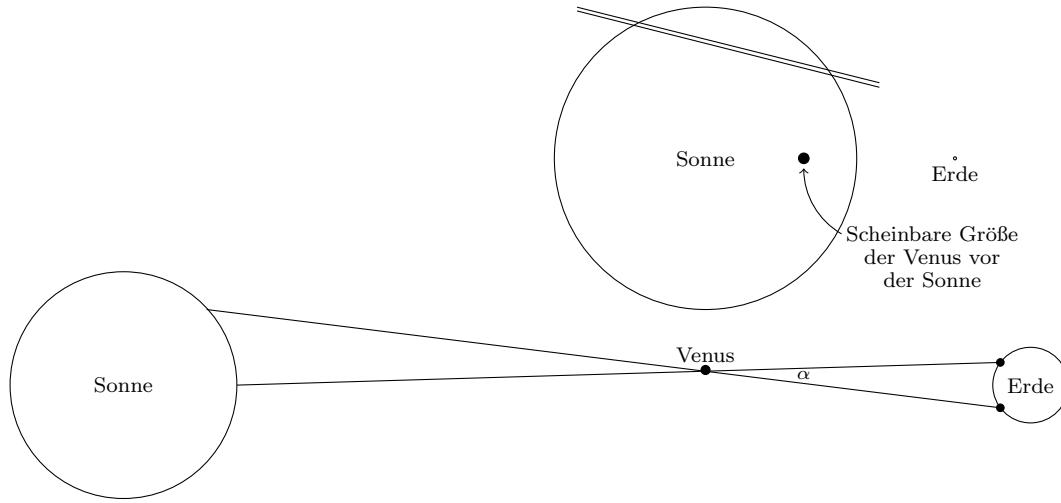


Abbildung 6: Die Parallaxenmessung bei einem Venus-Transit vor der Sonne. (unten) Die Verhältnisse sind übertrieben dargestellt. Halley hatte vorgeschlagen, von zwei verschiedenen Orten auf der Erde die genauen Zeiten zu messen, für die die Venus vor der Sonne sichtbar ist. (oben) Bis auf den Abstand Sonne-Erde sind hier die Größen ungefähr maßstabsgerecht dargestellt. Die Parallaxe der Venus vor der Sonne ist winzig und die scheinbaren Trajektorien der Venus vor der Sonne liegen sehr dicht beieinander, selbst wenn man sie von weit entfernten Orten auf der Erde beobachtet. Obwohl die Venus einen etwas kleineren Durchmesser hat als die Erde, erscheint sie vor der Sonne größer, da sie sich näher an der Erde befindet.

Die tatsächlichen Verhältnisse sind jedoch so, dass diese Zeiten mit einer sehr großen Genauigkeit bestimmt werden müssen, da die Venus-Trajektorien vor der Sonnenscheibe sehr dicht beieinander liegen (Abb. 6, oben). Der Winkelabstand der beiden Trajektorien ist kleiner als der Winkelabstand für den scheinbaren (projizierten) Durchmesser der Venus. Damit ist auch der Zeitpunkt schwer definierbar, wann genau die Venus in den Bereich der Sonnenscheibe eintritt oder

diesen verlässt (eine zusätzliche Problematik hierbei ist das sogenannte Tropfenphänomen: Aufgrund der endlichen Auflösung optischer Geräte scheint die Venus am inneren Rand der Sonne mit dem dunklen Hintergrund tropfenförmig zu verschmelzen).

In den Jahren 1761 und 1769 gab es Venus-Transite, und in diesen Jahren fanden Expeditionen statt, um die Venus-Parallaxe zu beobachten - die wichtigste im Jahr 1769, unter anderem mit James Cook in Tahiti am 17,6-ten südlichen Breitengrad. Der nördlichste Beobachtungsort war Vardø am 70-sten nördlichen Breitengrad in Norwegen. Insgesamt gab es bei den beiden Venus-Transiten weit über einhundert Beobachtungen, verteilt über die ganze Welt. Die Daten wurden von Jérôme Lalande zusammengetragen und ausgewertet. Der Fehler in der Bestimmung des Abstands Erde-Sonne betrug letztendlich weniger als 2%.

Heute ist die Astronomische Einheit definiert als die Länge  $1 \text{ AU} = 149\,597\,870\,700 \text{ m}$ . Da die Sonne jedoch ständig an Energie und damit Masse verliert, entfernen sich die Planeten von der Sonne (die Erde um rund 15 cm im Jahr). Daher wird der Sinn für eine derart festgelegte Konstante gelegentlich angezweifelt.

## 5 Parallaxenmessung der nächsten Sterne

Mit der Astronomischen Einheit sind nicht nur die Abstände der Planeten zur Sonne bzw. der Planeten untereinander sowie vieler weiterer Objekte in unserem Sonnensystem bestimmt, sondern damit steht auch eine neue Basis für die Messung von Parallaxen zu weiter entfernten Objekten zur Verfügung. Während man insbesondere in der populärwissenschaftlichen Literatur das Lichtjahr gerne als astronomische Entfernungseinheit verwendet, also die Distanz, die das Licht in einem Jahr zurücklegt, verwendet man in der Astrophysik eher das Parsec, die Parallaxensekunde (abgekürzt pc), als Entfernungseinheit.

Bei einer Geschwindigkeit von 300 000 km pro Sekunde benötigt das Licht rund 500 Sekunden von der Sonne bis zur Erde, das entspricht 8 Minuten und 20 Sekunden. In einem Jahr legt das Licht eine Strecke von  $9,4673 \cdot 10^{12} \text{ km}$  zurück. Das ist somit die Distanz, die einem Lichtjahr entspricht - ungefähr  $10^{13}$  Kilometer. Eine Parallaxensekunde ist definiert als der Abstand, bei dem die Astronomische Einheit, also der Radius der Erdbahn um die Sonne, unter einem Winkel von 1 Bogensekunde gesehen wird. Eine Bogensekunde sind  $1/3600$  Grad, und für eine Parallaxensekunde erhalten wir  $1 \text{ pc} = 1 \text{ AU} / \tan(1/3600) \approx 30,857 \cdot 10^{12} \text{ km}$ . Das entspricht 3,26 Lichtjahren.

Der nächste Stern (nach der Sonne) ist Proxima Centauri mit einem Abstand von 4,25 ly (Lichtjahren). Hierbei handelt es sich um einen Roten Zwerg, der mit bloßem Auge oder auch einem Fernglas nicht sichtbar ist. Im Abstand von rund 4,38 ly ist Alpha-Centauri, einer der hellsten Sterne am Himmel. Eigentlich handelt es sich um ein Doppelsternsystem, bestehend aus Alpha-Centauri A und Alpha-Centauri B, die jedoch nur mit einem Fernglas oder Fernrohr trennbar sind. Diese Objekte sind also schon weiter als eine Parallaxensekunde von der Erde entfernt und somit bedarf es sehr hochauflösender Teleskope, um die Entfernung zu diesen Sternen mit dem Verfahren der Parallaxe bestimmen zu können. Die erste Veröffentlichung einer solchen Messung stammt aus dem Jahre 1838, als Friedrich Wilhelm Bessel die Entfernung von Cygni-61 (einem bei guten Bedingungen mit bloßem Auge sichtbaren Stern im Sternbild Schwan) mithilfe einer Parallaxenmessung mit 10,4 ly angab (der tatsächliche Wert ist 11,4 ly). Allerdings hatte Thomas Henderson schon 1834 eine Parallaxenmessung von Alpha-Centauri vorgenommen, deren Ergebnisse er aber erst 1839, nachdem er von den Resultaten von Bessel erfuhr, veröffentlichte.

In beiden Fällen war bekannt, dass diese Sterne eine hohe Eigenbewegung haben. Man unterscheidet dabei die radiale Bewegung relativ zur Erde, die sich heute sehr gut mit spektroskopischen Mitteln (Doppler-Effekt) bestimmen lässt, und die tangentiale Bewegung relativ zur Erde, die nur

über mehrere Beobachtungen über einen längeren Zeitraum ermittelt werden kann. Diese tangentiale Bewegung wird meist in der Einheit „mas/yr“ (milliarcseconds per year) angegeben, da für ihren absoluten Wert (in km/s) die Entfernung bekannt sein müsste. Aus den hohen Eigenbewegungen von Cygni-61 und Alpha-Centauri schlossen sowohl Bessel als auch Henderson, dass diese Objekte uns vergleichsweise nahe sein müssten.

In den Jahren 1989 bis 1993 konnte der Satellit Hipparcos (für *High Precision Parallax Collecting Satellite*) die astrometrischen Daten (dazu zählen die Deklination, die Reklination, die Parallaxe - also der Abstand -, sowie die tangentialen und radialen Geschwindigkeiten) von fast 120 000 Sternen mit einer Genauigkeit von 0,001 Bogensekunden vermessen. Seit 2013 ist der Satellit Gaia (*Global Astrometric Interferometer for Astrophysics*) in Operation (vermutlich bis 2025). Für Sterne bis zu einer Magnitude von 7 soll hier eine Genauigkeit von 7 Mikrobogensekunden ( $\mu\text{as}$  - microarcseconds) erreicht werden. Der Abstand zu rund 20 Millionen Sternen wird dann mit einer Genauigkeit von unter 1% bekannt sein. Sämtliche Sterne mit einer Magnitude unter 20 und innerhalb eines Abstands von 30 000 ly werden mit einer Genauigkeit von unter 10% vermessen (das schließt unser galaktisches Zentrum mit ein).

## 6 Standardkerzen

Die Entfernungsbestimmung über eine Parallaxenmessung ist das einzige direkte Verfahren, die Entfernung zu Himmelsobjekten zu bestimmen. Die Abstände zu weiter entfernten Objekten lassen sich nur indirekt messen. Ein wichtiges Verfahren in diesem Zusammenhang beruht auf sogenannten Standardkerzen. Dabei handelt es sich um Objekte, bei denen die absolute Helligkeit aufgrund bestimmter Eigenschaften dieser Objekte bekannt ist, und aus der beobachteten scheinbaren Helligkeit dieser Objekte kann man ihre Entfernung bestimmen. Zur „Eichung“ dieses Verfahrens muss man die Entfernung zu einigen Vertretern dieser Standardkerzen jedoch direkt, d.h. über die Parallaxenmessungen, bestimmt haben. Aus diesem Grund sind die direkten Entfernungsbestimmungen solcher Standardkerzen auch von Bedeutung für unser kosmisches Weltbild.

### 6.1 Absolute und scheinbare Magnitude

Die Magnitude ist ein Helligkeitsmaß, das vermutlich auf Hipparch (um 190 - 120 v. Chr.) zurückgeht und ausführlich von Claudius Ptolemäus (um 100 n. Chr. bis rund 160 n. Chr.) verwendet wurde. Bei diesem Maß wurden ursprünglich die sichtbaren Sterne in 6 Helligkeitsklassen eingeteilt, wobei Klasse 1 die hellsten Sterne enthielt und Klasse 6 die Sterne, die unter guten Bedingungen gerade eben noch beobachtbar waren. Nach den Fechner-Weber'schen Gesetzen ist unsere subjektive Wahrnehmung proportional zum Logarithmus der Intensität der Einwirkung, wobei Intensität einer „Energie pro Fläche pro Zeiteinheit“ entspricht. Es handelt sich bei der Intensität also um eine Energie, die pro Zeiteinheit (damit erhalten wir eine Leistung) auf eine Flächeneinheit übertragen wird. Dies gilt nahezu unabhängig von dem Wahrnehmungssinn - also visuelle oder auditive Wahrnehmung, Schmerz- oder Kälteempfindung, etc. Insbesondere bedeutet dies, dass die subjektiv wahrgenommene Helligkeit proportional zum Logarithmus der Energie ist, die pro Zeiteinheit in unser Auge trifft.

Um einerseits ein objektiveres Helligkeitsmaß zu erhalten, andererseits möglichst nahe an dem Maß zu bleiben, das sich im Verlauf der Jahrhunderte in der Astronomie eingebürgert hatte, definiert man heute die Magnitude über folgende Beziehungen: Ganz grob entspricht der subjektiv wahrgenommene Unterschied zwischen der Helligkeitsstufe 1 und der Helligkeitsstufe 6 einem Faktor 100 in der Intensität der Strahlung. Sei  $I_i$  die Intensität zur Magnitude  $m = i$ , dann gilt somit  $I_1 = 100 I_6$  oder  $I_{i-1} = \sqrt[5]{100} I_i = 10^{2/5} I_i$ . Andererseits folgt aus dem Fechner-Weber'schen Gesetz

$m = \alpha \log I$  (mit zunächst unbekanntem Faktor  $\alpha$ ). Damit erhalten wir:

$$5 = m_6 - m_1 = \alpha \log I_6 - \alpha \log I_1 = \alpha \log \left( \frac{I_6}{I_1} \right) = \alpha \log \left( \frac{I_6}{100I_6} \right) = -\alpha \log 100 = -2\alpha \quad (8)$$

oder  $\alpha = -5/2 = -2,5$ . Die Magnitude  $m$  wird heute über die Intensität  $I$  einer Quelle durch die folgende Beziehung definiert:

$$m = -2,5 \log I/I_0, \quad (9)$$

wobei  $I_0$  eine willkürlich gewählte Referenzintensität der Magnitude 0 ist (früher definierte man die Magnitude des Sterns Wega im Sternbild Leier als Magnitude 0).

Wir messen hier auf der Erde von einem Stern bzw. einem astronomischen Objekt die sogenannte *scheinbare Helligkeit*, also die Lichtintensität, die hier auf der Erde ankommt. Nun ist bekannt, dass die beobachtete Intensität einer Lichtquelle wie  $1/r^2$  abnimmt, wobei  $r$  der Abstand von der Quelle ist. Die abgestrahlte Energie verteilt sich über eine Kugeloberfläche  $4\pi r^2$ , und da die Energie erhalten ist, nimmt die Energie pro Fläche wie  $1/r^2$  ab. Das setzt voraus, dass es keine absorbierenden Medien zwischen Quelle und Empfänger gibt, ansonsten nimmt die Intensität schneller ab.

Daraus folgt, dass sich die scheinbare Magnitude von zwei Quellen, welche dieselbe Intensität an Energie abstrahlen, sich aber in unterschiedlichen Entfernen  $r_1$  und  $r_2$  vom Beobachter befinden, um

$$m_1 - m_2 = -2,5 \log \left( \frac{I_1}{I_2} \right) = -2,5 \log \left( \frac{I}{r_1^2} \cdot \frac{r_2^2}{I} \right) = -2,5 \log \left( \frac{r_2^2}{r_1^2} \right) = -5 \log \left( \frac{r_2}{r_1} \right) \quad (10)$$

unterscheiden. Man definiert nun für ein astronomisches Objekt ein von der Entfernung unabhängiges Maß, die sogenannte *absolute Helligkeit*, durch die Bedingung: Die absolute Helligkeit  $I_0$  eines Objekts ist gleich der scheinbaren Helligkeit, die dieses Objekt hätte, wenn es sich in einer Entfernung von  $r_0 = 10$  pc (oder 32,6 Lichtjahren) befände. Kennen wir somit die Entfernung eines Objekts von der Erde, können wir seine absolute Helligkeit berechnen. Umgekehrt, kennen wir die absolute Helligkeit eines Objekts, können wir aus der scheinbaren Helligkeit auf seine Entfernung schließen. Sei die scheinbare, auf der Erde gemessene Magnitude eines Objekts  $m_1$  und sei  $m_0$  die (aus anderen Überlegungen bekannte) absolute Magnitude dieses Objekts, dann folgt:

$$m_1 - m_0 = -5 \log \left( \frac{r_0}{r} \right) \quad \text{oder} \quad r = r_0 \cdot 10^{\frac{(m_1 - m_0)}{5}} \quad \text{mit } r_0 = 10 \text{ pc}. \quad (11)$$

Standardkerzen sind nun Objekte, deren absolute Helligkeit bekannt ist, sodass wir aus der Beobachtung ihrer scheinbaren Helligkeit hier auf der Erde auf ihren Abstand schließen können.

## 6.2 Veränderliche

Es gibt sehr viele Formen von Veränderlichen, das sind Himmelsobjekte, deren Helligkeit Schwankungen unterworfen ist. Streng genommen gehört auch unsere Sonne dazu, die wegen ihrer periodischen Fluktuationen in den Sonnenflecken (mit einer Periode von 11 Jahren) auch in ihrer Helligkeit schwankt. Diese Schwankungen sind aber minimal. Es gibt jedoch Sterne, die in manchen Phasen ihrer Entwicklung deutlich größeren Helligkeitsschwankungen unterworfen sind. Diese Schwankungen beruhen sowohl auf Veränderungen in ihrer Größe als auch in ihrer Temperatur. Nicht-lineare Rückkopplungen in der Dynamik können solche Effekte hervorrufen: Ein Beispiel sind Sterne, bei denen die Temperatur der äußeren Hülle nahe der Ionisierungsenthalpie von Wasserstoff und Helium liegt (das sind Temperaturen zwischen 6000 und 9000 K). Die Lichtdurchlässigkeit solcher Schichten hängt sehr vom Ionisationsgrad ab: Mehr ionisierte Elemente bedeutet eine höhere Absorption von Licht, dadurch mehr Energieaufnahme und eine Zunahme der Ionisation, und entsprechend umgekehrt. Das

Zusammenspiel solcher Effekte kann zu periodischen Schwankungen in der Helligkeit von mehreren Magnituden führen.

Die vermutlich wichtigste Klasse von Veränderlichen, die als Standardkerzen dienen, sind die Cepheiden. Benannt sind sie nach  $\delta$  Cephei im Sternbild Kepheus. Die Variabilität dieses Sterns schwankt zwischen  $m = 3,48$  und  $4,37$  und ist seit Ende des 18. Jahrhunderts bekannt. Er ist ungefähr 800 Lichtjahre von uns entfernt (nach Parallaxenmessungen von Hipparcos) und die Pulsationsperiode beträgt rund 5,37 Tage.

Die Bedeutung der Cepheiden als Standardkerzen geht auf Henrietta Swan Leavitt (1868–1921) zurück. Sie arbeitete Anfang des 20. Jahrhunderts in der Gruppe der „Harvard Computers“, einer Gruppe von Frauen, die von dem Astrophysiker Edward Charles Pickering angeheuert worden war, um astronomische Daten auszuwerten. Sie untersuchte Veränderliche in der Magellanschen Wolke, die auf photographischen Platten registriert worden waren (damals durften Frauen noch keine Teleskope bedienen). Bei ihren sehr sorgfältigen Untersuchungen stellte sie fest, dass es eine logarithmische Beziehung zwischen der Periode und der mittleren Helligkeit dieser Veränderlichen gibt. Da sie davon ausgehen konnte, dass sich diese Objekte alle in mehr oder weniger derselben Entfernung von der Erde befinden (in der Magellanschen Wolke), erkannte sie, dass sich eine solche Beziehung zur Entfernungsmessung eignet. Nachdem der Abstand zu einigen Cepheiden in unserer Milchstraße mithilfe anderer Verfahren bestimmt worden war, konnte man mit ihrer Entdeckung nun auch den Abstand von Objekten in bis zu 20 Millionen Lichtjahren Entfernung bestimmen. Insbesondere konnte Edwin Hubble auf diese Weise zeigen, dass der Andromeda-Nebel nicht zu unserer Galaxie gehört - damit wurde die „Shapley-Curtis-Debatte“ oder auch „große Debatte“ von 1920 entschieden - und später wurde basierend auf ihrer Entdeckung ebenfalls von Hubble die Expansion des Universums entdeckt. Henrietta Leavitt wurde von einem Mitglied der Schwedischen Akademie der Wissenschaften für den Nobelpreis 1925 vorgeschlagen, allerdings stellt sich dann heraus, dass sie zu diesem Zeitpunkt schon seit drei Jahren tot war.

Anfang der 50er Jahre des 20. Jahrhunderts musste eine Korrektur in der Abstandsbestimmung vorgenommen werden, nachdem man erkannte, dass es zwei Arten von Cepheiden gibt, die sich hauptsächlich in Bezug auf ihr Alter unterscheiden. Die Entfernungsbestimmung nach der Parallaxenmethode war an „alten“ Cepheiden (heute W Virginis oder Typ II Cepheiden genannt) vorgenommen worden, wohingegen es sich bei den Cepheiden im Andromeda-Nebel um „junge“ Cepheiden (heute auch „klassische Cepheiden“ oder Typ I Cepheiden genannt) handelt, bei denen sich die Periode-Helligkeitsbeziehung um rund 1,5 Magnituden unterscheidet. Das führte im Wesentlichen dazu, dass fast alle extragalaktischen Entfernungsmeßungen um teilweise mehr als einen Faktor 2 falsch und entsprechend vergrößert werden mussten.

### 6.3 Statistische Verfahren

Statistische Verfahren beruhen nicht auf den Abstand-Helligkeits-Beziehungen einzelner Objekte sondern auf vergleichbaren Beziehungen für eine Verteilung von Objekten innerhalb einer bestimmten Population. Beispielsweise ist die Größenverteilung der Sterne und damit ihre Helligkeitsverteilung in verschiedenen Galaxien sehr ähnlich. Es gibt eine obere Grenze für die Größe und damit auch die Helligkeit eines Sterns, es gibt Beziehungen zwischen der Größe und der spektralen Energieverteilung eines Sterns etc. Solche statistischen Beziehungen kann man ausnutzen, um beispielsweise die Entfernung von Galaxien oder auch die Entfernung von Sternenhaufen (globulären Clustern) zu bestimmen.

## 6.4 Supernovae Typ Ia

Eine besonders wichtige Klasse von Standardkerzen sind sogenannte Typ Ia Supernovae. Bei einer Supernova handelt es sich um das explosive Endstadium eines Sterns: Der Stern im- bzw. explodiert unter seiner eigenen Schwerkraft, die nicht mehr durch andere Prozesse wie den Strahlungsdruck der zentralen Kernfusion aufgehalten wird, zu einem Neutronenstern oder auch zu einem schwarzen Loch. Bei solchen Prozessen sind auch die meisten Elemente schwerer als Eisen in unserem Universum entstanden.

Bei einer Supernova vom Typ Ia handelt es sich (vermutlich) um ein Doppelsternsystem, bei dem einer der Partner ein weißer Zwerg ist. Dieser weiße Zwerg entzieht seinem Partner - einem normalen Stern - Materie und wird dadurch langsam schwerer. Überschreitet die Masse eines solchen weißen Zwergs die kritische Grenze von 1,4 Sonnenmassen (die sogenannte Chandrasekhar-Grenze), kann die Wirkung der gravitativen Kraft nicht mehr aufgehalten werden: Es kommt zu nuklearen Reaktionen, bei denen Protonen und Elektronen sich zu Neutronen verbinden und schließlich ein Neutronenstern entsteht. Dieser Prozess findet bei nahezu denselben Ausgangsbedingungen (kritische Masse des weißen Zwergs) statt und verläuft daher unter denselben Bedingungen. Aus diesem Grunde glaubt man heute, dass die absolute Helligkeit solcher Typ Ia Supernovae auch nahezu konstant ist und sich daher als Standardkerze eignet.

Da die Helligkeit eines Sterns bei einer Supernova die Helligkeit von Billionen Sternen bzw. die Helligkeit einer ganzen Galaxie erreichen kann, sind solche Ereignisse auch in sehr großen Entfernung beobachtbar. Ob es sich bei einer Supernova um eine Supernova vom Typ Ia handelt, kann man an verschiedenen Parametern erkennen; ein wesentliches Anzeichen ist das Vorhandensein einer Siliziumlinie im Lichtspektrum. Außerdem kann man aus dem Verlauf der Helligkeitskurve als Funktion der Zeit Rückschlüsse auf die Art der Supernova schließen.

Die systematische Untersuchung solcher Supernovae vom Typ Ia in den 90er Jahren des letzten Jahrhunderts führte dazu, dass man Abweichungen vom linearen Hubble-Gesetz (siehe Abschnitt 7) erkannte: Der Abstand von sehr weit entfernten Galaxien nach der Rotverschiebung - also dem Hubble-Gesetz - stimmte nicht mit den Abstandsmessungen basierend auf beobachteten Supernovae Typ Ia überein. Daraus konnte man schließen, dass die Beziehung zwischen der Geschwindigkeit, mit der sich ein Objekt von uns entfernt, und seinem Abstand von der Erde nicht linear ist und somit nicht einer linearen Ausdehnungsbeziehung in unserem Universum entspricht. Die Untersuchungen zeigten, dass sich unser Universum seit rund acht Milliarden Jahren beschleunigt ausdehnt. Dies führte dazu, dass man von einer „dunklen Energie“ in unserem Universum ausgeht - im Wesentlichen erklärt man dies heute durch eine negative kosmologische Konstante - deren Wirkung darin besteht, dass sich der Raum beschleunigt ausdehnt.

## 7 Die kosmische Rotverschiebung

Fast jedes selbststrahlende Himmelsobjekt (Sterne oder Galaxien) zeigt in seinem elektromagnetischen Spektrum (erweitert um den infraroten Bereich und den UV-Bereich) charakteristische Linien, entweder als Absorptionslinien oder als Emissionslinien. Absorptionslinien zeigen sich in einer Spektralzerlegung als dunkle Streifen vor einem im Wesentlichen thermischen Spektrallhintergrund. Sie entstehen durch die Absorption bestimmter Frequenzen durch chemische Elemente in den äußeren Schichten dieser Objekte. Emissionslinien sieht man, wenn bestimmte Elemente dominant sind, so dass ihr emittiertes Licht den Hintergrund überstrahlt. In der Astronomie von Sternen oder Galaxien findet man hauptsächlich Absorptionslinien. In Abhängigkeit von der Natur der emittierenden Objekte (Temperatur, chemische Zusammensetzung, etc.) können diese Linien an unterschiedlichen

Stellen auftreten und unterschiedlichen Frequenzen entsprechen.

Bei einer Rotverschiebung sind die charakteristischen Linien in einem Spektrum systematisch zu längeren Wellenlängen verschoben. Im umgekehrten Fall - Verschiebung zu kürzeren Wellenlängen - spricht man von Blauverschiebung. Die klassische Ursache für solche Verschiebungen ist der Doppler-Effekt, der auftritt, wenn sich die strahlungsaussendenden Objekte relativ zum empfangenden Objekt bewegen. In der Astronomie findet man diesen Effekt bei der sogenannten Pekuliarbewegung oder Pekuliargeschwindigkeit eines Objekts, also seiner Eigenbewegung relativ zu seiner Umgebung. Beispielsweise bewegt sich die Andromedagalaxie auf uns zu und zeigt eine Blauverschiebung. Außerdem gibt es noch die gravitative Rotverschiebung, wenn sich Licht von einer gravitativen Quelle entfernt, d.h. in einem Bereich hohen Gravitationspotenzials emittiert und in einem Bereich niedrigeren Gravitationspotenzials registriert wird. Die dritte Ursache - und um die geht es hier - ist die kosmische Rotverschiebung. Sie tritt auf, wenn sich der Abstand zwischen dem Objekt, welches das Licht emittiert, und dem Objekt, welches das Licht registriert, aufgrund der Raumausdehnung verändert.

Für eine gegebene Spektralzerlegung kennzeichnet man die Rotverschiebung durch den Faktor

$$z = \frac{\Delta\lambda}{\lambda} = \frac{\lambda_r - \lambda_e}{\lambda_e} = \frac{\lambda_r}{\lambda_e} - 1, \quad (12)$$

wobei  $\lambda_r$  die beobachtete (registrierte) Wellenlänge und  $\lambda_e$  die von dem Objekt emittierte Wellenlänge sind. Bei einer Rotverschiebung ist die beobachtete Wellenlänge größer als die emittierte Wellenlänge, sodass  $z$  positiv ist.

Heute interpretieren wir die kosmische Rotverschiebung als Effekt der Raumausdehnung. Diese Ausdehnung streckt auch die Wellenlängen, sodass das Verhältnis  $\lambda_r/\lambda_e = z + 1$  direkt den Faktor angibt, um den die Ausdehnung (genauer der Skalenfaktor) des Universums zwischen Emission und Registrierung der Strahlung zugenommen hat.

Die ersten Beobachtungen von Rotverschiebungen bei Spiralgalaxien stammten von Vesto Slipher aus der Zeit zwischen 1912 und 1917. Damals war noch nicht klar, ob es sich bei diesen Objekten um Nebel in unserer Milchstraße oder „Inseluniversen“ handelt. Erst Edwin Hubble konnte um 1924 klären, dass viele der damals bekannten Nebel, einschließlich des Andromeda-Nebels und des Triangulum-Nebels, nicht zu unserer Milchstraße gehören. Er verwendete dazu die von Henrietta Swan Leavitt entdeckte Cepheiden-Methode, wobei er diese Standardkerzen zunächst an Cepheiden in unserer Galaxie eichen musste (wie sich später herausstellte, waren viele seiner Entfernungsmessungen teilweise um bis zu einem Faktor 7 falsch, allerdings waren die relativen Entfernungen im Wesentlichen korrekt). Auf diese Weise entdeckte Hubble zusammen mit seinem Assistenten Milton Lasell Humason, dass es näherungsweise eine Proportionalität zwischen der Rotverschiebung von Galaxien und ihrer Entfernung gab. Diese lineare Beziehung zwischen Rotverschiebung und Entfernung bezeichnet man heute als Hubble-Gesetz oder auch Hubble-Lemaître-Gesetz (der belgische Theologe und Astrophysiker Georges Edouard Lemaître hatte dieses Gesetz aus seiner Urknalltheorie zwei Jahre vor Hubble postuliert).

Die ursprüngliche Version des Hubble-Gesetzes lautet somit:

$$z \propto D, \quad (13)$$

wobei  $D$  der Abstand zwischen dem Objekt und uns ist und  $z$  die an diesem Objekt beobachtete Rotverschiebung. Interpretiert man die Rotverschiebung als einen Doppler-Effekt, und dies war zu Zeiten von Slipher und Hubble naheliegend, kann man ihr eine Geschwindigkeit zuordnen, wobei für nicht zu große Geschwindigkeiten eine lineare Beziehung,  $z = v/c$ , besteht. Für das Hubble-Gesetz definiert man formal eine Rotverschiebungsgeschwindigkeit  $v_{rs} = cz$  und gelangt somit zu dem Gesetz:

$$v_{rs} = HD. \quad (14)$$

$H$  bezeichnet man als die Hubble-Konstante, die allerdings zeitabhängig sein kann. Diese Formulierung des Hubble-Gesetzes ist in mehrfacher Hinsicht problematisch:

1.  $z$  kann Werte größer als 1 annehmen, womit  $v_{rs}$  größer als die Lichtgeschwindigkeit wird. Der Rekord einer gemessenen Rotverschiebung bei einer Galaxie mit dem Deep Space Telescope des Hubble Satelliten liegt derzeit bei  $z \approx 10$ , d.h., das Universum hat sich seit der Zeit, als dieses Licht ausgesandt wurde, um einen Faktor 11 ausgedehnt. Nimmt man ganz grob eine lineare Ausdehnung an, was allerdings bei großen  $z$ -Werten problematisch ist, schaut man hier über 12 Milliarden Jahre in die Vergangenheit. Ein genauerer Wert liegt bei 13,2 Milliarden Jahren und somit stammt unsere heutige Wahrnehmung aus einer Zeit, in der das Universum weniger als eine Milliarde Jahre alt war.

Gelegentlich verwendet man die Beziehung des relativistischen (longitudinalen) Doppler-Effekts zwischen  $z$  und  $v$ ,

$$z + 1 = \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}}, \quad (15)$$

um einer Rotverschiebung eine Geschwindigkeit kleiner als  $c$  zuzuschreiben, doch auch dies wird der kosmologischen Rotverschiebung nicht gerecht und sollte eher vermieden werden.

2. Wir sehen Objekte nicht nur in großer Entfernung sondern auch in der Vergangenheit und  $H$  ist in den meisten Modellen zeitabhängig. Damit erhebt sich die Frage, was überhaupt der Abstand  $D$  zwischen zwei kosmischen Objekten ist. Ein pragmatischer (operationaler) Zugang definiert den Abstand über ein Messverfahren. Hier zeigt sich jedoch, dass verschiedene Messverfahren in einem expandierenden Universum zu unterschiedlichen Ergebnissen führen können. Ein eher konzeptuelles Verfahren beruht auf der Annahme eines homogenen und isotropen aber nicht statischen Universums - unter diesen Umständen kann man Raumkoordinaten wählen, in Bezug auf die die Hubble-Konstante nicht ortsabhängig ist, und es gibt eine bevorzugte Zeitrichtung mit Zeitvariablen  $t$ . Dann ist es sinnvoll, von einem Abstand  $D(t)$  zwischen zwei Objekten zum Zeitpunkt  $t$  zu sprechen und die Geschwindigkeit  $v(t)$ , mit der sich diese Objekte voneinander entfernen, durch  $v(t) = dD(t)/dt$  zu definieren. In diesem Fall gilt für zwei weit entfernte Objekte ohne Pekuliarbewegung (mathematisch spricht man im Englischen auch von „comoving objects“ in Bezug auf dieses Koordinatensystem) die Beziehung:

$$H = \frac{\dot{a}(t)}{a(t)} = \frac{dD(t)}{dt} / D(t) \quad \text{bzw.} \quad v(t) = H(t)D(t). \quad (16)$$

Hierbei ist  $a(t)$  die Skala des Universums, also ein Maß für seine Ausdehnung, und diese ist direkt proportional zum Abstand  $D(t)$  von „comoving“ Objekten. Dies bezeichnet man ebenfalls als Hubble-Gesetz (obwohl Hubble es in dieser Form nicht verwendet hat). Die Identifizierung des sogenannten Hubble'schen Rotverschiebungsgesetzes (Gl. 13) und des Hubble'schen Geschwindigkeit-Abstands-Gesetzes (Gl. 16) führt oft zu Fehlvorstellungen. Die in obiger Gleichung ausgedrückte Beziehung zwischen der Geschwindigkeit und dem Abstand gilt exakt (sie ist praktisch die Definition der Hubble-Konstanten) in allen homogenen und isotropen kosmologischen Modellen. Die Beziehung 13 gilt nur für kleine  $z$ -Werte.

Wir interpretieren also die Rotverschiebung für sehr weit entfernte Objekte nicht als Doppler-Effekt, sondern als einen Effekt der allgemeinen Relativitätstheorie, der mit der Raumausdehnung zu tun hat. Die Objekte haben somit keine Geschwindigkeit, sondern der Raum zwischen den Objekten und uns nimmt zu. Daraus abgeleitete Größen wie Distanz oder Geschwindigkeit hängen vom gewählten Koordinatensystem ab.

Die Hubble-Konstante  $H$  hat die Dimension  $\text{s}^{-1}$ , also eine inverse Zeit. Oft gibt man sie jedoch in  $(\text{km/s})/\text{Mpc}$  an und ihr Wert liegt heute bei ungefähr  $H = 70 (\text{km/s})/\text{Mpc}$ . Dies kann man so interpretieren: Wenn der Abstand einer Galaxie um eine Megaparallaxensekunde zunimmt (das sind rund  $3,26 \cdot 10^6 \text{ ly}$ ), nimmt die formale Fluchtgeschwindigkeit dieses Objekts von der Erde um  $70 \text{ km/s}$  zu. Beispielsweise haben Objekte in einer Entfernung von 1 Milliarde Lichtjahren formal eine Fluchtgeschwindigkeit von knapp  $21.500 \text{ km/s}$ .

Hubble selbst glaubte nicht, dass sein Gesetz Indiz für eine Urknalltheorie sein könnte. Unter anderem führten die von ihm verwendeten falschen Entfernungswerte auf viel zu hohe Geschwindigkeiten für die Abstandszunahme zwischen den Galaxien und damit auf ein viel zu junges Universum (jünger als manche geologische Schätzungen für das Alter der Erde, wobei diese Überlegungen damals noch sehr umstritten waren).

Bis in die 90er Jahre des 20. Jahrhunderts war das Hubble'sche Rotverschiebungsgesetz

$$z = \frac{H}{c} D \quad (17)$$

nahezu die einzige Möglichkeit, die Entfernung  $D$  zu sehr weit entfernten Objekten, bei denen z.B. keine Cepheiden mehr beobachtet werden konnten, zu bestimmen. Hierbei wurde  $H$  mehr oder weniger als Konstante angenommen. Nachdem in den 90er Jahren eine systematische Neubestimmung der Entfernungswerte über Typ Ia Supernovae erfolgte, erkannte man Abweichungen von diesem Rotverschiebungsgesetz bzw. man konnte die Zeitabhängigkeit von  $H$  bestimmen. Dies führte zu der Entdeckung, dass sich das Universum seit rund 5 bis 8 Milliarden Jahren wieder beschleunigt ausdehnt und damit zur Entdeckung der dunklen Energie.

# Landkarten und der metrische Tensor

Thomas Filk, Universität Freiburg

Die fundamentale dynamische Größe in der Allgemeinen Relativitätstheorie ist die Metrik bzw. der metrische Tensor, oft geschrieben als  $g_{\mu\nu}(x)$ , wobei  $x$  die Punkte der Raumzeit parametrisiert. In der Mathematik, insbesondere der Differentialgeometrie, geht der Definition dieses Feldtensors eine von einer Einbettung unabhängige Definition einer Mannigfaltigkeit voran. Dies wird hier umgangen, auch wenn die Begriffe der Karte und des Atlas einer solchen Beschreibung schon sehr nahe kommen.

In diesem Kapitel wird der Begriff des metrischen Feldtensors anhand von Landkarten veranschaulicht. Landkarten sind Lösungen des Problems, Ausschnitte einer Kugeloberfläche in einer Ebene darzustellen. Das ist mit längentreuen Darstellungen, die größere Bereiche der Erdoberfläche wiedergeben, nicht möglich. Besonders deutlich wird dies an Weltkarten. Hier kommt es immer zu deutlichen Verzerrungen und die Kunst besteht darin, die größten Verzerrungen in solche Bereiche zu legen, die für die konkrete Anwendung weniger relevant sind. Außerdem ist man man je nach Anwendung daran interessiert, unterschiedliche Dinge auf einer Karte invariant zu lassen. Es gibt winkeltreue bzw. formtreue Darstellungen der Kugeloberfläche, die aber die Flächen unterschiedlich skalieren, und es gibt flächentreue Darstellungen, die aber die Formen sehr verzerrn. Und dann gibt es natürlich sehr viele Optionen dazwischen.

Bei einer Weltkarte kommen neben den Verzerrungen noch sogenannte Koordinatensingularitäten hinzu. Dabei handelt es sich um singuläre Punkte oder Linien (typischerweise am Rand der Karte, z.B. an den Polen), bei denen die Bijektivität zwischen der Karte und dem dargestellten Gebiet verloren geht. Solche Koordinatensingularitäten treten immer dann auf, wenn die Topologie der darzustellenden Mannigfaltigkeit - hier der Kugeloberfläche - nicht mit der Topologie der darstellenden Mannigfaltigkeit - hier dem zusammenhängenden Ausschnitt einer euklidischen Ebene - übereinstimmt. Das Gebiet selbst hat natürlich keine Singularität - die Oberfläche einer Kugel ist glatt - sondern nur die Karte. In der Allgemeinen Relativitätstheorie treten solche Koordinatensingularitäten ebenfalls auf (z.B. ist der Horizont eines Schwarzen Lochs in der üblichen Darstellung der Schwarzschild-Koordinaten eine Koordinatensingularität). Allerdings gibt es auch Lösungen der Einstein'schen Gleichungen mit wirklichen (physikalischen) Singularitäten, z.B. das Zentrum eines Schwarzen Lochs, bei dem die Gezeitenkräfte unendlich werden. Es ist nicht immer leicht, eine Koordinatensingularität von einer physikalischen Singularität zu unterscheiden.

In den kommenden ersten Abschnitten wird zunächst der Begriff der Metrik anhand eines ortsabhängigen Kartenmaßstabs erläutert. Der Rest des Kapitels besteht im Wesentlichen aus Beispielen.

## 1 Ortsabhängige Landkartenmaßstäbe

Eine Landkarte enthält gewöhnlich eine Maßstabsangabe, beispielsweise eine Wanderkarte 1:25 000 oder eine Straßenkarte 1:250 000. Das bedeutet, eine Längeneinheit auf der Karte (z.B. ein Zentimeter) entspricht 25 000 bzw. 250 000 dieser Längeneinheiten in Wirklichkeit, das sind somit 250 m bzw. 2,5 km. Eine Maßstabsangabe setzt somit eine Distanz auf der Karte mit einer Distanz auf der von der Karte dargestellten Mannigfaltigkeit in Beziehung. Nichts anderes macht der metrische Tensor in der Differentialgeometrie, allerdings kommt hier eine Komplikation hinzu, die man auch schon bei Landkarten findet: die Orts- und Richtungsabhängigkeit des Maßstabs.

Der Maßstab einer Landkarte kann nicht überall derselbe sein, und wenn diese Landkarte große Gebiete darstellt oder sehr präzise sein soll (z.B. eine Seekarte für die Schifffahrt), sollte angegeben sein, wie die Korrektur zum allgemeinen Maßstab aussieht. Bei groben Weltkarten werden diese Korrekturen oft nicht angegeben, obwohl sie offensichtlich vorhanden sind. Abbildung 1 zeigt eine Weltkarte in einer sogenannten quadratischen Zylinderprojektion. Die  $x$ - und  $y$ -Achse entsprechen den Längen- und Breitengraden. Der Abstand zwischen zwei eingezeichneten Längengraden am Äquator beträgt rund 556 km (die Karte ist in  $5^\circ$ -Abschnitte unterteilt). Dieser Abstand verringert sich mit dem Breitengrad  $\theta$  (vom Äquator aus gerechnet) um einen Faktor  $\cos \theta$ , beträgt also am 60. Breitengrad (das entspricht der Höhe von Helsinki) nur noch die Hälfte.

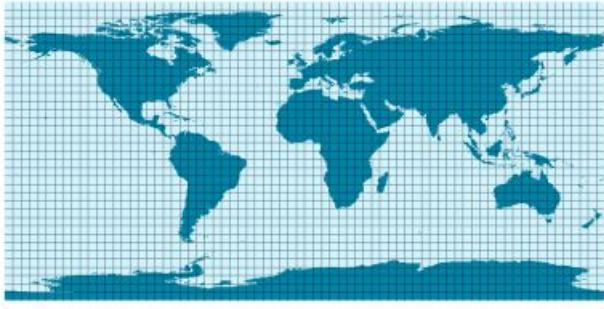


Abbildung 1: Eine Weltkarte in rechteckiger Form. Gleiche Breiten- und Längengrade entsprechen gleichen Abständen auf der Karte. Es ist offensichtlich, dass sich der Maßstab in Ost-West-Richtung in der Nähe der Pole ändert. Die Netzeinteilung ist in  $5^\circ$ -Schritten. Am Äquator entspricht einer Einheit rund 556 km, beim 60. Breitengrad ist es nur noch die Hälfte. (Quelle [1])

Der Maßstab einer solchen Karte hängt also sowohl vom Ort ab, an dem man die Beziehung zwischen Karte und Erdoberfläche bestimmen möchte, als auch von der Richtung. Bei der in Abb. 1 verwendeten Darstellung ändert sich der Maßstab nicht für Punkte auf demselben Längengrad, also in Nord-Süd-Richtung: Gleiche Breitengraddifferenzen entsprechen auch gleichen Abständen. Es ändern sich lediglich die Abstände zwischen den Längengraden als Funktion vom Breitengrad.

In dieser Darstellung wirken somit Länder am nördlichen Rand oder auch die Antarktis am südlichen Rand sehr in die Breite gestreckt. Andere Darstellungen versuchen diese Problematik zu beheben: Beispielsweise findet man bei der Mercator-Projektion (Abschnitt 5.2) dieselbe Streckung auch in Nord-Süd-Richtung, sodass die Proportionen in der Form wieder stimmen, allerdings wirken nun Gebiete in Polnähe wesentlich größer als in Äquatornähe, d.h. diese Karten skalieren die Flächen ungleich. Andere Projektionen, beispielsweise die Lambert-Projektion (Abschnitt 5.1), behalten die Flächeninhalte bei, sie stauchen aber zusätzlich die Abstände in Nord-Süd-Richtung, sodass die Gebiete noch flacher aussehen. Wir behandeln diese Darstellungen in den späteren Abschnitten. Hier soll als wesentlicher Punkt festgehalten werden, dass bei Landkarten die Maßstäbe (a) vom Ort abhängen können und (b) richtungsabhängig sein können. Außerdem muss die Streckung bzw. Stauchung nicht in Ost-West- oder Nord-Südrichtung erfolgen, sondern kann auch „schräg“ verlaufen. Der folgende Abschnitt beschreibt, wie man diese Art von Verzerrung lokal beschreiben kann und wie der Maßstab mit dem metrischen Tensor zusammenhängt.

## 2 Beschreibung einer Ellipse

In einem zweidimensionalen kartesischen Koordinatensystem lässt sich eine Ellipse algebraisch beispielsweise durch die folgenden zwei Formen darstellen:

$$\mathbf{x}(\varphi) = r(a \cos \varphi, b \sin \varphi), \quad \varphi \in [0, 2\pi] \quad \text{und} \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = r^2. \quad (1)$$

Die Äquivalenz dieser beiden Darstellungen erkennt man sofort, wenn man für  $x$  und  $y$  in der rechten Darstellung die Ausdrücke der linken Darstellung ( $x = ar \cos \varphi$  und  $y = br \sin \varphi$ ) einsetzt. Aus beiden

Darstellungen wird deutlich, dass es sich bei einer Ellipse um einen gestauchten oder gestreckten Kreis handelt: Multipliziert man die  $y$ -Komponente mit  $a/b$  erhält man die Gleichung eines Kreises. Sofern  $a > b$  gilt, sind  $ar$  und  $br$  die große und die kleine Halbachse der Ellipse. Der zusätzliche Parameter  $r$  hat folgenden Grund:  $x$  und  $y$  sollen die Dimension einer Länge haben (hier handelt es sich um Abstände auf einer Karte), ebenso soll  $r$  die Dimension einer Länge haben (hier handelt es sich um einen Abstand auf der Erdkugel); die beiden Parameter  $a$  und  $b$  sollen dimensionslose Skalierungsfaktoren sein. Beispielsweise wären bei einer normalen Karte mit dem Maßstab 1:25 000 diese Parameter  $a = b = 1/25\,000$ .

Die Halbachsen sind entlang der Koordinatenachsen ausgerichtet. Für eine allgemeine Darstellung einer Ellipse kann man die Koordinaten noch um den Winkel  $\alpha$  drehen, d.h.

$$x \mapsto x \cos \alpha - y \sin \alpha \quad \text{und} \quad y \mapsto x \sin \alpha + y \cos \alpha \quad (2)$$

und erhält die Form:

$$\frac{x^2 \cos^2 \alpha + y^2 \sin^2 \alpha - 2xy \cos \alpha \sin \alpha}{a^2} + \frac{x^2 \sin^2 \alpha + y^2 \cos^2 \alpha + 2xy \cos \alpha \sin \alpha}{b^2} = r^2 \quad (3)$$

oder

$$g_{11}x^2 + (g_{12} + g_{21})xy + g_{22}y^2 = r^2 \quad (4)$$

mit

$$g_{11} = \frac{\cos^2 \alpha}{a^2} + \frac{\sin^2 \alpha}{b^2} \quad g_{22} = \frac{\sin^2 \alpha}{a^2} + \frac{\cos^2 \alpha}{b^2} \quad g_{12} = g_{21} = \left( \frac{1}{b^2} - \frac{1}{a^2} \right) \cos \alpha \sin \alpha. \quad (5)$$

Aus dieser Darstellung finden wir:

$$g_{11} + g_{22} = \frac{1}{a^2} + \frac{1}{b^2} \quad \text{und} \quad g_{11}g_{22} - g_{12}^2 = \frac{1}{a^2 b^2}. \quad (6)$$

Wir erkennen somit: Gleichung 4 lässt sich in der Form

$$\mathbf{x}^T \cdot g\mathbf{x} = (x, y) \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = r^2 \quad (7)$$

schreiben, wobei Gl. 6 zum Ausdruck bringt, dass die Matrix  $g$  die Eigenwerte  $\lambda_1 = 1/a^2$  und  $\lambda_2 = 1/b^2$  hat (in Gl. 6 steht links die Spur - also die Summe der Eigenwerte - und rechts die Determinante von  $g$  - also das Produkt der Eigenwerte). Die Bilinearform in Gl. 3 bzw. 4 wird durch eine Drehung um  $-\alpha$  diagonalisiert und auf die Form in Gl. 1 gebracht - da kam sie schließlich einmal her.

### 3 Tissot-Indikatrix und die Metrik

Wie wir in Abschnitt 1 gesehen haben, benötigen die meisten Landkarten (insbesondere alle Landkarten, die ein großes Gebiet der Erdkugel darstellen) einen ortsabhängigen Maßstab. Dieser Maßstab bringt zum Ausdruck, wie ein Kreis auf der Erdkugel in der Karte wiedergegeben wird, wobei diese Darstellung in führender Ordnung (d.h., wenn dieser Kreis klein ist im Vergleich zu der Skala, auf der sich der Maßstab verändert) den Kreis zu einer Ellipse verformt. Einen ortsabhängigen Maßstab können wir also dadurch kennzeichnen, dass wir an jedem Punkt der Karte die Parameter  $g_{ij}$  angeben, die nach Gl. 4 eine Ellipse charakterisieren. Genau dies ist aber der metrische Feldtensor. Das soll in diesem Abschnitt erläutert werden.

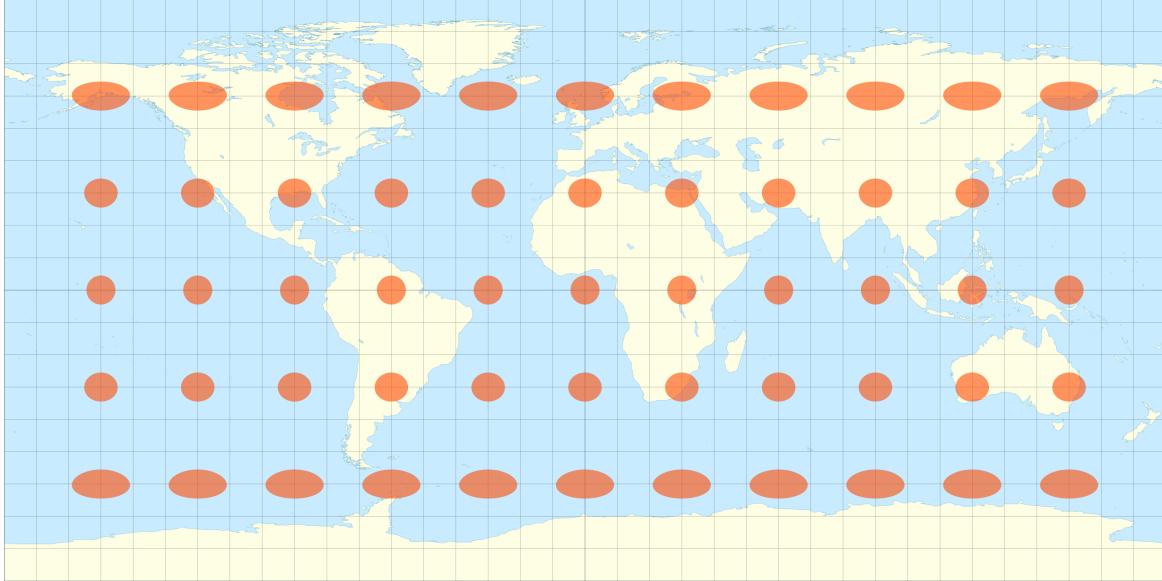


Abbildung 2: Tissot'sche Indikatrix zur quadratischen Zylinderprojektion. Die Ellipsen entsprechen den Darstellungen von Kreisen auf der Kugeloberfläche in der Karte (Quelle [2]).

Abbildung 2 zeigt nochmals eine quadratische Zylinderprojektion, bei der die Abschnitte entlang der Längengrade (also in Nord-Süd-Richtung) einen konstanten Abstand haben, d.h., hier werden die Längen- und Breitengrade in einem kartesischen Koordinatensystem dargestellt. Wie schon besprochen, führt diese Darstellung zu einer Verzerrung in Ost-West-Richtung in Abhängigkeit vom Breitengrad: Je näher der Breitengrad an den Polen ist, umso größer ist die Dehnung im Vergleich zum Äquator, wo der Maßstab in alle Richtungen derselbe ist. Ein hypothetischer Kreis auf der Erde (in diesem Fall ein Kreis mit einem Radius von rund  $r = 500$  km) wird in der Karte durch eine Ellipse wiedergegeben. Diese Ellipse entartet am Äquator zu einem Kreis (dort sind die Maßstäbe in alle Richtungen gleich) und wird in Ost-West-Richtung gestreckt, je weiter man sich vom Äquator entfernt. Auf dem 60. Breitengrad ist die große Halbachse bereits doppelt so groß wie die kleine Halbachse. Beide Halbachsen entsprechen jedoch immer noch auf der Erdkugel einer Strecke von 500 km. Man bezeichnet diese Darstellung des ortsbildenden Kartenmaßstabs auch als Tissot'sche Indikatrix. Wir nennen die Kartenabbildung von Kreisen auf der Kugeloberfläche dann Tissot-Ellipsen. Die Charakterisierung einer Tissot-Ellipse an einem bestimmten Ort auf der Karte durch die Parameter der Ellipse in Form der Matrix  $g_{ij}$  bezeichnet man als metrischen Feldtensor. „metrisch“ bedeutet, dass dieses Objekt die Maßstäbe zur Bestimmung von Entfernungen kodiert, „Feld-“ bedeutet, dass dieses Objekt an jedem Ort definiert und von Ort zu Ort verschieden sein kann, und „Tensor“ bedeutet, dass es sich um eine Matrix handelt, die die Richtungsabhängigkeit angibt.

Zur Bestimmung der Parameter  $a$  und  $b$  können wir folgendermaßen vorgehen (ein allgemeines Verfahren, wie man aus einer Parameterdarstellung der Kugeloberfläche diese Parameter gewinnt, beschreiben wir in Abschnitt 4). Die Karte habe eine Breite  $B$  (hier ungefähr  $B = 150$  mm) und eine Höhe  $H = B/2$  (da die Ost-West-Richtung in 360 Grade, die Nord-Süd-Richtung aber nur in 180 Grade unterteilt ist, und beide Richtungen dieselbe Skala haben sollen). Der Erdumfang am Äquator beträgt  $U = 40\,000$  km (wir gehen hier von einer idealen Kugel mit dieser Länge eines Großkreises aus). Allgemein: Wenn eine Strecke auf einer Kugel (hier der Umfang) die Länge  $U$  hat und auf einer Karte im Abstand  $B$  (hier der Breite der Karte) dargestellt wird, dann hat die Karte einen Maßstab von  $1 : U/B$ .

Entfernt man sich nun vom Äquator in Richtung der Pole ändert sich dieser Maßstab in Ost-West-Richtung (d.h. für Punkte auf demselben Breitengrad) und der Abstand zwischen zwei Längengraden bei einem Breitengrad  $\theta$  wird um den Faktor  $\cos \theta$  kürzer. Wir erhalten also für den Maßstab in Ost-West-Richtung am Breitengrad  $\theta$  den Wert  $(U/B) \cos \theta$ . Damit folgt:

$$a(\theta, \varphi) = \frac{B}{U \cos \theta} \approx \frac{1}{266\,666\,667 \cos \theta} \quad \text{und} \quad b(\theta, \varphi) = \frac{B}{U} \approx \frac{1}{266\,666\,667}. \quad (8)$$

Diese Parameter hängen nicht von  $\varphi$  (dem Längengrad) ab sondern nur vom Breitengrad. Der Maßstab am Äquator für obige Karte (mit der Breite 150 mm) wäre somit 1 : 266 666 667. Der metrische Feldtensor wäre

$$g = \begin{pmatrix} \left(\frac{U}{B}\right)^2 \cos^2 \theta & 0 \\ 0 & \left(\frac{U}{B}\right)^2 \end{pmatrix}. \quad (9)$$

Man erkennt, dass  $g_{11}$  bei  $\theta = \pm 90^\circ$  (also an den Polen) verschwindet. Hierbei handelt es sich um eine typische Koordinatensingularität, bei der ein einzelner Punkt (der Nord- bzw. der Südpol) auf eine ganze Achse (den oberen bzw. unteren Rand der Karte) abgebildet wird. Der Nord- bzw. Südpol sind auf der Kugel natürlich vollkommen reguläre Punkte.

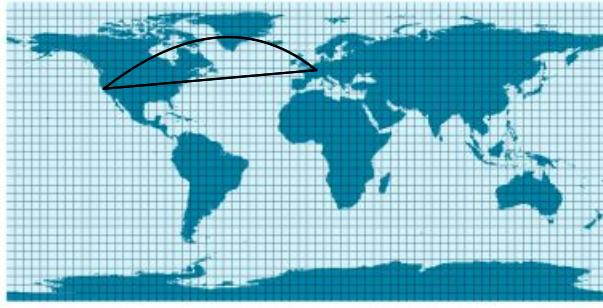


Abbildung 3: Die kürzeste Verbindung auf der Kugeloberfläche zwischen zwei Punkten (z.B. Frankfurt und San Francisco) erscheint auf einer Weltkarte als gekrümmte Linie. Allerdings bedarf es weniger Tissot-Ellipsen, um diese gekrümmte Linie zu überdecken als bei der geraden Verbindungsstrecke. (Abbildungsquelle [1])

## 4 Parameterdarstellungen

Unter einer Parameterdarstellung einer (offenen Teilmenge einer) 2-dimensionalen Mannigfaltigkeit (z.B. einer Kugeloberfläche), eingebettet in den 3-dimensionalen Raum, versteht man eine Abbildung der Form  $(u, v) \mapsto \mathbf{x}(u, v)$ , wobei diese Abbildung lokal bijektiv sein soll. Wir betrachten zunächst nochmals das Beispiel der Kugeloberfläche, die durch die Längen- und Breitengrade parametrisiert wird.

Eine Kugel lässt sich in Kugelkoordinaten durch die beiden Winkel  $\theta$  und  $\varphi$  beschreiben:

$$\mathbf{x}(\theta, \varphi) = (R \cos \varphi \cos \theta, R \sin \varphi \cos \theta, R \sin \theta). \quad (10)$$

Im Gegensatz zur üblichen Wahl von Kugelkoordinaten wurde hier die Parametrisierung so gewählt, dass der Winkel  $\theta = 0$  dem Äquator entspricht und  $\theta = \pm \frac{\pi}{2}$  dem Nord- bzw. Südpol.  $\varphi$  kann die Werte  $-\pi$  bis  $+\pi$  annehmen, wobei  $\varphi = 0$  den 0-Meridian bezeichnet. Damit entsprechen  $\theta$  und  $\varphi$  dem Breiten- bzw. Längengrad.  $R$  ist der Radius der Kugel, bei der Erde ist somit  $R = 40\,000$  km.

Eine sehr kleine (infinitesimale) Verschiebung  $\Delta \mathbf{x}$  auf der Erdkugel, bedeutet für die Längen- und Breitengrade:

$$\Delta \mathbf{x} = \frac{\partial \mathbf{x}}{\partial \theta} \Delta \theta + \frac{\partial \mathbf{x}}{\partial \varphi} \Delta \varphi, \quad (11)$$

bzw.

$$(\Delta \mathbf{x})^2 = \left( \frac{\partial \mathbf{x}}{\partial \theta} \cdot \frac{\partial \mathbf{x}}{\partial \theta} \right) (\Delta \theta)^2 + 2 \left( \frac{\partial \mathbf{x}}{\partial \varphi} \cdot \frac{\partial \mathbf{x}}{\partial \theta} \right) \Delta \theta \Delta \varphi + \left( \frac{\partial \mathbf{x}}{\partial \varphi} \cdot \frac{\partial \mathbf{x}}{\partial \varphi} \right) (\Delta \varphi)^2. \quad (12)$$

Durch Vergleich mit Gl. 4 erkennen wir die folgenden Beziehungen:

$$g_{\theta\theta} = \left( \frac{\partial \mathbf{x}}{\partial \theta} \cdot \frac{\partial \mathbf{x}}{\partial \theta} \right) \quad g_{\theta\varphi} = g_{\varphi\theta} = \left( \frac{\partial \mathbf{x}}{\partial \varphi} \cdot \frac{\partial \mathbf{x}}{\partial \theta} \right) \quad g_{\varphi\varphi} = \left( \frac{\partial \mathbf{x}}{\partial \varphi} \cdot \frac{\partial \mathbf{x}}{\partial \varphi} \right). \quad (13)$$

Berechnen wir die beiden Vektoren

$$\frac{\partial \mathbf{x}}{\partial \theta} = R(-\cos \varphi \sin \theta, -\sin \varphi \sin \theta, \cos \theta) \quad \text{und} \quad \frac{\partial \mathbf{x}}{\partial \varphi} = R(-\sin \varphi \cos \theta, \cos \varphi \cos \theta, 0) \quad (14)$$

so folgt:

$$g_{\theta\theta} = R^2 \quad g_{\theta\varphi} = g_{\varphi\theta} = 0 \quad g_{\varphi\varphi} = R^2 \cos^2 \theta. \quad (15)$$

Damit haben wir bezüglich der Parametrisierung durch die Längen- und Breitengrade den metrischen Feldtensor gefunden. Allerdings haben wir in Abschnitt 3 die Parametrisierung durch einen Maßstab auf unserer Karte angeben, d.h. die  $x$ - und  $y$ -Koordinaten auf der Karte. Wir müssen also noch die Beziehungen zwischen den Längen- und Breitengraden und den  $x$ - und  $y$ -Koordinaten der Karte finden. Die Breite  $B$  der Karte entspricht einem Vollkreis von  $360^\circ$  oder  $2\pi$ , daher folgt

$$\varphi = \frac{2\pi}{B}x \quad \text{und} \quad \theta = \frac{2\pi}{B}y \quad \text{bzw.} \quad \Delta \varphi = \frac{2\pi}{B}\Delta x \quad \text{und} \quad \Delta \theta = \frac{2\pi}{B}\Delta y. \quad (16)$$

Setzen wir diese Beziehungen in Gl. 12 ein und nutzen aus, dass  $U = 2\pi R$ , folgen die Beziehungen aus Abschnitt 3.

Parametrisieren wir die Kugeloberfläche (oder ganz allgemein eine 2-dimensionale Mannigfaltigkeit im 3-D-Raum) in der Form  $(u, v)$ , dann folgt

$$\Delta \mathbf{x} = \frac{\partial \mathbf{x}}{\partial u} \Delta u + \frac{\partial \mathbf{x}}{\partial v} \Delta v, \quad (17)$$

bzw.

$$(\Delta \mathbf{x})^2 = \left( \frac{\partial \mathbf{x}}{\partial u} \cdot \frac{\partial \mathbf{x}}{\partial u} \right) (\Delta u)^2 + 2 \left( \frac{\partial \mathbf{x}}{\partial u} \cdot \frac{\partial \mathbf{x}}{\partial v} \right) \Delta u \Delta v + \left( \frac{\partial \mathbf{x}}{\partial v} \cdot \frac{\partial \mathbf{x}}{\partial v} \right) (\Delta v)^2 \quad (18)$$

und somit:

$$g_{uu} = \left( \frac{\partial \mathbf{x}}{\partial u} \cdot \frac{\partial \mathbf{x}}{\partial u} \right) \quad g_{uv} = g_{vu} = \left( \frac{\partial \mathbf{x}}{\partial u} \cdot \frac{\partial \mathbf{x}}{\partial v} \right) \quad g_{vv} = \left( \frac{\partial \mathbf{x}}{\partial v} \cdot \frac{\partial \mathbf{x}}{\partial v} \right). \quad (19)$$

Die beiden Vektoren  $\frac{\partial \mathbf{x}}{\partial u}$  und  $\frac{\partial \mathbf{x}}{\partial v}$  spannen eine Fläche auf. Für das Quadrat dieser Fläche gilt:

$$\left( \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \right)^2 = \sum_m \left( \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \right)_m \left( \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \right)_m \quad (20)$$

$$= \sum_{m,i,j,k,l} \epsilon_{mij} \frac{\partial x_i}{\partial u} \frac{\partial x_j}{\partial v} \epsilon_{mkl} \frac{\partial x_k}{\partial u} \frac{\partial x_l}{\partial v} \quad (21)$$

$$= \sum_{i,j,k,l} (\delta_{ik} \delta_{jl} - \delta_{il} \delta_{jk}) \frac{\partial x_i}{\partial u} \frac{\partial x_j}{\partial v} \frac{\partial x_k}{\partial u} \frac{\partial x_l}{\partial v} \quad (22)$$

$$= \sum_{ij} \left( \frac{\partial x_i}{\partial u} \frac{\partial x_j}{\partial v} \frac{\partial x_i}{\partial u} \frac{\partial x_j}{\partial v} - \frac{\partial x_i}{\partial u} \frac{\partial x_j}{\partial v} \frac{\partial x_j}{\partial u} \frac{\partial x_i}{\partial v} \right) \quad (23)$$

$$= \left( \sum_i \frac{\partial x_i}{\partial u} \frac{\partial x_i}{\partial u} \right) \left( \sum_j \frac{\partial x_j}{\partial v} \frac{\partial x_j}{\partial v} \right) - \left( \sum_i \frac{\partial x_i}{\partial u} \frac{\partial x_i}{\partial v} \right) \left( \sum_j \frac{\partial x_j}{\partial v} \frac{\partial x_j}{\partial u} \right) \quad (24)$$

$$= g_{uu} g_{vv} - g_{uv} g_{vu} = \det g. \quad (25)$$

Das bedeutet,  $\sqrt{\det g}$  gibt den Faktor zwischen einer Fläche auf der Kugeloberfläche und einer Fläche auf der Karte an: Seien  $\Delta\mathbf{x}_u$  und  $\Delta\mathbf{x}_v$  zwei infinitesimale Verschiebungen auf der Kugeloberfläche, dann gilt:

$$\Delta\mathbf{x}_u = \frac{\partial\mathbf{x}}{\partial u} \Delta u \quad \text{und} \quad \Delta\mathbf{x}_v = \frac{\partial\mathbf{x}}{\partial v} \Delta v \quad (26)$$

und somit:

$$|\Delta\mathbf{x}_u \times \Delta\mathbf{x}_v| = \left| \frac{\partial\mathbf{x}}{\partial u} \times \frac{\partial\mathbf{x}}{\partial v} \right| \Delta u \Delta v = \sqrt{\det g} \Delta u \Delta v. \quad (27)$$

Wenn  $\sqrt{\det g}$  konstant ist (also nicht von dem Ort  $u, v$  auf der Karte abhängt), bezeichnet man die Karte als flächentreu. In diesem Fall haben gleiche infinitesimale Flächen  $\Delta u \Delta v$  auf der Karte auch gleiche Flächen auf der Kugeloberfläche (und umgekehrt). Eine Karte heißt winkeltreu oder formtreu bzw. konform, wenn

$$g = a(u, v) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (28)$$

In diesem Fall werden Kreise wieder auf Kreise abgebildet (vgl. Gl. 4), die allerdings um den möglicherweise ortsabhängigen Faktor  $a(u, v)$  skaliert sind.

## 5 Zylinder-Projektionen

Unter einer Zylinderprojektion versteht man eine Abbildung der Kugeloberfläche auf einen Zylinder, der am Äquator (gelegentlich auch an anderen Großkreisen) um die Kugel gelegt wird. Die  $x$ -Achse entspricht immer den Längengraden und auf der  $y$ -Achse sind die Breitengrade aufgetragen. Verschiedene Zylinderprojektionen unterscheiden sich im Wesentlichen in der Skala, die für die Breitengrade gewählt wird. Die quadratische Zylinderprojektion wählt die  $y$ -Achse direkt proportional zu den Breitengraden und zwar mit derselben Skala, wie die Längengrade. Daher hat eine solche Karte immer das Verhältnis Breite:Höhe=2:1.

Der Nachteil einer solchen Karte ist, dass die Umrisse von Flächen verzerrt werden - die Flächen erscheinen an den Polen in die Breite gezogen - und dass die Flächen zu den Polen hin größer erscheinen. Beide „Fehler“ lassen sich nicht gleichzeitig beheben. Es gibt aber zwei bekannte Zylinderprojektionen, bei denen die Fehler einzeln behoben werden: Die Mercator-Projektion ist „formtreu“ oder auch konform, d.h., die Form der Flächen bleibt erhalten, allerdings werden die Flächen zu den Polen hin immer größer; die Lambert-Projektion ist „flächentreu“, d.h., der Flächeninhalt bleibt erhalten, allerdings werden die Formen der Flächen zu den Polen hin verzerrt.

### 5.1 Die Lambert-Projektion

Bei der Lambert-Projektion (genauer sollte man von der zylindrischen Lambert-Projektion sprechen, da Lambert dieses Darstellungsverfahren auch auf Kegelmäntel erweitert hat) handelt es sich um eine rechteckige Zylinderprojektion, bei der ein Punkt der Kugeloberfläche senkrecht, ausgehend von der Erdachse, auf die Zylinderfläche projiziert wird. Seine Höhe auf der Zylinderfläche entspricht also seiner  $z$ -Koordinate in Kugelkoordinaten. Die Gleichungen 16 werden nun abgewandelt zu:

$$\varphi = \frac{2\pi}{B}x \quad \text{und} \quad \sin \theta = \frac{2\pi}{B}y \quad \text{bzw.} \quad \Delta\varphi = \frac{2\pi}{B}\Delta x \quad \text{und} \quad \cos \theta \Delta\theta = \frac{2\pi}{B}\Delta y. \quad (29)$$

Damit folgt nun:

$$g_{yy} = \frac{U^2}{B^2} \frac{1}{\cos^2 \theta} \quad g_{xy} = g_{yx} = 0 \quad g_{xx} = \frac{U^2}{B^2} \cos^2 \theta. \quad (30)$$

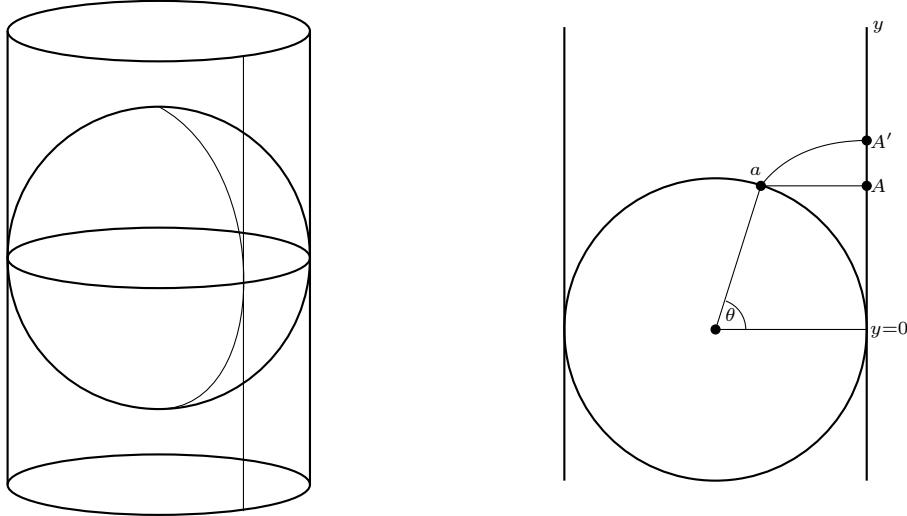


Abbildung 4: Zylinderprojektionen. (links) Bei einer Zylinderprojektion wird eine Kugeloberfläche auf einen Zylinder projiziert, der um einen Großkreis (meist den Äquator) der Kugel gelegt wird, sodass Längengrade in äquidistante senkrechte Geraden und Breitengrade in Graden parallel zu der Projektion des Großkreises abgebildet werden. (rechts) Die einzige Freiheit besteht in den Abständen der Breitengrade, d.h. in der Beziehung zwischen  $\theta$  und der  $y$ -Achse. Die Lambert-Projektion ( $a \mapsto A$ ) projiziert Punkte senkrecht auf die Zylinderfläche, d.h. sie behalten ihre Höhe. Bei der quadratischen Zylinderprojektion ( $a \mapsto A'$ ) wird der Längengrad „abgerollt“.

Wir erkennen, dass die Wurzel aus der Determinante von  $g$ , die in zwei Dimensionen die Änderung in der Skala für infinitesimale Flächen angibt (vgl. Gl.27), konstant ist (Faktor  $U^2/B^2$ ).<sup>1</sup> Das bedeutet, eine infinitesimale Fläche  $df$  auf der Erdkugel ist um den konstanten Faktor  $R^2/B^2$  größer, als die entsprechende Fläche auf der Karte. Infinitesimal gleiche Flächen auf der Erdkugel werden somit durch gleiche Flächen auf der Karte wiedergegeben. In diesem Sinne sagt man, die Lambert-Projektion ist flächenerhaltend oder flächentreu.

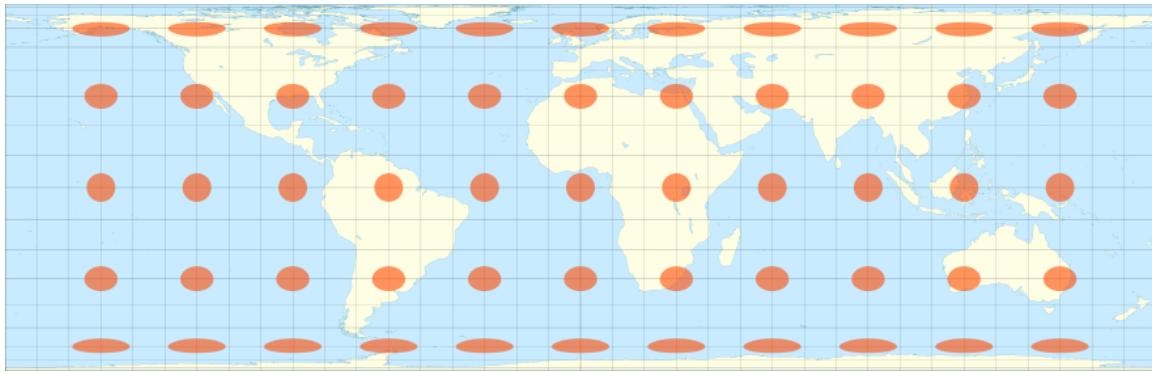


Abbildung 5: Tissot-Darstellung der zylindrischen Lambert-Projektion. (Abbildungsquelle [3])

Abbildung 5 zeigt eine zylindrische Lambert-Projektion der Erdkugel mit Tissot-Ellipsen. Im Vergleich zu Abb. 2 fällt auf, dass die Ellipsen in Polnähe nun flacher sind. Die Ellipsen nehmen in der

<sup>1</sup>Die Beziehung zwischen der Wurzel der Determinante und dem Skalenfaktor eines infinitesimalen (Hyper-)Volumens gilt in allen Dimensionen. Daher findet man bei invarianten Volumenintegralen in  $d$  Dimensionen auch immer  $\sqrt{\det g} d^d x$  als Integrationsmaß.

Höhe um denselben Faktor ab, um den sie in der Breite zunehmen. Dadurch bleibt der Flächeninhalt der Ellipsen überall derselbe. Allerdings werden die Gebiete in Polnähe auch stärker in der Höhe zusammengepresst und im Vergleich zu einer formgetreuen Darstellung verzerrt. Nun werden beide diagonalen Komponenten im metrischen Tensor an den Polen singulär.

## 5.2 Die Mercator-Projektion

Obwohl man die Mercator-Projektion als rechteckige Zylinderprojektion bezeichnet, handelt es sich im strengen Sinne nicht um eine Projektion, da die Abbildung eines Punkts auf der Kugeloberfläche auf einen Punkt auf der Zylinderoberfläche keine geometrische Konstruktion ist. Trotzdem besteht auch hier die einzige Veränderung zur quadratischen Zylinderprojektion in der Beziehung zwischen der  $y$ -Achse und dem Breitengrad.

Die Mercator-Projektion ist lokal winkel- und formtreu. Diese beiden Begriffe sind insofern äquivalent, als aus lokaler Winkeltreue die lokale Formtreue folgt und umgekehrt: Wenn zwei Dreiecke dieselben Winkel haben, haben sie auch die gleiche Form bzw. sind sich ähnlich, d.h., die Verhältnisse von je zwei Seitenlängen sind gleich. Die Mercator-Projektion ist nach Gerhard Mercator (1512-1597) benannt, der diese Projektionen um 1570 zum ersten Mal für Weltkarten verwendete. Die lokale Winkeltreue der Karte war früher in der Seefahrt von Vorteil, da ein bestimmter Kurs nach dem Kompass (der den Winkel zu einem Längengrad angibt) einer geraden Linie entspricht. Über große Abstände ist das aber nicht die kürzeste Verbindung zwischen zwei Punkten. Großkreise (die die kürzeste Verbindung darstellen) werden auf Mercator-Karten nicht als Geraden dargestellt.



Abbildung 6: Eine Weltkarte in Mercator-Darstellung. In Äquator Nähe gleicht diese Karte der quadratischen Zylinderprojektion (Abb. 1). Allerdings wird der Maßstab in Polnähe nicht nur in die Breite sondern auch in die Höhe gestreckt. Dadurch erscheinen die Umrisse von kleineren Ländern zwar ähnlich wie auf einer lokalen Projektion, also entsprechend ihrer lokalen Form, doch wirken die Länder im Vergleich zu Gebieten am Äquator übertrieben groß. Afrika ist in Wirklichkeit über fünfzehnmal größer als Grönland, Australien ist viermal größer. Afrika ist mehr als doppelt so groß wie das Landgebiet der Antarktis. (Quelle [4])

Damit eine Karte formtreu ist, müssen lokal die Abstände in  $x$ -Richtung und in  $y$ -Richtung um denselben Maßstab verändert werden, d.h., die beiden diagonalen Komponenten in der Metrik sind gleich (aber ortsabhängig). Da wir bei der Parametrisierung nur die Beziehung zwischen dem Breitengrad  $\theta$  und der Höhe  $y$  verändern können, ist eine Beziehung gesucht, sodass

$$g = \frac{U^2}{B^2} \cos^2 \theta \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{bzw.} \quad \Delta\theta = \frac{2\pi}{B} \cos \theta \Delta y. \quad (31)$$

Zur Bestimmung von  $y(\theta)$ , der  $y$ -Koordinate auf der Karte als Funktion des Breitengrads, haben wir somit das Integral

$$y(\theta) = \frac{B}{2\pi} \int_0^\theta \frac{1}{\cos \theta'} d\theta' \quad (32)$$

zu lösen. Die Lösung lautet:

$$y(\theta) = \frac{B}{2\pi} \ln \sqrt{\frac{1 - \sin \theta}{1 + \sin \theta}}. \quad (33)$$

Zur Lösung des Integrals: Man erweitere den Integranden im Zähler und Nenner um  $\cos \theta'$ , ersetze im Zähler  $\cos \theta' d\theta' = d \sin \theta'$  und im Nenner  $\cos^2 \theta' = 1 - \sin^2 \theta'$ . Mit einer Partialbruchzerlegung,  $\frac{1}{(1-\sin^2 \theta')} = \frac{1}{2} \left( \frac{1}{1-\sin \theta'} + \frac{1}{1+\sin \theta'} \right)$  kann man das Integral leicht lösen.

Für kleine Werte von  $\theta$ , also in Polnähe, verhält sich obige Beziehung wie in Gl. 16, aber für  $\theta \rightarrow \frac{\pi}{2}$  divergiert dieser Ausdruck, d.h., der Nord- bzw. Südpol sind auf einer Mercator-Karte im Unendlichen. Daher hören die meisten Mercator-Karten auch etwas oberhalb des 80. Breitengrads auf (Abb. 6 endet ungefähr beim 83. Breitengrad)

## 6 Das UNO-Emblem



Abbildung 7: Das Logo der Vereinten Nationen. In diesem Fall handelt es sich um eine sogenannte Azimutalprojektion. Es wird eine Ebene an einen Punkt der Kugel gelegt (in diesem Fall den Nordpol) und die Erdkugel wird in Polarkoordinaten um diesen Punkt herum dargestellt. Im vorliegenden Fall sind die Breitengrade von 90°-Nord bis rund 50°-Süd wiedergegeben. Die Einteilung der Breitengrade ist in 30°-Schritten und die Breitengrade sind äquidistant dargestellt. (Quelle [5])

Die Flagge der Vereinten Nationen (Abb. 7) verwendet eine Darstellung der Erdkontinente in Polarkoordinaten - eine sogenannte Azimutalprojektion. In diesem Fall wird eine Ebene tangential an einen Punkt der Kugel gelegt (sehr häufig, wie auch bei dem UN-Logo, an den Nordpol) und die Kugeloberfläche wird in Polarkoordinaten auf diese Fläche projiziert. Diese Darstellung (Nordpol als zentraler Punkt) wird auch hier gewählt. Der Polarwinkel  $\varphi$  entspricht dem Längengrad (allerdings wird im UN-Logo der Längengrad 0 nach unten projiziert). Die Breitengrade sind dann konzentrische Kreise um den Nordpol. Der Abstand zwischen Breitengräden ist der einzige Freiheitsgrad, der hier gewählt werden kann. Im UN-Logo sind die Breitengrade äquidistant angeordnet. Es gibt auch flächentreue Darstellungen, bei denen der Abstand zwischen Breitengräden von Nord nach Süd abnimmt. Theoretisch gibt es auch eine konforme Abbildung, bei der die Flächenform erhalten bleibt, diese würde sich aber nach Unendlich erstrecken und die Länder südlich des Äquators wären übertrieben groß dargestellt. Azimutale Projektionen sind nur an einem Punkt der Kugeloberfläche singulär, in diesem Fall am Südpol.

Polarkoordinaten sind 2-dimensionale Koordinaten in der Ebene, gegeben durch

$$\mathbf{x}(r, \varphi) = (r \cos \varphi, r \sin \varphi). \quad (34)$$

(Wir wählen hier die übliche Konvention, bei der  $\varphi = 0$  dem Punkt  $(x, y) = (1, 0)$  entspricht. Die UN-Darstellung erhält man aus der Koordinatenwahl  $(x, y) = (r \sin \varphi, -r \cos \varphi)$ .) Der Winkel  $\varphi$  entspricht direkt dem Längengrad. Der Breitengrad  $\theta$  auf der Kugeloberfläche entspricht hier dem Radius, d.h. je nach Wahl der Darstellung ist  $r = r(\theta)$  eine andere Funktion.

Für Polarkoordinaten gilt die Beziehung:

$$(\Delta s)^2 = (\Delta r)^2 + r^2(\Delta\varphi)^2. \quad (35)$$

Damit ist

$$g = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix} \quad \text{bzw.} \quad (\Delta s)^2 = g_{rr}(\Delta r)^2 + (g_{r\varphi} + g_{\varphi r})\Delta r\Delta\varphi + g_{\varphi\varphi}(\Delta\varphi)^2. \quad (36)$$

Wir erhalten diese Metrik wieder aus der Forderung

$$g_{uu} = \frac{\partial \mathbf{x}(u,v)}{\partial u} \cdot \frac{\partial \mathbf{x}(u,v)}{\partial u} \quad g_{uv} = g_{vu} = \frac{\partial \mathbf{x}(u,v)}{\partial u} \cdot \frac{\partial \mathbf{x}(u,v)}{\partial v} \quad g_{vv} = \frac{\partial \mathbf{x}(u,v)}{\partial v} \cdot \frac{\partial \mathbf{x}(u,v)}{\partial v} \quad (37)$$

mit den Tangentialvektoren:

$$\frac{\partial \mathbf{x}(r,\varphi)}{\partial r} = (\cos\varphi, \sin\varphi) \quad \frac{\partial \mathbf{x}(r,\varphi)}{\partial \varphi} = (-r\sin\varphi, r\cos\varphi). \quad (38)$$

Wir müssen nun noch die Beziehung zwischen  $r$  auf unserer Karte (in Polarkoordinaten) und den Breitengraden  $\theta$  auf der Kugeloberfläche herstellen. Die Winkel  $\varphi$  sind in beiden Fällen gleich. Wenn wir den Durchmesser der Karte mit  $D$  bezeichnen, entspricht  $D$  einem Vollkreis, sodass  $r = D/(2\pi)\theta$ . Umgekehrt entspricht auf der Erdkugel der Differenz von Breitengraden  $\Delta\theta$  eine Strecke von  $\Delta l = R\Delta\theta = (U/2\pi)\Delta\theta$ . Insgesamt erhalten wir somit:

$$\Delta r = \frac{D}{2\pi}\Delta\theta = \frac{D}{2\pi}\frac{2\pi}{U}\Delta l = \frac{D}{U}\Delta l, \quad (39)$$

und misst man die Länge  $l$  vom Nordpol aus zu einem Punkt auf der Erdkugel, gilt auch  $r = \frac{D}{U}l$ .

## 7 Kuriositäten

### 7.1 Tissot-Figuren in höheren Dimensionen

In drei Dimensionen wird eine Tissot-Figur zu einem Ellipsoid. Ein Ellipsoid ist gekennzeichnet durch die drei Hauptachsen sowie die Lage im Raum (nochmals drei Winkel, d.h. sechs Parameter). Dies lässt sich durch eine symmetrische  $3 \times 3$ -Matrix  $g_{ij}$  charakterisieren, sodass die Ellipsoid-Gleichung die Form

$$(\Delta s)^2 = \sum_{i,j=1}^3 g_{ij}\Delta x_i\Delta x_j \quad (40)$$

nimmt. Diese Gleichung bleibt unverändert auch in höheren Dimensionen, lediglich die Indizes durchlaufen eine größere Indexmenge. Da  $g$  symmetrisch (und damit selbst-adjungiert) ist, kann man es durch eine Rotation diagonalisieren. Die Eigenwerte sind  $\lambda_i = 1/a_i^2$ , wobei  $a_i$  die Hauptachsen des verallgemeinerten Ellipsoids sind, und die Rotation charakterisiert die Lage dieser Hauptachsen im Raum.

### 7.2 Tissot-Hyperbeln in Minkowski-Räumen

In  $(1+1)$ -Raumzeit-Dimensionen handelt es sich bei den Tissot-Figuren um Hyperbeln und die Vorgabe der Lichtkegelstruktur. Die Kreisgleichung wird ersetzt durch

$$(\Delta t)^2 - (\Delta x)^2 = \text{const.} \quad (41)$$

Ist die Konstante positiv, spricht man von zeitartigen Ereignissen (die sich in ihrer Lage um  $\Delta t$  und  $\Delta x$  unterscheiden). Ist sie negativ nennt man die Ereignisse raumartig, und ist sie null, sind die Ereignisse lichtartig. In diesen Koordinaten werden die Lichtkegel durch Diagonalen dargestellt und die Skala ist in zeitartige und raumartige Richtungen dieselbe. In einer allgemeinen Karte können die Lichtkegel lokal gedreht sein und die Skalen auch unterschiedlich. In höher dimensional Räumen werden die Lichtkegel zu verallgemeinerten Kegelmantelflächen, die zeitartigen „Hyperbeln konstanter Eigenzeiten“ werden zu „Hyperbelschalen konstanter Eigenzeiten“, die raumartigen Hyperbeln konstanter Abstände werden zu Rotationskörpern, die durch Drehung um die Zeitachse entstehen.

### 7.3 Die Lambert-Karte und ein Theorem von Archimedes

Eines der drei bekannten mathematischen Probleme der Antike war die geometrische Konstruktion - nur mit Zirkel und Lineal - eines Quadrats mit derselben Fläche wie ein vorgegebener Kreis bzw. letztendlich die Konstruktion der Zahl  $\pi$  aus einer Einheitslänge. Der Beweis, dass dies nicht möglich ist, erfolgte erst 1882 durch Ferdinand Lindemann. Genauer hat Lindemann bewiesen, dass  $\pi$  transzendent ist (also keine Lösung einer algebraischen Gleichung mit rationalen Koeffizienten ist); dass sich transzendente Zahlen nicht geometrisch mit Zirkel und Lineal konstruieren lassen, war schon vorher bekannt.

Archimedes konnte jedoch sehr viele Theoreme beweisen, bei denen krummlinige Flächen mit Quadraten oder Rechtecken in Beziehung gebracht wurden. Eines dieser Theoreme besagt, dass die Oberfläche einer Kugel genauso groß ist wie die Mantelfläche eines Zylinders, der am Äquator um die Kugel gewickelt ist und dieselbe Höhe wie die Kugel hat. Heute würden wir das folgendermaßen beweisen: Die Oberfläche einer Kugel vom Radius  $R$  ist  $4\pi R^2$ , ein um die Kugel gewickelter Zylinder hat die Grundseite  $U = 2\pi R$  (der Umfang der Kugel) und die Höhe  $2R$  und damit die Fläche  $2R \cdot 2\pi R = 4\pi R^2$ . Manchmal heißt es auch, dass die Gesamtfläche des umschriebenen Zylinders gleich  $3/2$  mal die Kugeloberfläche ist: Die beiden Deckel haben zusammen eine Fläche von  $2 \cdot \pi R^2$ ; womit man auch dieses Ergebnis leicht erhält.

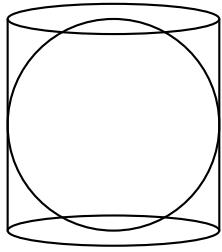


Abbildung 8: Die Archimedes-Figur. Angeblich wollte Archimedes, dass diese Figur auf seinem Grabstein abgebildet wird. Dargestellt ist eine Kugel, die einem Zylinder gleicher Höhe eingeschrieben ist. Archimedes konnte beweisen, dass die Oberfläche der Kugel gleich der Mantelfläche des Zylinders ist.

Angeblich wollte Archimedes auf seinem Grabstein die Figur aus Abb. 8 abgebildet haben, weil er die Beziehung zwischen der Kugeloberfläche und der Mantelfläche des Zylinders als seine größte Entdeckung ansah. Eigentlich hat Archimedes sogar mehr bewiesen, als dass die Gesamtflächen gleich sind; er hat bewiesen, dass kleine Ausschnitte der Kugeloberfläche, die man im Sinne der Lambert-Projektion von der Zylinderachse aus auf die Zylinderfläche projiziert, auf Flächen derselben Größe abgebildet werden. Damit hat er die Flächentreue der Lambert-Projektion bewiesen.

Archimedes hat bei vielen seiner mathematischen Beweise sehr physikalisch gedacht und oft infinitesimale Flächen in Gedanken auf eine Balkenwaage gelegt und die Hebelgesetze genutzt, um Beziehungen zwischen diesen Flächen abzuleiten. Sehr oft kann man solche Operationen mit dem Strahlensatz und dem Satz von den gleichen Verhältnissen von sich entsprechenden Seitenlängen in ähnlichen Dreiecken in Verbindung bringen. Der folgende Beweis nutzt nur diese beiden Sätze.

In Abbildung 9 (rechts) sind die beiden Dreiecke  $ABC$  und  $A'B'C'$  ähnlich. Das bedeutet:

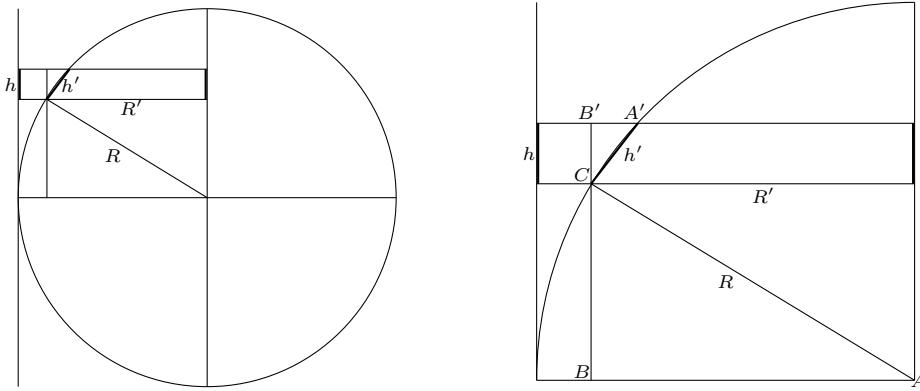


Abbildung 9: Zum geometrischen Beweis der Flächentreue der Lambert-Projektion. Die rechte Seite zeigt den Ausschnitt links-oben vergrößert. Da die Strecke  $A'C$  senkrecht auf  $AC$  steht, sind die Dreiecke  $ABC$  und  $A'B'C$  ähnlich. Daraus folgt  $R'/R = h/h'$ .

$h/h' = R'/R$ . Stellen wir uns nun vor, die infinitesimale Strecke  $h$  auf dem Zylinder wird einmal um die zentrale Zylinderachse rotiert, dann ist die überstrichene Fläche gleich  $2\pi Rh$ . Die entsprechende Fläche für das Streckenstück  $h'$  ist  $2\pi h'R'$ . Doch wegen  $h/h' = R'/R$  folgt, dass diese beiden Flächen gleich sind. Diese Aussage gilt nicht nur für den vollen Rotationskörper, sondern auch für die überstrichenen Flächen bei beliebig kleinen Rotationswinkel.

## Literatur

- [1] Wikipedia „Kartennetzentwurf“:  
[https://commons.wikimedia.org/wiki/File:Zylinderprojektion\\_quadratische\\_plattkarte\\_kl.jpg](https://commons.wikimedia.org/wiki/File:Zylinderprojektion_quadratische_plattkarte_kl.jpg)
- [2] Wikipedia „Tissot’sche Indikatrix“:  
[https://de.wikipedia.org/wiki/Tissotsche\\_Indikatrix](https://de.wikipedia.org/wiki/Tissotsche_Indikatrix)
- [3] Wikipedia „Projection équivalente cylindrique de Lambert“;  
[https://upload.wikimedia.org/wikipedia/commons/thumb/c/cb/Tissot\\_indicatrix\\_world\\_map\\_Lambert\\_cyl\\_equal-area\\_proj.svg/880px-Tissot\\_indicatrix\\_world\\_map\\_Lambert\\_cyl\\_equal-area\\_proj.svg.png](https://upload.wikimedia.org/wikipedia/commons/thumb/c/cb/Tissot_indicatrix_world_map_Lambert_cyl_equal-area_proj.svg/880px-Tissot_indicatrix_world_map_Lambert_cyl_equal-area_proj.svg.png)
- [4] Wikipedia „Gerhard Mercator“;  
<https://upload.wikimedia.org/wikipedia/commons/f/fa/Mercator-proj.png>
- [5] UN-Logo:  
[https://pngimg.com/uploads/un/un\\_PNG20.png](https://pngimg.com/uploads/un/un_PNG20.png)

# Grundlagen der SRT

Thomas Filk; Universität Freiburg

Gegen Ende des 19. Jahrhunderts mehrten sich die Anzeichen, dass irgendetwas im Weltbild der Physik nicht stimmen konnte. Die Newton'sche Mechanik war sehr erfolgreich bei der Beschreibung der Bewegungen von materiellen Körpern, insbesondere den Bewegungen der Planeten. Andererseits war die Theorie Maxwell's ebenso erfolgreich bei der Beschreibung der Phänomene im Zusammenhang mit elektrischen und magnetischen Feldern. Doch die beiden Theorien passten nicht zusammen: Die Newton'sche Theorie ist fundamental Galilei-invariant, d.h. mit jeder Bahnkurve  $x(t)$ , die eine Lösung der Bewegungsgleichungen ist, ist auch eine Galilei-transformierte Bahnkurve,  $\hat{x}(t) = x(t) + vt + a$ , eine Lösung der Bewegungsgleichungen. Hierbei ist  $v$  eine konstante Geschwindigkeit und  $a$  eine konstante Verschiebung.  $x(t)$  kann sich auch auf mehrere Komponenten und mehrere Objekte beziehen.

Die Maxwell'sche Theorie enthält jedoch als Parameter eine Geschwindigkeit (die Lichtgeschwindigkeit im Vakuum) und scheint somit ein spezielles Ruhesystem auszuzeichnen. Dieses Ruhesystem dachte man sich gleichzeitig als Ruhesystem eines Äthers, dessen oszillatorische Anregungen den elektromagnetischen Wellen entsprechen sollten, ähnlich wie die Schwingungen von Luft den Schallwellen entsprechen. Das Experiment von Michelson und Morley (siehe Abschnitt 2) war dazu gedacht, dieses Ruhesystem experimentell zu bestimmen.

Oftmals wird der Ausgang des Michelson-Morley-Experiments als Beweis dafür gewertet, dass es den Äther bzw. das ausgezeichnete Ruhesystem nicht gibt. Das ist streng genommen nicht richtig: Die dynamischen Erklärungen von Lorentz und Fitzgerald (wonach sich Gegenstände bei einer Bewegung relativ zum Äther verkürzen und alle Zeitaläufe entsprechend verlangsamen) kann sämtliche Phänomene ebenso erklären wie die heute vorherrschende Interpretation von Einstein und Minkowski, bei der kein Ruhesystem ausgezeichnet ist und bei der die geometrischen Eigenschaften der Raumzeit für die beobachteten Effekte verantwortlich gemacht werden. Man kann sogar sagen, dass jede Lorentz-invariante Theorie *per definitionem* auch die Interpretation von Lorentz und Fitzgerald zulässt. Experimentell lässt sich zwischen diesen beiden Interpretationen nicht unterscheiden (Näheres siehe Kapitel Philosophischer Hintergrund der SRT). Die Einstein'sche Interpretation hat lediglich den Vorteil, auf unbeobachtbare Entitäten wie das Ruhesystem eines Äthers und den Äther selbst verzichten zu können.

Heute verbindet man die spezielle Relativitätstheorie in erster Linie mit dem Namen Albert Einsteins, doch man sollte nicht vergessen, dass insbesondere Hendrik Antoon Lorentz (1853–1928) und Jules Henri Poincaré (1854–1912) wesentliche Vorarbeiten geliefert haben. Was die entscheidenden Schritte zur speziellen Relativitätstheorie betrifft, so werden heute meist drei Arbeiten zitiert (aus [13], S. 2 und [3], S. 408):

1. H.A. Lorentz; *Electromagnetic phenomena in a system moving with any velocity smaller than that of light* ([8]). Eingereicht hatte er diese Arbeit am 27.5.1904.
2. J.H. Poincaré; *Sur la dynamique de l'électron* ([14]). Diese Arbeit wurde bei der Französischen Akademie der Wissenschaften am 5.6.1905 eingereicht.
3. A. Einstein; *Zur Elektrodynamik bewegter Körper* ([5]). Diese Arbeit wurde am 30.6.1905 eingereicht.

Eine Diskussion, welchem Autor was zuzuschreiben ist, findet man bei Pauli [13] (S. 2/3).

# 1 Der Äther

Den Begriff des Äthers gab es in unterschiedlichen Bedeutungen und Bezeichnungen schon im Altertum. Bei den Griechen bezog sich dieser Begriff je nach Autor auf eine „leuchtende Substanz“, „Sitz der Götter“, „Urmaterie und Quintessenz (fünftes Element neben den vier bekannten Elementen)“ etc. [Brockhaus]. Eine konkretere Wiederbelebung erfuhr der Äther bei Descartes zur Erklärung der Planetenbahnen (allgemeiner zur Erklärung der Gravitation) [4] und bei Huygens als Träger der Lichtwellen. Newton setzte sich in seiner Optik ([11], Frage 18ff, besonders Frage 22: „... Äther (denn so will ich ihn nennen) ...“) mit der Ätherhypothese auseinander.

Eine klare Definition von Äther bzw. der Ätherhypothese zu geben fällt schwer, da sich die Bedeutung des Wortes wie auch die ihm zugesprochenen Eigenschaften oft gewandelt haben. Meist verstand man aber unter Äther eine „schwerelose, durchsichtige, reibungslose, chemisch oder physikalisch nicht nachweisbare und alle Materie und den gesamten Raum durchdringende Substanz“ ([3]; Stichwort ‘Ether’). Manchmal schienen die oben genannten Eigenschaften jedoch auch im Widerspruch zu den Beobachtungen zu stehen. Um den großen Wert der Lichtgeschwindigkeit erklären zu können, musste man beispielsweise eine sehr hohe Dichte des Äthers annehmen. 1816 zeigten Augustin Jean Fresnel (1788–1827) und François Arago (1786–1853), dass zwei senkrecht zueinander polarisierte Strahlen nicht interferieren und 1817 erklärte Thomas Young (1773–1829) diese Erscheinung durch die Annahme transversaler Schwingungen.

Transversale Schwingungen setzen jedoch voraus, dass es in dem Trägermedium Scherkräfte gibt, die einer transversalen Auslenkung entgegenwirken. Damit schied aber ein Äther mit den Eigenschaften von Flüssigkeiten oder Gasen (in denen nur longitudinale Wellen existieren) aus (vgl. Born [2], S. 3). Die fehlende longitudinale Polarisation konnte sogar nur erklärt werden, wenn man dem Äther die Eigenschaften eines unendlich dichten Festkörpers zuschrieb. Andererseits sollten sich aber die Planeten nahezu reibungslos durch dieses Medium bewegen können.

Eine Theorie von George Gabriel Stokes (1819–1903) zur Erklärung dieser scheinbaren Widersprüche erscheint uns heute eher absurd: Er schrieb dem Äther die Eigenschaften bestimmter nicht-newtonscher Fluide zu, wie sie beispielsweise bei Pech, Siegellack oder nassem Sand beobachtet wurden. Von diesen Stoffen war bekannt, dass sie einerseits recht schneller Schwingungen fähig sind (also die hohe Lichtgeschwindigkeit und die fehlende longitudinale Polarisation erklärbar wurde), andererseits aber auch gegenüber langsamen Beanspruchungen völlig nachgiebig sind (und dadurch die vergleichsweise langsame, nahezu reibungslose Planetenbewegung möglich war).

Im 19. Jahrhundert wurden viele Experimente unternommen, den Äther nachzuweisen. Als Beweis für die Existenz des Äthers wurde oft ein Experiment von Armand Hypolit Louis Fizeau (1819–1896) gewertet, der die Lichtgeschwindigkeit  $c'$  in einer bewegten Flüssigkeit gemessen und festgestellt hatte, dass sich die Geschwindigkeit von Licht in der ruhenden Flüssigkeit (d.h.  $c/n$ , wobei  $n$  der Brechungsindex der Flüssigkeit ist) und die Geschwindigkeit der Flüssigkeit  $v$  nicht addieren, sondern  $v$  um einen vom Brechungsindex abhängigen Faktor verringert werden muss ([3]; S. 400):

$$c' = \frac{c}{n} + \left(1 - \frac{1}{n^2}\right)v. \quad (1)$$

Diese Ergebnis konnte unter der Annahme einer partiellen, von der optischen Dichte  $n$  abhängigen Mitführung des Äthers durch die Flüssigkeit erklärt werden ([3], S. 400). Erst das verallgemeinerte Additionstheorem für Geschwindigkeiten in der Relativitätstheorie konnte diese Erscheinung auch ohne Ätherhypothese erklären. Danach erhält man (vgl. Pauli [13], S. 114):

$$c' = \frac{\frac{c}{n} + v}{1 + \frac{cv}{nc^2}} = \frac{c}{n} + v \left(1 - \frac{1}{n^2}\right) \frac{1}{1 + \frac{v}{nc}}. \quad (2)$$

In führender Ordnung von  $v/c$  stimmt dieses Ergebnis mit dem alten Resultat überein. In seiner bekannten „Geschichte der Physik“ schreibt Max von Laue ([7], S. 63) in diesem Zusammenhang:

Der Fizeausche Versuch galt lange als der schlagende Beweis für die Existenz eines Äthers, der alle Körper durchdringen sollte, ohne an ihrer Bewegung teilzunehmen. Denn nur so konnte man diesen verkleinerten Faktor verstehen. ... So ist die Geschichte des Fizeau-Versuchs ein lehrreiches Beispiel dafür, wie weit in die Deutung jedes Versuchs schon theoretische Elemente hineinspielen; man kann sie gar nicht ausschalten. Und wenn dann die Theorien wechseln, so wird aus einem schlagenden Beweise für die eine leicht ein ebenso starkes Argument für eine ganz entgegengesetzte.

Im 19. Jahrhundert war die Ätherhypothese auch Grundlage vieler Modelle von Raum, Zeit und Materie, die weit über die einfache Erklärung der Wellennatur von Licht hinausgingen. Ein interessantes Modell stammt beispielsweise von William Thomson (1824–1907), dem späteren Lord Kelvin of Largs. 1866 hatte er unter dem Eindruck der bahnbrechenden Arbeiten von Hermann von Helmholtz (1821–1894) zur Theorie der Vortizes in einem idealen Fluid (1858, [6]) – insbesondere ihrer erstaunlichen Stabilität, der Möglichkeit elastischer Stoßprozesse zwischen Vortizes und der Komplexität ihrer Strukturen – eine Theorie aufgestellt, wonach der Äther in unserem Kosmos nicht nur für die optischen, elektrischen und magnetischen Phänomene verantwortlich ist, sondern darüberhinaus auch die Atome – die Bausteine der Materie – als Verknotungen von Vortizes in diesem Äther beschreibt. Die einzelnen Atomarten entsprechen dabei topologisch verschiedenen Knotentypen. Sämtliche Naturgesetze sollten sich somit aus den statischen und dynamischen Eigenschaften des Äthers als einem idealen Fluid ableiten lassen. Dieses Modell würde sogar erklären, weshalb der Raum eines „nicht leeren“ Universums dreidimensional sein muss, denn nur in drei Dimensionen sind Knoten topologisch stabil. (Lit.: Encyclopaedia Britannica [3], Macropaedia, Stichwort ‘Helmholtz’, Bd. 20, S. 564-2b.)

## 2 Das Experiment von Michelson und Morley

Wenn der Äther tatsächlich existierte und wenn er, wie das Experiment von Fizeau andeutete, die Körper durchdringt, ohne unmittelbar an ihrer Bewegung teilzuhaben, dann sollte die Geschwindigkeit der Erde relativ zum Äther – und damit relativ zum absoluten Raum – bestimmbar sein. Auf diese Möglichkeit hatte auch bereits Maxwell hingewiesen. Da die von Maxwell, Hertz und Lorentz entwickelte Theorie des Elektromagnetismus die Lichtgeschwindigkeit  $c$  als Konstante enthielt, galt es als sicher, dass die Maxwell'schen Gleichungen nur in dem Bezugssystem gelten, in dem Licht diese Geschwindigkeit hat, d.h. dem System, in dem der Äther als Träger der Lichtwellen ruht.

Das Schlüsselexperiment zum Nachweis des Äthers sollte das Experiment von Albert Abraham Michelson (1852–1931) und Edward Williams Morley (1838–1923) werden. Der entsprechende Versuch wurde 1881 von Michelson, dann 1887 nochmals von ihm gemeinsam mit Morley durchgeführt. Mit Hilfe eines Interferometers (Abb. 1(a)) wurde die Laufzeit von Licht entlang zweier aufeinander senkrecht stehender Richtungen  $l_l$  und  $l_t$  verglichen.  $l_l$  bezeichnet dabei die Distanz in longitudinaler Richtung, d.h. der Richtung der vermuteten Erdbewegung relativ zum Äther, und  $l_t$  eine dazu senkrechte Distanz.

Relativ zum Äther hat Licht immer die Geschwindigkeit  $c$ . Für die longitudinale Richtung berechnen wir die Laufzeit am einfachsten im Laborsystem. Je nachdem, ob sich die experimentelle Anordnung in oder entgegen der Ausbreitungsrichtung des Lichts bewegt, hat das Licht im Laborsystem die Geschwindigkeit  $c + v$  bzw.  $c - v$ . Die Summe der Zeiten zur Durchquerung der Strecke  $l_l$  in beide Richtungen ist somit

$$t_l = \frac{l_l}{c+v} + \frac{l_l}{c-v} = \frac{2l_l}{c} \frac{1}{1 - \frac{v^2}{c^2}} . \quad (3)$$

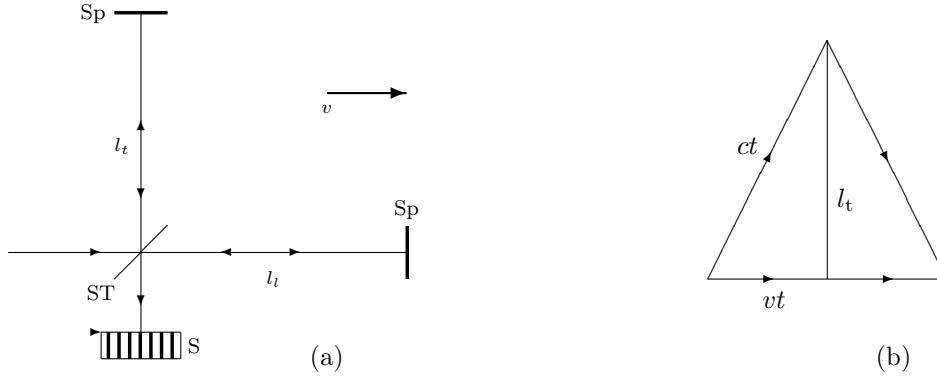


Abbildung 1: Das Michelson-Morley-Interferometer. (a) Ein Lichtstrahl trifft auf einen Strahlteiler (ST) und die beiden Anteile breiten sich in orthogonale Richtungen entlang der Strecken  $l_t$  und  $l_l$  aus. Sie werden an Spiegeln (Sp) reflektiert, treffen wieder auf den Strahlteiler und können auf dem Schirm (S) interferieren. (b) Der Strahl transversal zur Bewegungsrichtung relativ zum Äther legt die Strecke  $2ct$  zurück, während sich die Apparatur um die Strecke  $2vt$  weiterbewegt hat.

Für die transversale Richtung berechnen wir die Laufzeit im Ruhesystem des Äthers (vgl. Abb. 1(b)). Das Labor bewegt sich in diesem System mit der Geschwindigkeit  $v$  und das Licht „schräg“ dazu mit der Geschwindigkeit  $c$ , sodass die Geschwindigkeitskomponente von Licht parallel zum Laborsystem ebenfalls gleich  $v$  ist. Wir berechnen zunächst die Zeit  $t$ , die das Licht bis zum Umkehrpunkt benötigt, also die Hälfte der Zeit  $t_t$  zum Durchlaufen der gesamten Strecke. Für die vom Licht und vom Bezugssystem (Erde) zurückgelegten Strecken, bis das Licht am Umkehrpunkt ist, gilt:

$$(vt)^2 + l_t^2 = (ct)^2$$

d.h.

$$t = \frac{l_t}{\sqrt{c^2 - v^2}}$$

und damit insgesamt

$$t_t = \frac{2l_t}{c} \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} . \quad (4)$$

Durch Drehung der Apparatur um  $90^\circ$  konnten die Rollen von  $l_t$  und  $l_l$  vertauscht werden. Außerdem wurde das Experiment zu verschiedenen Jahreszeiten wiederholt, falls zu einem Zeitpunkt des Experiments die Erde zufällig relativ zum Äther ruhen sollte.

Wäre die Ätherhypothese richtig gewesen, hätte man im Rahmen einer Newton'schen Beschreibung eine Differenz zwischen der longitudinalen und der transversalen Richtung finden müssen. Das Experiment zeigte aber keine solche Differenz.

Zunächst war man derart von der Richtigkeit der Ätherhypothese überzeugt, dass man nach anderen Erklärungen für den negativen Ausgang des Michelson-Morley-Experiments suchte. Eine naheliegende Erklärung war, dass die Erde den Äther in ihrer Umgebung gleichsam mitschleppt, sodass an der Erdoberfläche die Geschwindigkeit des Äthers relativ zur Erde immer Null ist. Eine solche Erklärung widersprach aber nicht nur dem Fizeau'schen Experiment (wonach der Mitführungsterm von der optischen Dichte abhängen sollte und somit für Luft nahezu verschwindet), sondern auch der 1728 von James Bradley (1692–1762) entdeckten Aberration des Lichtes. Darunter versteht man den Effekt, dass ein Fernrohr relativ zur Richtung zu einem Stern etwas vor bzw. nachgestellt werden muss, je nach der senkrechten Geschwindigkeit der Erde relativ zu dieser Richtung ([3], S. 400). Der

Effekt beruht darauf, dass das Licht wegen der Endlichkeit der Lichtgeschwindigkeit auch eine endliche Zeit benötigt, um das Fernrohr zu durchqueren. Die Aberration ließ sich am einfachsten durch die Annahme erklären, dass die Erde den Äther nicht mitführt.

Ein interessanter Vorschlag kam 1892 von Hendrik Antoon Lorentz (1853–1928) und gleichzeitig von George Francis Fitzgerald (1851–1901). Nach ihrer Hypothese sollte jeder Maßstab als Folge der Wechselwirkung mit dem Äther in Richtung der relativen Bewegung zum Äther eine sogenannte Lorentz-Fitzgerald-Kontraktion erfahren. Diese Kontraktion bzw. Verkürzung von Längenmaßstäben sollte gerade einem Faktor  $\sqrt{1 - \beta^2}$  (mit  $\beta = v/c$ ) entsprechen. Wie ein Vergleich der Gleichungen 3 und 4 zeigt, werden die beiden Laufzeiten  $t_1$  und  $t_t$  gleich, wenn man  $l_1$  mit diesem Faktor multipliziert. Lorentz konnte in den folgenden Jahren seine Theorie soweit ausbauen, dass er nicht nur die bekannten Phänomene beschreiben sondern sogar die Transformationsgesetze formulieren konnte, die sich später aus der speziellen Relativitätstheorie ergeben sollten. Für eine widerspruchsfreie Theorie musste neben der Kontraktion von Längen auch noch angenommen werden, dass die Zeitskalen sämtlicher physikalischer Phänomene bei einer Bewegung relativ zum Äther um einen entsprechenden Faktor größer werden. Seine Überlegungen basierten jedoch immer noch auf der Annahme eines ausgezeichneten Bezugssystems, in welchem der Äther ruhte. Diese Annahme hat er auch nachdem die Relativitätstheorie ihre Triumpfe feierte nur langsam und zögerlich aufgegeben.

### 3 Axiomatische Formulierung der speziellen Relativitätstheorie

Wie schon aus den Titeln der drei in der Einleitung zu diesem Kapitel zitierten Arbeiten deutlich wird, nahm die Relativitätstheorie ihren Ausgang von der Elektrodynamik. Auch Lorentz hat sich die Frage gestellt, wie sich physikalische Systeme (z.B. solche, die wir als Uhren und Maßstäbe verwenden) verhalten, wenn ihre elementaren Bestandteile durch elektromagnetische Kräfte zusammengehalten werden. Auf diese Weise konnte er den Faktor für die Längenkontraktion aus den Maxwell'schen Gleichungen ableiten. In der Folgezeit wurde jedoch versucht, die Annahmen, die zur Herleitung der speziellen Relativitätstheorie führen, auf ein Minimum zu reduzieren. Man kann zeigen, dass die folgenden drei Axiome bereits die Struktur der speziellen Relativitätstheorie – und damit auch die Lorentz-Invarianz der fundamentalen Gleichungen – implizieren:

1. Der Raum ist homogen (überall gleich) und isotrop (es ist keine Richtung ausgezeichnet).
2. Es gilt das Relativitätsprinzip.
3. Die Konstanz der Lichtgeschwindigkeit ist unabhängig vom Bewegungszustand der Lichtquelle.

Axiom 1 wird zunächst als Erfahrungstatsache angesehen. Axiom 2 war für Einstein eine Konsequenz der fehlgeschlagenen Versuche, den Äther bzw. Bewegungen relativ zu dem ausgezeichneten Ruhesystem des Universums nachzuweisen. Wenn sich experimentell kein ausgezeichnetes Ruhesystem nachweisen lässt, dann sollte die Annahme eines absoluten Raumes oder einer absoluten Zeit auch aus der Theorie verschwinden.

Diese ersten beiden Axiome gelten auch für die Newton'sche Theorie. Es gibt also kein „Relativitätsprinzip der Relativitätstheorie“ oder relativistisches Relativitätsprinzip. Inertialsysteme sind solche Bezugssysteme, in denen die kräftefreie Bewegung geradlinig und gleichförmig verläuft. Das Relativitätsprinzip besagt, dass die Physik in allen Inertialsystemen gleich ist.

Axiom 3 ist das Minimum, auf das sich die Aussagen der Maxwell-Gleichungen reduzieren lassen, so dass zusammen mit den ersten beiden Axiomen die spezielle Relativitätstheorie eindeutig wird.

Die Konstanz der Lichtgeschwindigkeit bedeutet, dass jeder inertiale Beobachter die Wellenfronten einer punktförmigen Lichtquelle als konzentrische (gleichzentrierte) Sphären beobachtet.

Streng genommen gelten die ersten beiden Axiome nur in einem lokalen Sinne: Die Mikrowellenhintergrundstrahlung bzw. die sichtbare Masse im Universum zeichnen ein Ruhesystem aus. Solange wir aber nicht das Universum als Ganzes bzw. kosmologische Probleme betrachten, sind die ersten beiden Axiome hinreichend gut erfüllt.

Nicht alle Schritte zur Herleitung der Lorentz-Transformationen werden in voller mathematischer Strenge durchgeführt (siehe beispielsweise Pauli [13] oder Sexl und Urbantke [15]). Die folgende Herleitung umgeht die meisten mathematischen Feinheiten.

1. Aus dem Relativitätsprinzip (Axiom 2) folgt insbesondere, dass geradlinige Bewegungen wieder in geradlinige Bewegungen übergehen müssen, also Geraden in der Raum-Zeit in Geraden transformiert werden. Diese Aussage bedeutet, dass der Übergang von einem Koordinatensystem zu einem anderen nur durch eine lineare Transformation gegeben sein kann:

$$\begin{pmatrix} ct' \\ \mathbf{x}' \end{pmatrix} = \Lambda(\mathbf{v}) \begin{pmatrix} ct \\ \mathbf{x} \end{pmatrix},$$

wobei  $\Lambda(\mathbf{v})$  eine  $4 \times 4$  Matrix ist, die von der relativen Geschwindigkeit  $\mathbf{v}$  zwischen den beiden Koordinatensystemen abhängen kann. An dieser Stelle haben wir außerdem angenommen, dass beide Koordinatensysteme denselben Ursprung haben, d.h., dass die Koordinaten  $(0,0,0,0)$  für den Ursprung in beiden Systemen dasselbe Ereignis O beschreiben. Allgemeiner sind es die (bijektiven) affinen Transformationen, die sämtliche Geraden wieder in Geraden überführen. Hier beschränken wir uns auf Geraden durch den Koordinatenursprung und somit auf Transformationen, die diesen Ursprung invariant lassen.

Dieser Schritt ist übrigens mathematisch am schwierigsten zu beweisen.

2. Da die Lichtgeschwindigkeit in jedem Intertialsystem dieselbe sein soll (Axiom 3), folgt aus  $|\mathbf{x}|/t = \pm c$  auch  $|\mathbf{x}'|/t' = \pm c$ , bzw.

$$(ct)^2 - (\mathbf{x})^2 = 0 \iff (ct')^2 - (\mathbf{x}')^2 = 0. \quad (5)$$

Hierbei stellen wir uns vor, dass bei dem Ereigniss O (das für beide Koordinatensysteme dasselbe ist) ein Lichtsignal ausgesandt wurde. In Gl. 5 sollen sich  $(t, \mathbf{x})$  bzw.  $(t', \mathbf{x}')$  auf ein Ereignis A beziehen, das von diesem Lichtsignal „getroffen“ wird (siehe Abb. 2).<sup>1</sup>

Betrachten wir nun ein beliebiges Ereignis (nicht notwendigerweise auf dem Lichtkegel) mit den Koordinaten  $(t, \mathbf{x})$  bzw.  $(t', \mathbf{x}')$ , so folgt zusammen mit der Linearität der Transformation, dass sich die beiden Ausdrücke nur um einen Faktor unterscheiden können:

$$(ct)^2 - (\mathbf{x})^2 = \alpha((ct')^2 - (\mathbf{x}')^2).$$

3. Der Faktor  $\alpha$  kann noch von der Geschwindigkeit  $\mathbf{v}$  abhängen, mit der sich das eine System gegen das andere bewegt:  $\alpha = \alpha(\mathbf{v})$ . Wegen der Isotropie des Raumes (Axiom 1) sollte  $\alpha(-\mathbf{v}) = \alpha(\mathbf{v})$  gelten, und aus der Tatsache, dass die zu  $\mathbf{v}$  inverse Transformation durch  $-\mathbf{v}$  gegeben ist, folgt  $\alpha(-\mathbf{v})\alpha(\mathbf{v}) = 1$ , insgesamt also  $\alpha(\mathbf{v}) = \pm 1$ . Aus der Stetigkeit als Funktion von  $\mathbf{v}$  (sowie  $\alpha(0) = 1$ ) können wir schließen:  $\alpha(\mathbf{v}) = 1$  oder

$$(ct)^2 - (\mathbf{x})^2 = (ct')^2 - (\mathbf{x}')^2. \quad (6)$$

---

<sup>1</sup>Man beachte, dass es sich bei  $t, t'$  bzw.  $\mathbf{x}$  und  $\mathbf{x}'$  um Differenzen handelt, die sich auf zwei Ereignisse beziehen: das Ereignis A und das Referenzereignis O. Trotzdem werde ich die umständlichere Notation  $\Delta t$ ,  $\Delta \mathbf{x}$  etc. vermeiden.

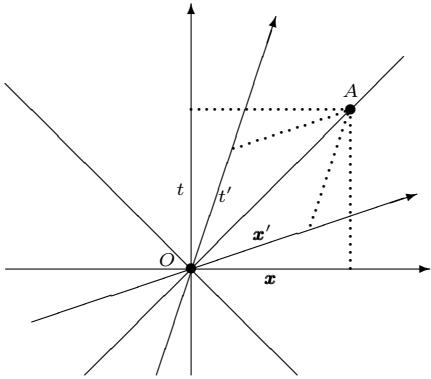


Abbildung 2: Die beiden Ereignisse  $O$  und  $A$  werden durch einen Lichtstrahl verbunden (sie sind lichtartig). Für die beiden Koordinatensysteme sei  $O$  ein Ereignis im Ursprung.  $\mathbf{x}, t$  und  $\mathbf{x}', t'$  sind jeweils die Koordinaten von Ereignis  $A$  in den beiden Koordinatensystemen.

4. Gesucht sind also alle linearen Transformationen  $\Lambda$ , welche die Kombination  $(ct)^2 - \mathbf{x}^2$  invariant lassen.

## 4 Lorentz-Transformationen

Wie allgemein üblich führen wir nun die 4-Koordinaten  $x^0 = ct$  und  $x^i$  ein und bezeichnen mit  $x$  (ohne Vektorpfeil) einen 4-Vektor:  $x = (x^0, x^1, x^2, x^3)$ , wobei  $\mathbf{x} = (x^1, x^2, x^3)$ . Die Komponenten eines 4-Vektors bezeichnen wir mit griechischen Buchstaben ( $\mu, \nu$ , etc.) und sie nehmen die Werte 0,1,2,3 an. Für die Indizes von räumlichen Komponenten verwenden wir weiterhin lateinische Buchstaben.

Dass die Indizes für die Komponenten von Vektoren hochgestellt sind, ist eine Konvention. Wir bezeichnen damit die Koordinaten von Vektoren. Wie aus der Linearen Algebra bekannt, gibt es zu jedem Vektorraum auch einen Dualraum (der Raum der linearen Abbildungen von dem Vektorraum in den jeweiligen Zahlenkörpern). Die Komponenten von Elementen des Dualraums kennzeichnen wir durch untenstehende Indizes. Außerdem verwenden wir im Folgenden noch die *Einstein'sche Summenkonvention*: Über doppelt auftretende Indizes, einmal oben und einmal unten, auf einer Seite einer Gleichung wird summiert. Diese Konvention macht viele Formeln wesentlich übersichtlicher.

Wir definieren nun ein symmetrisches, bilineares Produkt<sup>2</sup>

$$(x, y) := \eta_{\mu\nu} x^\mu y^\nu = x^0 y^0 - \sum_{i,j=1}^3 x^i y^j. \quad (7)$$

Wir bezeichnen dieses Produkt manchmal als Skalarprodukt, obwohl es nicht positiv definit und damit in der üblichen mathematischen Sprechweise kein Skalarprodukt ist. Oft nennen wir es auch Minkowski-Produkt. Dieses Produkt ist nicht-entartet, d.h., es gibt keine nicht-verschwindenden Vektoren  $y$ , sodass  $(x, y)$  für alle Vektoren  $x$  gleich null ist. Die Matrix

$$\eta = \text{diag}(1, -1, -1, -1) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad (8)$$

bezeichnen wir manchmal als Minkowski-Metrik. Auch dieser Ausdruck ist strenggenommen irreführend, da von einer Metrik üblicherweise verlangt wird, dass Abstände nie negativ werden können, was hier aber nicht der Fall ist. Daher spricht man manchmal auch von einer *Pseudo-Metrik*.

<sup>2</sup>Die Vorzeichen sind Konvention und werden insbesondere in der relativistischen Feldtheorie im Vergleich zur Relativitätstheorie oft unterschiedlich gewählt. Bei der hier angegebene Konvention ist das Skalarprodukt von zeitartigen Vektoren mit sich selbst positiv.

Durch die Bilinearform  $\eta$  können wir jedem Vektor mit Komponenten  $\{x^\mu\}$  einen dualen Vektor mit den Komponenten  $\{x_\mu\}$  zuordnen:

$$x_\mu = \eta_{\mu\nu} x^\nu. \quad (9)$$

Beim dualen Vektoren kehren sich also alle Vorzeichen der räumlichen Komponenten um.

Die Lorentz-Transformationen  $\Lambda$  bestehen aus allen linearen Transformationen, welche die Minkowski-Metrik invariant lassen. Das bedeutet

$$(x, y) = (\Lambda x, \Lambda y) \quad (10)$$

für alle 4-Vektoren  $x$  und  $y$ . Ausgedrückt in Komponenten bedeutet diese Bedingung

$$\Lambda^\alpha{}_\mu \eta_{\alpha\beta} \Lambda^\beta{}_\nu = \eta_{\mu\nu} \quad (11)$$

oder komponentenunabhängig

$$\Lambda^T \eta \Lambda = \eta. \quad (12)$$

Wir lösen diese Gleichungen für eine Raumdimension, also für  $2 \times 2$  Matrizen. Die Matrixgleichung

$$\begin{pmatrix} A & C \\ B & D \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (13)$$

führt auf drei algebraische Gleichungen,

$$A^2 - C^2 = D^2 - B^2 = 1 \quad AB = CD, \quad (14)$$

die (bis auf Vorzeichen) eine einparametrische Schar an Lösungen zulassen. Eine mögliche Parametrisierung dieser Lösungen ist:

$$A = D = \cosh \phi \quad B = C = \sinh \phi. \quad (15)$$

Man bezeichnet  $\phi$  auch manchmal als Rapidität.

Wir können aber auch eine anschaulichere Parametrisierung wählen, die sich aus folgender Überlegung ergibt: Die Weltlinie des räumlichen Ursprungs des  $(t', x')$ -Systems, also die Gerade zu  $x' = 0$ , soll sich für den anderen Beobachter als die Gerade  $x = vt$  darstellen. Damit erhält die Geschwindigkeit  $v$  erst ihre Bedeutung. Das bedeutet aber, dass

$$x' = \gamma(v)(x - vt) = \gamma(v) \left( x^1 - \frac{v}{c} x^0 \right) \quad (16)$$

sein muss, mit einem noch zu bestimmenden ( $v$ -abhängigen) Faktor  $\gamma(v)$ . Durch Vergleich mit den obigen Transformationen folgt  $A = D = \gamma$  und  $B = C = -\gamma \frac{v}{c}$ , und aus  $A^2 - C^2 = 1$  ergibt sich

$$\gamma(v) = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad (17)$$

Damit erhalten wir für die Lorentz-Transformationen in einer Raumdimension:

$$\Lambda(v) = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \begin{pmatrix} 1 & -\frac{v}{c} \\ -\frac{v}{c} & 1 \end{pmatrix}. \quad (18)$$

Für die weiteren Überlegungen wird meist diese Form der Lorentz-Transformation ausreichen. Man bezeichnet sie auch als „Boost“. Die Matrixdarstellung eines allgemeinen Boosts (für eine beliebige dreidimensionale Geschwindigkeit  $\mathbf{v}$ ) erhält man am einfachsten, indem man die Raumkoordinaten

in zur Geschwindigkeit  $\mathbf{v}$  parallele und senkrechte Komponenten aufspaltet und berücksichtigt, dass sich die senkrechten Komponenten nicht ändern (vgl. z.B. die englische Wikipedia-Seite „Lorentz transformation“):

$$\Lambda(\mathbf{v}) = \begin{pmatrix} \gamma & -\gamma\beta^T \\ -\gamma\beta & \mathbb{I} + (\gamma - 1)\beta\beta^T/\beta^2 \end{pmatrix}, \quad (19)$$

wobei  $\beta = \mathbf{v}/c$  ist,  $\beta^T$  der zugehörige transponierte (Zeilen)-Vektor und  $\beta\beta^T$  die  $3 \times 3$  Matrix mit den Komponenten  $\beta_i\beta_j$ .  $\mathbb{I}$  ist die  $3 \times 3$  Identitätsmatrix und  $\beta^2 = v^2/c^2$ .

Diese Matrizen bilden noch keine Gruppe. Die gewöhnlichen dreidimensionalen Drehungen  $R \in \text{SO}(3)$  lassen die Minkowski-Metrik ebenfalls invariant. Erst die Boosts zusammen mit den Drehungen bilden eine (sechsparametrische) Gruppe, die Lorentz-Gruppe  $\text{SO}(1,3)$ .<sup>3</sup> Man kann die Gruppe noch um räumliche Spiegelungen (Paritätstransformationen) und die zeitliche Umkehr  $t \rightarrow -t$  erweitern. Da der Minkowski-Raum ein affiner Raum ohne ausgezeichneten Raum-Zeit-Ursprung ist, erhält man insgesamt als Invarianzgruppe der Speziellen Relativitätstheorie die Poincaré-Gruppe, bestehend aus den Transformationen  $\tilde{\Lambda} + \mathbf{a}$  wobei  $\tilde{\Lambda}$  eine Lorentz-Transformation (eventuell plus Paritätstransformation oder Zeitumkehr) ist und  $\mathbf{a}$  ein beliebiger 4-dimensionaler Translationsvektor.

Abschließend soll noch das Geschwindigkeitadditionstheorem in seiner einfachsten Form (parallele Geschwindigkeiten) abgeleitet werden. Dazu betrachten wir einfach das Produkt zweier Lorentz-Boosts, für die gelten soll:

$$\gamma(v_1) \begin{pmatrix} 1 & -v_1/c \\ -v_1/c & 1 \end{pmatrix} \gamma(v_2) \begin{pmatrix} 1 & -v_2/c \\ -v_2/c & 1 \end{pmatrix} = \gamma(v_{\text{ges}}) \begin{pmatrix} 1 & -v_{\text{ges}}/c \\ -v_{\text{ges}}/c & 1 \end{pmatrix} \quad (20)$$

Durch direktes Nachrechnen (am einfachsten bildet man das Produkt auf der linken Seite und erhält  $-v_{\text{ges}}/c$  aus dem Verhältnis eines Nebendiagonalelements mit einem Diagonalelement) findet man:

$$v_{\text{ges}} = \frac{v_1 + v_2}{1 + \frac{v_1 v_2}{c^2}}. \quad (21)$$

Für nicht-relativistische Geschwindigkeiten  $v_i \ll c$  erhält man das klassische Ergebnis der Newton'schen Mechanik – die Gesamtgeschwindigkeit ist die Summe der Einzelgeschwindigkeiten.  $v_{\text{ges}}$  kann jedoch nie größer als  $c$  werden und setzt man z.B.  $v_1 = c$  so erhält man auch  $v_{\text{ges}} = c$ . Gleichung 21 hatten wir schon bei der Herleitung des Fresnel'schen Mitführungs faktors in Gl. 2 verwendet.

## 5 Die Minkowski-Raumzeit

1908 hatte Hermann Minkowski (1864–1909) die 4-dimensionale Raumzeit eingeführt und damit den Formalismus der speziellen Relativitätstheorie wesentlich vereinfacht. Berühmt geworden sind die Anfangsworte zu einem seiner Vorträge, gehalten auf der „80. Versammlung Deutscher Naturforscher und Ärzte zu Köln“ am 21. September 1908 (aus [1], S. 123):

Meine Herren! Die Anschauungen über Raum und Zeit, die ich Ihnen entwickeln möchte, sind auf experimentell-physikalischem Boden erwachsen. Darin liegt ihre Stärke. Ihre Tendenz ist eine radikale. Von Stund an sollen Raum für sich und Zeit für sich völlig zu Schatten herabsinken und nur noch eine Art Union der beiden soll Selbständigkeit bewahren.

Doch worin bestand das eigentlich Neue?

---

<sup>3</sup>Zur Notation: Die Gruppe  $\text{SO}(n,m)$  ist die Gruppe aller reellen linearen Transformationen mit Determinante 1, welche die  $(n+m) \times (n+m)$  Matrix  $\eta = \text{diag}(1, \dots, 1, -1, \dots, -1)$  invariant lassen, wobei die ersten  $n$  Einträge +1 und die letzten  $m$  Einträge -1 sind. Für die üblichen speziellen orthogonalen Gruppen ist  $m=0$  und man schreibt einfach  $\text{SO}(n)$ .

## 5.1 Die Geometrie des Minkowski-Raums

Die Besonderheit der 4-dimensionalen Minkowski-Raumzeit ergibt sich nicht einfach aus der Zusammenfassung des 3-dimensionalen gewöhnlichen Raums mit einer 1-dimensionalen Zeitkoordinate. Dies ist auch in der gewöhnlichen Newton'schen Mechanik möglich. Die Besonderheit ergibt sich aus den Invarianzen dieses Raumes bzw. der Art von Geometrie, welche durch die Minkowski-Metrik auf ihm definiert wird.

Abgesehen von räumlichen und zeitlichen Translationen ist der Newton'sche Raum invariant unter Galilei-Transformationen: dreidimensionale Rotationen sowie die speziellen Galilei-Transformationen

$$t \rightarrow t' = t \quad \text{und} \quad \mathbf{x} \rightarrow \mathbf{x}' = \mathbf{x} + \mathbf{v}t. \quad (22)$$

Dass beide Koordinatensysteme über eine affine Transformation zusammenhängen, folgt wiederum aus dem Relativitätsprinzip, das auch in der Newton'schen Mechanik gilt (Geraden in einem Inertialsystem sind Geraden in allen Inertialsystemen). Während die ersten beiden Axiome in unveränderter Form auch in der Newton'schen Mechanik gelten, kann man dort das 3. Axiom ersetzen durch: Zwei Ereignisse haben in allen Inertialsystemen denselben zeitlichen Abstand (daraus folgt  $\Delta t = \Delta t'$ ) und zwei gleichzeitige Ereignisse haben in allen Inertialsystemen denselben räumlichen Abstand (daraus folgt  $\Delta \mathbf{x}^2 = (\Delta \mathbf{x}')^2$ ). Damit liegen die Galilei-Transformationen als Invarianzgruppe fest.

Demgegenüber ist die Geometrie des Minkowski-Raums durch die Invarianz von  $(ct)^2 - \mathbf{x}^2$  bestimmt. Genauer bedeutet dies Folgendes: Zwei Ereignisse A und B werden von zwei Inertialsystemen durch die Koordinaten  $(t_A, \mathbf{x}_A)$  bzw.  $(t_B, \mathbf{x}_B)$  sowie  $(t'_A, \mathbf{x}'_A)$  und  $(t'_B, \mathbf{x}'_B)$  beschrieben. Dann gilt

$$c^2(t_A - t_B)^2 - (\mathbf{x}_A - \mathbf{x}_B)^2 = c^2(t'_A - t'_B)^2 - (\mathbf{x}'_A - \mathbf{x}'_B)^2. \quad (23)$$

Durch diese Invariante wird zwei Ereignissen ein „Abstand“ zugeschrieben, der nicht vom Koordinatensystem abhängt. Diesen Abstand werden wir später nutzen, um Linien (insbesondere Weltlinien) eine „Länge“ zuzuschreiben und Geometrie zu betreiben. Diese Minkowski-Geometrie ist zunächst etwas ungewohnt, sodass ich einige Aspekte betonen möchte.

## 5.2 Darstellung des Minkowski-Raums

Im Folgenden werden wir sehr oft Gebrauch von geometrischen Konstruktionen machen. Wir stellen dabei die Raumzeit meist vereinfacht durch eine Ebene dar, die einer Raum- und einer Zeitdimension entspricht. Die Punkte dieser Ebene repräsentieren Ereignisse und damit physikalische Tatsachen, die nicht von irgendeinem Koordinatensystem oder einer anderen Wahl der Beschreibung abhängen (vgl. Abb. 3). Ob sich zwei Personen am selben Ort treffen, oder der Zeiger einer bestimmten Uhr auf die 12 springt, oder eine Lampe an- und wieder ausgeknippt wird, oder eine Rakete dicht an einem bestimmten Satelliten vorbeifliegt – das sind Tatsachen, die für alle Beobachter gleichermaßen existent sind.

Sehr oft handelt es sich bei Ereignissen um den Schnittpunkt von Weltlinien, wobei Weltlinien von Objekten bestimmte kontinuierliche Folgen von Ereignissen und somit ebenfalls unabhängig von einem Koordinatensystem sind. Auch wenn der Zeiger einer Uhr auf 12 zeigt, schneiden sich im Prinzip zwei Weltlinien: die Weltlinie der Zeigerspitze und die Weltlinie der Markierung für die 12. Die Weltlinien von Objekten, auf die keine Kräfte wirken, werden in Minkowski-Diagrammen durch Geraden dargestellt. Insbesondere verläuft der räumliche Ursprung eines Inertialsystems entlang einer geraden Weltlinie. (Dies wird bei allgemeinen Raumzeit-Diagrammen und in der Allgemeinen Relativitätstheorie nicht immer der Fall sein.)

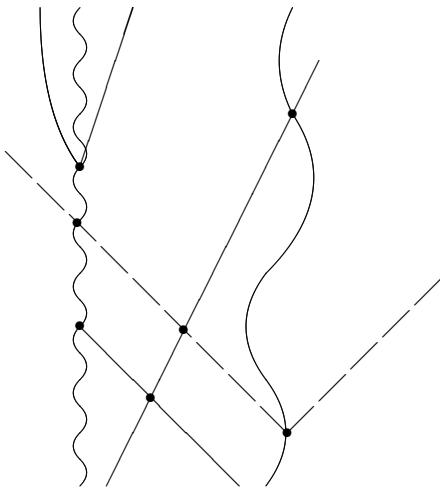


Abbildung 3: Die Raumzeit ist die Menge aller Ereignisse. Klassische Weltlinien sind kontinuierliche Folgen von Ereignissen (nicht zu verwechseln mit den „Weltlinien“ in Feynman-Graphen, hierbei handelt es sich um Repräsentationen von Propagatoren bzw. Green'schen Funktionen, nicht um reale, „faktische“ Weltlinien).

### 5.3 Die kausale Struktur

Zu jedem Ereignis können wir den Zukunfts- und den Vergangenheitslichtkegel angeben. Der Zukunftslichtkegel besteht aus allen Ereignissen, die von einem Lichtblitz, der bei dem betreffenden Ereignis ausgesandt wird, überstrichen wird. Dabei stellen wir uns vor, dass sich das Licht von diesem Ereignis aus kugelförmig in alle Richtungen ausbreitet. Die Zeitspanne des Lichtblitzes sei vernachlässigbar kurz. Der Vergangenheitslichtkegel besteht aus allen Lichtstrahlen, die das betreffende Ereignis treffen.

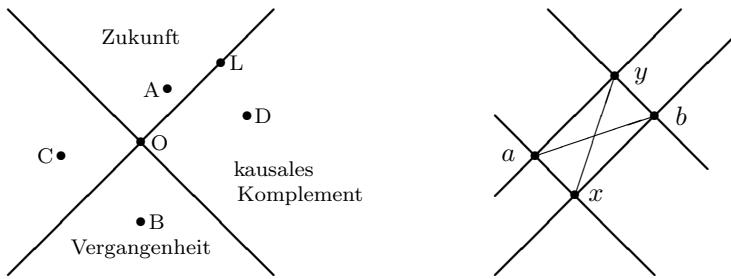


Abbildung 4: (links) Der Zukunfts- und Vergangenheitslichtkegel zu einem Ereignis O und sein kausales Komplement. (rechts) Ein „Diamant“ zu zwei Ereignissen  $x$  und  $y$  besteht aus allen Ereignissen, die sowohl in der Zukunft von  $x$  als auch in der Vergangenheit von  $y$  liegen. Er definiert zwei raumartige Ereignisse  $a$  und  $b$ , die „gerade eben noch“ im Diamanten liegen. Bis auf ein Vorzeichen ist der Abstand  $\overline{xy}$  gleich dem Abstand  $\overline{ab}$  (für  $c = 1$ ); damit lassen sich die Messungen räumlicher Abstände auf die Messungen von zeitlichen Abständen reduzieren.

Für alle Ereignisse auf dem Lichtkegel gilt (in jedem Inertialsystem)  $(ct)^2 - \mathbf{x}^2 = 0$  (wobei wir das Ereignis O als Ursprung (0,0) des Koordinatensystems gewählt haben, ansonsten sind  $t$  und  $\mathbf{x}$  entsprechend durch  $\Delta t$  und  $\Delta \mathbf{x}$  zu ersetzen). Zwei Ereignisse, die direkt durch einen Lichtstrahl verbunden werden können, also auf dem Lichtkegel des jeweils anderen Ereignisses liegen, bezeichnet

man als *lichtartig*. In Abb. 4 sind O und L lichtartig.

Für Ereignisse innerhalb des Zukunfts- oder Vergangenheitslichtkegels gilt  $(ct)^2 - \mathbf{x}^2 > 0$ . Solche Ereignispaare bezeichnet man als *zeitartig*. Die Ereignisse A und B sind zeitartig zu O. Ereignisse außerhalb des Lichtkegels (z.B. die Ereignisse C und D) bezeichnet man als relativ zu O *raumartig*. Für solche Ereignisse gilt  $(ct)^2 - \mathbf{x}^2 < 0$ .

Für zeit- und lichtartige Ereignispaare kann man eindeutig angeben, welches der beiden Ereignisse in der Zukunft relativ zu dem anderen Ereignis liegt. (Ist  $c|t| \geq |\mathbf{x}|$  und  $t > 0$ , so gibt es keine Lorentz-Transformation, für die  $t'$  negativ wird.) Diese Relation – „A liegt in der Zukunft von B“, geschrieben als  $A > B$  – ist antisymmetrisch (wenn  $A > B$ , gilt nicht  $B > A$ ). Außerdem ist diese Relation transitiv: Aus  $A > B$  und  $B > C$  folgt  $A > C$ . Bei dieser Relation handelt es sich also um eine Teilordnung.

Für raumartige Ereignisse ist eine allgemein gültige zeitliche Ordnung nicht möglich. Man kann immer Bezugssysteme finden, in denen  $\Delta t > 0$  ist, und andere Bezugssysteme, für die  $\Delta t' < 0$  ist, d.h., während in dem einen System  $C$  scheinbar später als  $O$  liegt, findet es in dem anderen System früher statt. Solche Ereignisse können sich gegenseitig nicht kausal beeinflussen.

Die beiden Lichtkegel (meist spricht man einfach von *dem* Lichtkegel) zu einem Ereignis O unterteilen also die Menge aller Ereignisse in drei Klassen: (1) die Menge der zukünftigen Ereignisse, die von O theoretisch kausal beeinflusst werden können, (2) die Menge der Ereignisse in der Vergangenheit, von denen O theoretisch kausal beeinflusst werden kann, sowie (3) die raumartigen Ereignisse, die in *keinem* kausalen Zusammenhang zu O stehen. Die für die Newton'sche Raumzeit noch sinnvolle Relation „gleichzeitig“ gibt es in der Relativitätstheorie nicht mehr in einem absoluten Sinne.

## 5.4 Inertialsysteme

In einem Inertialsystem werden alle kräftefreien Bewegungen (im Sinne der speziellen Relativitätstheorie – wir werden in der allgemeinen Relativitätstheorie auch die Geodäten in gekrümmten Raumzeiten als kräftefreie Bahnkurven ansehen) durch Geraden dargestellt. Damit bewegt sich auch der räumliche Ursprung eines Intertialsystems entlang einer Geraden. Die Koordinate entlang dieser Geraden bezeichnen wir als Zeitkoordinate. Sie wird realisiert durch die Weltline einer idealen Uhr (näherungsweise z.B. durch eine Cs-Uhr), die sich im Ursprung des Systems befindet.

Wenn wir einem Ereignis A in einem solchen Inertialsystem die Koordinaten  $(t, \mathbf{x})$  zuordnen, ist damit operational Folgendes gemeint: Wir denken uns den gesamten Raum des Inertialsystems mit Uhren ausgeplastert, die alle synchronisiert sind. (Auf die Problematik der Synchronisation von Uhren werden wir in Abschnitt 3 näher eingehen, in einer speziellen Form schon im nächsten Abschnitt. An dieser Stelle soll genügen, dass eine solche globale Synchronisation möglich ist und dass sie beispielsweise durch den langsamen Transport von Uhren – alle Uhren wurden in der fernen Vergangenheit im Ursprung auf dieselbe Zeit eingestellt und dann langsam an ihren Platz gebracht – realisiert werden kann.<sup>4)</sup>) Außerdem können wir die räumliche Lage von jeder Uhr in diesem System durch einen Vektor  $\mathbf{x}$  kennzeichnen, dessen Komponenten wir durch Anlegen eines geeichten Längenmaßstabes bestimmen können.<sup>5)</sup> Alle Uhren bewegen sich im Raumzeit-Diagramm auf parallelen Weltlinien und halten untereinander ihren Abstand. Die  $\mathbf{x}$ -Koordinate eines Ereignisses A ist dann gleich der räumlichen Koordinate der Uhr, bei der das Ereignis A stattfindet. Die  $t$ -Koordinate

<sup>4)</sup>Später werden wir zeigen, dass die so genannte Einstein-Synchronisation, bei der Uhren durch Austausch von Lichtsignalen synchronisiert werden, zu dieser Vorschrift identisch ist.

<sup>5)</sup>Auch hier ist der Austausch von Lichtsignalen und die Messung des räumlichen Abstands durch die Messung der Zeit für Hin- plus Rückweg ein praktikableres Verfahren. Das Anlegen eines Längenmaßstabes ist äquivalent und dient hier nur der Veranschaulichung.

des Ereignisses A ist gleich der Zeitanzeige dieser Uhr.

Zwei Ereignisse ereignen sich *in diesem Inertialsystem* gleichzeitig, wenn die Uhren an den jeweiligen Punkten, an denen die Ereignisse stattfinden, dieselbe Zeit anzeigen. Für ein gegebenes Inertialsystem ist es also sinnvoll, von der Gleichzeitigkeit zweier Ereignisse zu sprechen. Allerdings muss betont werden, dass die physikalische Realisation dieser Gleichzeitigkeit nur durch die Synchronisation von Uhren möglich ist, die sich bei den jeweiligen Ereignissen befinden, und eine solche globale Synchronisation ist nur in einem Inertialsystem realisierbar. Insbesondere wird der Gleichzeitigkeitsbegriff problematisch für allgemeine Bezugssysteme (die keine Inertialsysteme sind, also beispielsweise beschleunigt werden) oder auch in der allgemeinen Relativitätstheorie.

Abschließend noch eine Anmerkung zur Sprechweise. Sehr oft liest man, dass ein Beobachter ein Ereignis A zu einem bestimmten Zeitpunkt  $t$  an einem bestimmten Ort  $\mathbf{x}$  „sieht“ oder „wahrnimmt“. Dieses „sehen“ oder „wahrnehmen“ hat jedoch meist nichts mit einem physiologischen Sehen oder Wahrnehmen zu tun, sondern bedeutet, dass das Ereignis in dem Inertialsystem des Beobachters von einer entsprechenden Uhr (im oben beschriebenen Sinne) zum Zeitpunkt  $t$  am Ort  $\mathbf{x}$  registriert wird. Der Beobachter im Ursprung des Systems erfährt möglicherweise erst sehr viel später durch einen Datenaustausch von dem Ereignis und seinen Koordinaten. Für ein tatsächliches Sehen muss erst ein Lichtsignal von dem Ereignis A zu dem Beobachter gelangen. Dieses Sehen findet also im Allgemeinen später statt, außerdem kann es sein, dass unterschiedliche Ereignisse, die in dem Inertialsystem zwar gleichzeitig stattfinden (und damit dieselbe  $t$ -Koordinate haben) von dem Beobachter im Ursprung zu unterschiedlichen Zeiten gesehen werden und umgekehrt.

## Literatur

- [1] Peter C. Aichelburg (Hrsg.); *Zeit im Wandel der Zeit*; Verlag Vieweg, Braunschweig, Wiesbaden, 1988.
- [2] Max Born; *Optik*; Springer-Verlag, Berlin, Heidelberg, 1972.
- [3] Encyclopaedia Britannica; 15.th edition, 1988.
- [4] René Descartes; *Die Prinzipien der Philosophie*; Felix Meiner Verlag, Hamburg, 1992; übersetzt von Artur Buchenau.
- [5] Albert Einstein; *Zur Elektrodynamik bewegter Körper*; Annalen der Physik, Leipzig, 17 (1905) 891.
- [6] Hermann von Helmholtz; *Über Wirbelbewegungen, Über Flüssigkeitsbewegungen*, 1858; in Ostwalds Klassiker der exakten Wissenschaften Bd. 1; Verlag Harri Deutsch, Frankfurt, 1996.
- [7] Max von Laue; *Geschichte der Physik*; Universitäts-Verlag Bonn, 1947.
- [8] Hendrik Antoon Lorentz; *Electromagnetic phenomena in a system moving with any velocity smaller than that of light*; Proc. Acad. Sci., Amsterdam, 6 [1904], S. 809.
- [9] Peter Mittelstaedt; *Der Zeitbegriff in der Physik*; BI-Wissenschaftsverlag, 1989.
- [10] Peter Mittelstaedt; *Philosophische Probleme der modernen Physik*; BI-Wissenschaftsverlag, 1989.
- [11] Isaac Newton; *Über die Gravitation...*; Klostermann Texte Philosophie; Vittorio Klostermann, Frankfurt, 1988; übersetzt von Gernot Böhme.

- [12] Isaac Newton; *Optik oder Abhandlung über Spiegelungen, Brechungen, Beugungen und Farben des Lichts*; I., II. und III. Buch (1704); aus dem Englischen übersetzt von W. Abendroth; Ostwalds Klassiker der exakten Wissenschaften, Verlag Harri Deutsch 1998.
- [13] Wolfgang Pauli; *Theory of Relativity*; Dover Publications, New York, 1981.
- [14] Jules Henri Poincaré; *Sur la dynamique de l'électron*, C.R. Acad. Sci., Paris, 140 (1905) S. 1504; und Rendiconti del Circolo Matematico di Palermo, Bd. 21 (1906) S. 129.
- [15] Roman U. Sexl, Helmuth K. Urbantke; *Relativität, Gruppen, Teilchen*; Springer-Verlag, Wien, New York, 1992.
- [16] Károly Simonyi; *Kulturgeschichte der Physik*; Verlag Harri Deutsch, Thun, Frankfurt am Main, 1990.

# Philosophischer Hintergrund der SRT

Thomas Filk; Universität Freiburg

Betrachtet man vor dem Hintergrund der speziellen Relativitätstheorie das Modell von Lorentz, so wirkt die mechanische Kontraktion physikalischer Systeme und die zeitliche Verzögerung physikalischer Prozesse bei einer Bewegung relativ zum Äther eher befremdlich. Insbesondere der material-unabhängige, universelle Verkürzungsfaktor  $\sqrt{1 - (v/c)^2}$  erscheint auf den ersten Blick willkürlich. Man gewinnt den Eindruck, als ob eine entsprechende universelle Wechselwirkung zwischen Materie und Äther, die diesen Verkürzungsfaktor erklärt, nur in einem sehr komplizierten und unnatürlichen Modell beschrieben werden kann.

Dies ist jedoch nicht der Fall. Im Gegenteil, die Maxwell-Theorie oder auch unser heutiges Standardmodell sind ebenfalls in der Lage, diesen Verkürzungsfaktor zu erklären. Der Unterschied zwischen der speziellen Relativitätstheorie im Sinne von Einstein und der Lorentz'schen Theorie (Äther, Verkürzungsfaktor, etc.) erweist sich nur als eine Frage der Interpretation. Der mathematische Formalismus ist derselbe und kein Experiment kann zwischen den beiden Interpretationen unterscheiden. Die heute verbreitete Interpretation, nach der die Effekte der speziellen Relativitätstheorie als Folgerung der Geometrie der Minkowski-Raumzeit gedeutet werden, hat den Vorteil, keine unbeobachtbaren Elemente - die Entität des Äthers oder ein ausgezeichnetes Ruhesystem - zu enthalten. Man spricht in diesem Zusammenhang auch gerne von Ockhams Rasiermesser (Ockham's Razor): Hierbei handelt es sich um ein nach Wilhelm von Ockham (1288-1347) benanntes Prinzip der Wissenschaftstheorie, wonach ein Erklärungsmodell möglichst einfach sein sollte, bzw. zwischen zwei Modellen dasjenige den Vorzug hat, das mit weniger Variablen auskommt.

Andererseits ist es oft von Vorteil, möglichst viele gleichberechtigte Perspektiven auf einen Sachverhalt zu haben, und es gibt durchaus Situationen, in denen die Lorentz'sche Interpretation der Relativitätstheorie eine intuitivere Vorstellung vermittelt als die Einstein'sche. Daher beginnt dieser Abschnitt mit einer Beschreibung dieser Sichtweise.

## 1 Eine Kette gekoppelter Pendel

Wir betrachten zunächst ein einfaches mechanisches Modell, in dem die Längenmaßstäbe eine Lorentz-Kontraktion erfahren, wenn sie sich mit einer Geschwindigkeit  $v$  relativ zu dem absolut ruhenden Bezugssystem bewegen. Auch die anderen bekannten Beziehungen aus der speziellen Relativitätstheorie wie beispielsweise die Zeitdilatation werden in diesem Modell wiedergegeben.

Als physikalisches Modell stelle man sich eine Kette harmonisch gekoppelter Pendel in einem konstanten Gravitationsfeld vor. Die Pendel seien mit einem Index  $i$  durchnummieriert, wobei wir zunächst  $i$  die ganzen Zahlen durchlaufen lassen. Der Freiheitsgrad des  $i$ -ten Pendels ist der Winkel  $\varphi_i$  relativ zur herabhängenden Ruhelage. Die Wirkung dieses Modells lautet:

$$S = \frac{1}{2} \int dt \sum_i \left[ \left( \frac{\partial \varphi_i(t)}{\partial t} \right)^2 - D [\varphi_{i+1}(t) - \varphi_i(t)]^2 + 2g \cos \varphi_i(t) \right],$$

und die zugehörige Bewegungsgleichung des  $i$ -ten Pendels ist:

$$\frac{\partial^2 \varphi_i(t)}{\partial t^2} - D[\varphi_{i+1}(t) - 2\varphi_i(t) + \varphi_{i-1}(t)] + g \sin \varphi_i = 0. \quad (1)$$

Suchen wir nach Lösungen, bei denen sich die Winkel benachbarter Pendel nicht wesentlich unterscheiden, muss  $g \ll D$  gelten. In diesem Fall können wir das diskrete Modell durch ein kontinuierliches

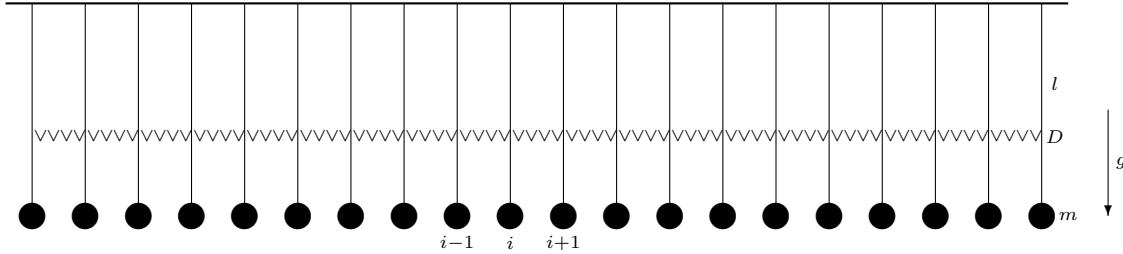


Abbildung 1: Eine Kette harmonisch gekoppelter Pendel im Schwerkraftfeld der Erde. Die Pendel können senkrecht zur Darstellungsebene schwingen. Ihr Freiheitsgrad ist der Auslenkungswinkel relativ zur Senkrechten. Die Pendelkugeln haben eine Masse  $m$ , die harmonische Feder zwischen den Pendeln die Federkonstante  $D$ . Die Länge der Pendel sei  $l$  und  $g$  sei die Schwerebeschleunigung der Erde.

feldtheoretisches Modell mit der Wirkung

$$S = \frac{1}{2} \int dt dx \left[ \left( \frac{\partial \varphi(x,t)}{\partial t} \right)^2 - \left( \frac{\partial \varphi(x,t)}{\partial x} \right)^2 + g \cos \varphi(x,t) \right]$$

und den Feldgleichungen

$$\frac{\partial^2 \varphi}{\partial t^2} - \frac{\partial^2 \varphi}{\partial x^2} + g \sin \varphi = 0 \quad (2)$$

approximieren. Hier wurde  $D = 1$  gesetzt und der kontinuierliche Parameter  $x$ , der den Ort entlang der Aufhängung bezeichnet, ersetzt die Nummerierung der Pendel. Die folgenden Überlegungen gehen immer von diesem kontinuierlichen Modell aus. Die Kennzeichnung der Raum-Zeit-Punkte  $(x,t)$  bezieht sich jedoch auf eine feste, Newton'sche Hintergrundsraumzeit. Daher ist es auch ganz instruktiv, sich die Kette gekoppelter Pendel vorzustellen, da in diesem Fall der klare – Newton'sche – Charakter von „Raum“ und „Zeit“ deutlicher wird.

Wir wissen, dass die Kontinuumsfeldgleichung (2) invariant unter Lorentz-Transformationen ist. Eine solche Feststellung ist zunächst keine Aussage über die zugrundeliegende Raum-Zeit-Struktur, sondern bezeichnet eine Eigenschaft der Lösungsmenge der Gleichungen: Wenn  $\varphi_0(x,t)$  eine Lösung der Gleichung (2) ist, dann ist auch

$$\varphi_v(x,t) = \varphi_0(\gamma(v)(x-vt), \gamma(v)(t-vx)) \quad \text{mit} \quad \gamma(v) = \frac{1}{\sqrt{1-v^2}} \quad (3)$$

eine Lösung dieser Gleichung für beliebiges  $-1 < v < 1$ . Solange  $g \ll 1$ , bzw. der Wert für  $v$  so eingeschränkt wird, daß auch  $g\gamma(v) \ll 1$  (in unserer Normierung ist  $c = 1$ ), werden die Lösungen der diskretisierten Gleichung (1) durch die Lösungen der Kontinuumsgleichung angenähert. Im Rahmen dieser Näherung werden die diskretisierten Lösungen daher auch dieselbe Invarianzeigenschaft (3) zeigen.

Da wir untersuchen wollen, wie sich Längenmaßstäbe und Uhren verhalten, wenn man sie gegen das Ruhesystem bewegt, müssen wir zunächst intrinsische „Lineale“ und „Uhren“ definieren. Dazu benutzen wir die besonderen Lösungen der Sinus-Gordon-Gleichung: die Soliton-Lösung und die Lösung zu gebundenen Zuständen von zwei Solitonen – die Breather-Lösungen. Die genaue Form der Lösungen spielt für das Folgende keine Rolle, wird aber (um wirklich explizit zu sein) angegeben. Wie wir sehen werden, ist nur die oben erwähnte Invarianz der Lösungsmenge von Bedeutung.

Die Soliton-Lösung entspricht einer Konfiguration, bei der sich die Pendel in der Nähe eines Punktes einmal um ihre Aufhängung herumwinden, sich anderenfalls aber im Wesentlichen in der senkrecht herabhängenden Ruhelage befinden (siehe Abb. 2). Diese Lösung ist stabil. Die statische

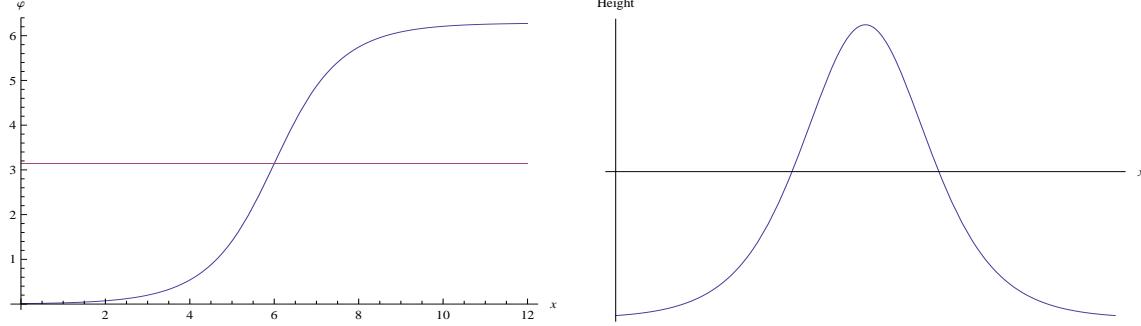


Abbildung 2: Die Soliton-Lösung: (oben links) Der Auslenkungswinkel  $\varphi$  als Funktion der Koordinate  $x$  entlang der Aufhängung, (oben rechts) die Seitenansicht.

Lösung des Kontinuumsmodells ist durch

$$\varphi_0(x) = 4 \tan^{-1} (\exp \pm \sqrt{g}(x - x_0)) \quad (4)$$

gegeben, wobei die Integrationskonstante  $x_0$  die Position des Solitons angibt, d.h. den Punkt, bei dem  $\varphi = \pi$ . Das + Zeichen in (4) entspricht einer Lösung, für die  $\lim_{x \rightarrow -\infty} \varphi(x) \rightarrow 0$  und  $\lim_{x \rightarrow +\infty} \varphi(x) \rightarrow 2\pi$ . Das - Zeichen beschreibt eine sogenannte Anti-Soliton-Lösung mit  $\lim_{x \rightarrow -\infty} \varphi(x) \rightarrow 2\pi$  und  $\lim_{x \rightarrow +\infty} \varphi(x) \rightarrow 0$ . In beiden Fällen nähern sich die Lösungen ihrem asymptotischen Wert für große  $|x - x_0|$  exponentiell. Die Reichweite der Lösung entspricht dabei

$$\Delta L = 1/\sqrt{g}. \quad (5)$$

Diese Größe ist ein Maß für die halbe Breite des Solitons und soll uns als Längenskala dienen.

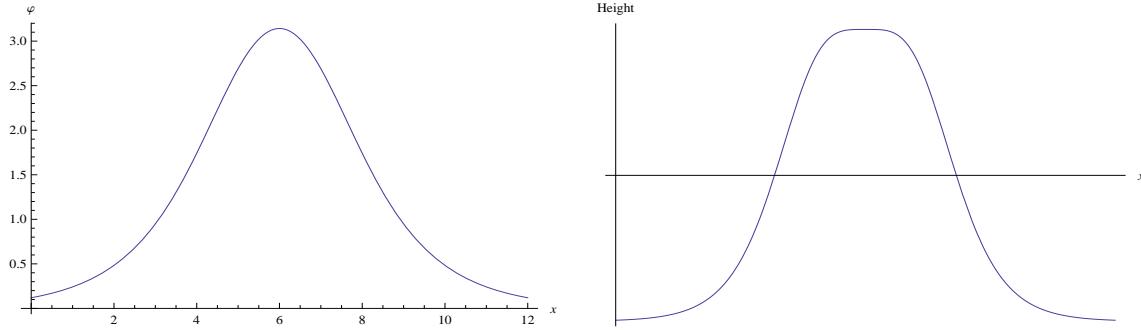


Abbildung 3: Die Breather-Lösung: (oben links) Der Auslenkungswinkel  $\varphi$  als Funktion der Koordinate  $x$  entlang der Aufhängung im Augenblick maximaler Auslenkung, (oben rechts) die Seitenansicht.

Die Sinus-Gordon Theorie besitzt ebenfalls Lösungen, die gebundenen Zuständen eines Solitons und eines Anti-Solitons entsprechen, die sogenannten Breather-Lösungen (vgl. [1]; Darstellung in Abb. 3):

$$\varphi_0(x,t) = -4 \tan^{-1} \left[ \frac{m}{\sqrt{1-m^2}} \frac{\sin \sqrt{g(1-m^2)}(t-t_0)}{\cosh m\sqrt{g}(x-x_0)} \right]. \quad (6)$$

Der Parameter  $m$  muss der Bedingung  $0 < m^2 < 1$  genügen, ist ansonsten aber beliebig. Die Schwingungsperiode ist

$$\Delta T = \frac{2\pi}{\sqrt{g(1-m^2)}}. \quad (7)$$

Diese Lösung ist metastabil. Es handelt sich um einen gebundenen Zustand zwischen einem Teilchen und seinem Antiteilchen, und kleine Störungen lassen dieses System in „Photonen“ zerfallen, d.h. einfache Schwingungen der Pendelkette. Der freie Parameter  $m$  hängt mit der Bindungsenergie zusammen und legt sowohl die Amplitude als auch die Periode fest. Zur Festlegung einer Zeitskala muss daher ein spezieller Wert für  $m$  gewählt werden. Eine mögliche Wahl wäre, dass  $\varphi$  zwischen  $-\pi$  und  $+\pi$  variiert, d.h.  $m^2 = 1/2$  oder  $\Delta T = \sqrt{8\pi^2/g}$ . (Dieser Wert für  $m$  wurde in Abb. 3 gewählt.)

Wir wollen nun untersuchen, wie sich die so definierten Längen- und Zeitmaßstäbe transformieren, wenn man sie relativ zum „Äther“, d.h. dem durch die Pendel definierten Ruhesystem, bewegt. Wir betrachten zunächst ein Soliton mit einer Geschwindigkeit  $v$ . Diese Lösung erhalten wir aus der statischen Lösung nach Gleichung (3). Auch die Breite des propagierenden Solitons lässt sich daraus ablesen.

Die statische Lösung  $\varphi_0(x)$  habe ihr Zentrum bei  $x = 0$ . Dann bestimmt die Bedingung

$$\varphi_0(\pm\Delta L) = \pi \pm \alpha$$

einen Wert für  $\alpha$ , bei dem wir die Breite von  $\varphi_0$  messen. Für die propagierende Lösung

$$\varphi_v(x,t) = \varphi_0(\gamma(v)(x-vt)) ,$$

bewegt sich das Zentrum  $x_0(t)$  nach der Gleichung  $x_0(t) = vt$ . Die Breite dieser Lösung ist gleich dem Wert  $\Delta L_v$ , der der Bedingung

$$\varphi_v(x_0(t) \pm \Delta L_v, t) = \pi \pm \alpha$$

genügt. Dies ist offensichtlich der Fall für

$$\gamma(v) \Delta L_v = \Delta L \quad \text{oder} \quad \Delta L_v = \sqrt{1-v^2} \Delta L . \quad (8)$$

Die Breite der propagierenden Lösung ist daher um einen Faktor  $1/\gamma(v)$  kontrahiert. Diese „Lorentz-Kontraktion“ kann von einem außenstehenden Beobachter, der sieht, wie sich die Solitonen entlang der Pendelkette ausbreiten, tatsächlich wahrgenommen werden.

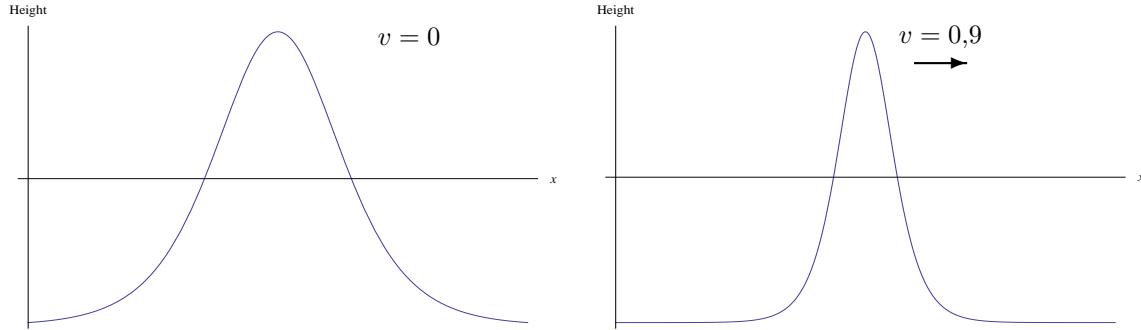


Abbildung 4: Die Soliton-Lösung in Bewegung: (links) Die ruhende Soliton-Lösung aus Abb. 2, (rechts) die zugehörige transformierte Lösung zu einer Geschwindigkeit  $v = 0,9$ .

Auch die Dilatation der Schwingungsperiode des gebundenen Zustands lässt sich leicht herleiten. In Ruhe gilt für die Breather-Lösung:

$$\varphi_0^b(x,t) = \varphi_0^b(x,t + \Delta T) .$$

Die zugehörige propagierende Lösung erfüllt

$$\varphi_v^b(x,t) = \varphi_v^b(x + v\Delta T_v, t + \Delta T_v) .$$

Da

$$\varphi_v^b(x,t) = \varphi_0^b(\gamma(v)(x-vt), \gamma(v)(t-vx))$$

erhalten wir

$$\gamma(v)(t + \Delta T_v - v(x + v\Delta T_v)) = \gamma(v)(t - vx) + \gamma(v)(1 - v^2)\Delta T_v$$

und somit:

$$\Delta T = \gamma(v)(1 - v^2)\Delta T_v \quad \text{oder} \quad \Delta T_v = \frac{1}{\sqrt{1 - v^2}}\Delta T. \quad (9)$$

Die Schwingungsdauer eines sich bewegenden gebundenen Zustands ist also tatsächlich um einen Faktor  $\gamma(v)$  im Vergleich zum ruhenden Zustand verlängert.

Solitonen und gebundene Zustände von Solitonen können an einem festen Ende der Kette ohne Energieverlust reflektiert werden. Durch eine solche experimentelle Anordnung kann man das „Zwillingssparadoxon“ aufzeigen (vgl. Abschnitt 8): Ein gebundener Zustand, der sich entlang der Kette bewegt und nach einer Reflexion am Ende der Kette zurückkehrt, hat weniger Schwingungen ausgeführt, als ein ruhender gebundener Zustand. Offensichtlich kann dieser Effekt nicht der „Beschleunigung“ am Ende der Kette zugeschrieben werden.

## 2 Von der Ätherhypothese zur Relativitätstheorie

Zunächst hat es vielleicht den Anschein, als ob das oben diskutierte Modell der harmonisch gekoppelten Pendel sehr speziell sei. Hinsichtlich der Existenz von Soliton- und Breather-Lösung mag das richtig sein, nicht aber hinsichtlich der Tatsache, dass dieses Modell den richtigen Verkürzungsfaktor und die richtige Zeitdilatation liefert. Das einzige, was wir zur Herleitung dieser Faktoren benutzt haben, ist die Lorentz-Invarianz der Feldgleichungen. *Jede Lorentz-invariante Feldgleichung hat somit die Eigenschaft, dass ihre Lösungen, die eine Längen- oder Zeitskala definieren, sich genau so transformieren, dass die bewegten Maßstäbe bzw. Uhren mit dem Lorentz'schen Verkürzung- bzw. Dilatationsfaktor multipliziert werden.*

Insbesondere haben auch die Maxwell-Gleichungen diese Eigenschaft. Genau das hatte Lorentz gezeigt, wobei er allerdings noch die falschen Transformationsgesetze für die Ladungen und Ströme verwendet hatte. Diesen Fehler hat Poincaré korrigiert. Unter der Annahme, dass auf atomarer Ebene die Materie, und damit auch Längenmaßstäbe, durch die elektromagnetische Wechselwirkung zusammengehalten wird, war daher die Lorentz-Kontraktion nicht nur plausibel sondern sogar eine Folgerung aus den Maxwell-Gleichungen.

Vor diesem Hintergrund können wir auch den Übergang von der Lorentz'schen Sichtweise zur Einstein'schen Sichtweise und der speziellen Relativitätstheorie leichter nachvollziehen. Aus der Lorentz-Invarianz der fundamentalen Gleichungen folgen ja nicht nur die Lorentz-Kontraktion und die Zeitdilatation, sondern sämtliche Phänomene, die sich im Rahmen der speziellen Relativitätstheorie erklären lassen.

Oftmals wird behauptet, der Übergang zur Sichtweise der speziellen Relativitätstheorie bestünde in einer Neuinterpretation von „Raum“ und „Zeit“. Doch über Raum und Zeit als abstrakte Entitäten können wir eigentlich nichts sagen. In der Physik interessieren wir uns nur für räumliche und zeitliche *Distanzen*, und diese Distanzen messen wir mit physikalischen Instrumenten – Längenmaßstäben und Uhren –, die selbst den physikalischen Gesetzen unterliegen.

Bisher bezog sich das Symbol  $x$  auf eine Koordinate im Ruhesystem des Äthers – im obigen Modell war dies das Ruhesystem der Pendelaufhängung. Dies entspricht dem „Lineal“ eines externen Beobachters, der außerhalb unserer Welt steht und alles mit seinen Maßstäben ausmessen kann –

wie wir bei den gekoppelten Pendeln. Entsprechend bezog sich auch das Symbol  $t$  auf die Uhr eines externen Beobachters. Wir als externe Beobachter können die Längenkontraktion der Solitonen nachweisen, indem wir einfach ein „externes“ Lineal neben das Soliton halten. Ebenso können wir den propagierenden gebundenen Zustand beobachten und die Dilatation in der Schwingungsperiode mit unserer externen Uhr vergleichen. Für uns als externe Beobachter sind Raum (gemessen entlang der Pendelkette) und Zeit absolut.  $c = 1$  (die Geschwindigkeit von Wellen bzw. kleinen Störungen entlang der Kette) ist für uns keine obere Grenzgeschwindigkeit, und wenn wir uns entlang der Kette bewegen, ist die Ausbreitungsgeschwindigkeit einer Störung entlang der Kette (das entspricht der Ausbreitung des Lichts) in unserem System für die Vorwärts- und Rückwärtsrichtung verschieden: Das Ruhesystem der Pendelkette ist ein ausgezeichnetes System.

Stellen wir uns nun jedoch (1+1)-dimensionale Wesen vor, die in dieser „Soliton-Welt“ leben. Ihnen stehen nur die Solitonen bzw. andere Lösungen ihrer „universellen Bewegungsgleichungen“ als Längen- und Zeitmaßstäbe zur Verfügung. Sie werden daher die Breite eines Solitons und die Schwingungsdauer der Breather-Lösung als Längen- und Zeitmaßstab zur Beschreibung der physikalischen Phänomene benutzen. Doch wenn sie ein bewegtes Soliton mit der Breite eines anderen bewegten Solitons vergleichen, „messen“ sie keine Verkürzung, entsprechend messen sie auch mit bewegten Breather-Lösungen keine Veränderungen in der Zeitskala eines entsprechend bewegten Systems. Erst wenn sie die Breite oder Zeitskalen von bewegten Systemen mit denen von anders bewegten oder ruhenden Systemen vergleichen, messen sie Unterschiede.

Der Übergang zur Minkowski-Welt erfolgt gerade dadurch, dass die externen Längen- und Zeitmaßstäbe durch interne Längen- und Zeitmaßstäbe, die denselben physikalischen Gesetzen unterliegen, ersetzt werden. Es ist der Übergang von einer externen zu einer internen Perspektive. Die zugrundeliegende diskrete Struktur und damit das ausgezeichnete Ruhesystem (der „Äther“) zeigen sich erst, wenn  $v$  so groß wird, dass die Breite der Solitonen mit der Größenordnung des Pendelabstands (bzw. der Gitterstruktur) vergleichbar wird. Ist die fundamentale Theorie eine Kontinuumstheorie, so tritt der „Äther“ für einen internen Beobachter überhaupt nicht in Erscheinung.

Wenn wir von der Lorentz-Invarianz bzw. allgemeiner Poincaré-Invarianz der Raum-Zeit sprechen, sollten wir eigentlich zwei Schritte bzw. Aspekte unterscheiden. Wir haben oben die Invarianzeigenschaft der Bewegungsgleichungen als eine Eigenschaft der Lösungsmenge dieser Gleichungen interpretiert: Mit jeder Lösung  $\varphi(x)$  ist auch  $\varphi^{(\Lambda,a)}(x) = \varphi(\Lambda x - a)$  eine Lösung der Bewegungsgleichung. Gibt es nun ausgezeichnete Lösungen, die Längenmaßstäbe oder Uhren definieren, so bedeutet diese Invarianz, dass bewegte Längenmaßstäbe kürzer, bewegte Uhren langsamer erscheinen. Dies gilt zunächst bezüglich der festen „Hintergrunds-Raum-Zeit“. Diesen Schritt hatte auch Lorentz erkannt. Der wesentliche zweite Schritt aber blieb Einstein vorbehalten. Er erkannte nämlich, dass die Poincaré-Invarianz einer Gleichung auch bedeutet, dass bezüglich der intrinsischen Maßstäbe das Relativitätsprinzip gilt und damit kein Inertialsystem mehr ausgezeichnet ist.

Statt die Lorentz-Invarianz der Gleichungen als Eigenschaft der Lösungsmenge anzusehen, was beispielsweise auch für die diskretisierte Pendelkette einfach zu interpretieren war, können wir die Lorentz-Invarianz auch als eine Freiheit der Koordinatenwahl auffassen. Wenn wir nämlich die Raum- und Zeitkoordinaten einer Lorentz-Transformation unterwerfen, für eine Raumdimension somit die Ersetzung

$$x \rightarrow x' = \frac{x - vt}{\sqrt{1 - \beta^2}} \quad \text{und} \quad t \rightarrow t' = \frac{t - \frac{v}{c^2}x}{\sqrt{1 - \beta^2}} \quad \left( \beta = \frac{v}{c} \right) \quad (10)$$

vornehmen, dann bleiben die Feldgleichungen unverändert. Ordnen wir nun die Koordinaten  $(x,t)$  bzw.  $(x',t')$  jeweils Beobachtern zu, deren Weltlinie durch  $x = 0$  bzw.  $x' = 0$  und deren „gleichzeitige Ereignisse“ durch die Bedingungen  $t = \text{const.}$  bzw.  $t' = \text{const.}$  definiert sind, dann bedeutet die Invarianz der Feldgleichungen, dass beide Beobachter dieselbe Physik sehen, d.h., es gilt das

Relativitätsprinzip.

Die Lorentz-Invarianz der Bewegungsgleichungen können wir offensichtlich physikalisch auf zwei vollkommen unterschiedliche Weisen interpretieren. Einmal als Eigenschaft der Lösungsmenge, die es uns erlaubt, aus bestimmten Lösungen andere zu gewinnen. Diese Interpretation bezieht alles auf einen festen Satz von Koordinaten  $(x,t)$  (vgl. (3)), der als das Koordinatensystem eines ausgezeichneten Ruhesystems interpretiert werden kann. Bezuglich dieser Koordinaten beobachten wir die Lorentz-Kontraktion und die Zeitdilatation.

Die andere Interpretation impliziert das Relativitätsprinzip. Wenn sich zwei Beobachter relativ zueinander mit einer konstanten Geschwindigkeit  $v$  bewegen, können sie unterschiedliche Koordinatensysteme verwenden, in denen sie sich jeweils im Ursprung (und damit in Ruhe) befinden. Jeder verwendet zur Angabe von räumlichen und zeitlichen Distanzen Werte, die mit Instrumenten in seinem System gemessen wurden. Die von den beiden Beobachtern verwendeten Koordinaten hängen über die Gleichungen (10) miteinander zusammen. In diesem Fall erfahren beide Beobachter dieselbe Physik. In dieser Interpretation gibt es kein ausgezeichnetes Ruhesystem – der Äther ist verschwunden. Der Preis ist eine neue Vorstellung von Gleichzeitigkeit. Dieser letzte Schritt zeichnete Einstein vor Lorentz und Poincaré aus.

### 3 Die Synchronisation von Uhren

Ein ungewohnter (und in manchen Kreisen immer noch umstrittener) Aspekt der speziellen Relativitätstheorie ist ihr Konzept von Gleichzeitigkeit. Dabei wird oft vergessen, dass es sich bei der Behauptung, zwei Ereignisse seien gleichzeitig, nicht um eine „faktische“ Aussagen handelt, sondern eher um eine Konvention.<sup>1</sup> Ob zwei Ereignisse für ein gegebenes Inertialsystem gleichzeitig stattfinden oder nicht, können wir experimentell nur entscheiden, wenn wir eine Vorschrift angeben, was wir unter „gleichzeitig“ verstehen wollen. Diese Vorschrift wird letztendlich darin bestehen, dass wir zwei Ereignisse als gleichzeitig ansehen, wenn die Uhren am jeweiligen Ort dieser Ereignisse dieselbe Zeit anzeigen, womit wir das Problem auf die Synchronisation von Uhren an verschiedenen Orten (aber im selben Inertialsystem) reduzieren haben. Die einzige Einschränkung ist, dass zwei gleichzeitige Ereignisse  $A$  und  $B$  *nicht* in einer kausalen Abhängigkeit stehen sollten. Grundsätzlich können wir eine beliebige raumartige Hyperfläche als gleichzeitig definieren.

Die folgenden Überlegungen gelten für den flachen Minkowski-Raum. Wir konzentrieren uns ausschließlich auf den Aspekt der Synchronisation von Uhren für ein gegebenes Inertialsystem. Eine ausführlichere Darstellung dieser Problematik findet man bei Mittelstaedt [2] sowie in den beiden Abhandlungen *Philosophie der Raum-Zeit-Lehre* [3] und *Axiomatik der relativistischen Raum-Zeit-Lehre* [4] von Hans Reichenbach (Hans Friedrich Herbert Günther Reichenbach, geb. 26.9.1891 in Hamburg; gest. 9.4.1953 in Los Angeles).

#### 3.1 Synchronisation durch Lichtsignale

Zunächst ist es recht hilfreich, sich bei der operationalen Definition von Gleichzeitigkeit auf ein allgemeines Verfahren zu beschränken, das allerdings noch keine Einschränkung an die Möglichkeiten darstellt. Wir wollen Uhren durch Austausch von Lichtsignalen synchronisieren. Beobachter  $A$  sendet zum Zeitpunkt  $t_1$  ein Lichtsignal aus. Dieses wird von Beobachter  $B$  reflektiert und erreicht Beobachter  $A$  zum Zeitpunkt  $t_2$ . Bezeichnen wir den Zeitpunkt des Ereignisses der Reflektion des Lichtstrahls bei Beobachter  $B$  mit  $t$ , so möchte Beobachter  $A$  nun definieren, welches Ereignis auf seiner Weltlinie

---

<sup>1</sup>Versteht man unter „faktisch“, dass sich alle Beobachter unabhängig von ihrem Bezugssystem in einer Aussage einig sind, so ist die Gleichzeitigkeit in der Newton'schen Physik faktisch, in der Relativitätstheorie jedoch nicht.

zu dem Moment der Reflektion des Lichtstrahls bei Beobachter  $B$  gleichzeitig war, d.h. dem Zeitpunkt  $t$  entspricht. Die einzige Einschränkung liefert dabei die Forderung der Kausalität:  $t_1 < t < t_2$ . Beobachter  $A$  definiert nun:

$$t = t_1 + \epsilon(A,B)(t_2 - t_1)$$

mit

$$0 < \epsilon(A,B) < 1 .$$

(Der Einfachheit halber soll das Synchronisationsverfahren nicht von der Zeit abhängen, d.h.  $\epsilon(A,B)$  hängt nicht zusätzlich noch von  $t$  bzw.  $t_1$  ab.)

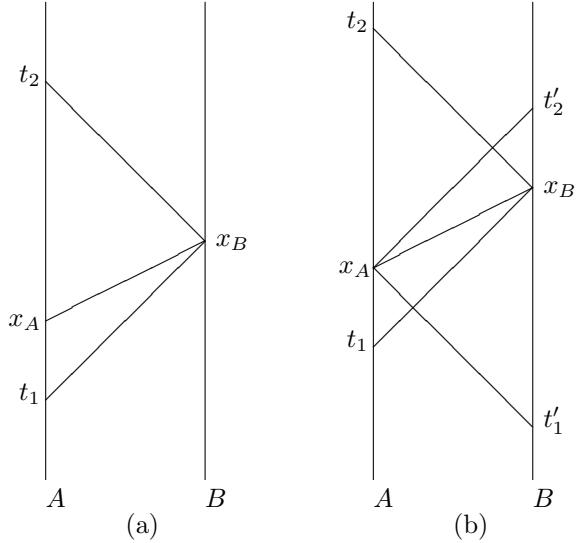


Abbildung 5: Synchronisation von Uhren durch Lichtstrahlen. Die senkrechten Linien entsprechen den Weltlinien der Beobachter  $A$  und  $B$ , die zueinander einen konstanten Abstand halten. (a) Zum Zeitpunkt  $t_1$  sendet  $A$  ein Lichtsignal aus, das von  $B$  im Ereignis  $x_B$  reflektiert wird und bei  $t_2$  wieder Beobachter  $A$  erreicht. Durch Vorgabe von  $\epsilon_A$  konstruiert  $A$  das Ereignis  $x_A$ , das zu  $x_B$  gleichzeitig ist. (b) Soll die Gleichzeitigkeit der Ereignisse  $x_A$  und  $x_B$  auch für Beobachter  $B$  gelten, so muss er sein Verfahren zur Bestimmung der Gleichzeitigkeit (ausgedrückt durch  $\epsilon_B$ ) so wählen, dass  $\epsilon_B = 1 - \epsilon_A$ .

Zunächst kann man sich leicht überzeugen, dass durch geeignete Wahl von  $\epsilon$  tatsächlich je zwei raumartige Ereignisse  $x_A$  und  $x_B$  auf der Weltlinie von  $A$  und  $B$  als gleichzeitig definiert werden können (vgl. Abb. 5). Dieses Verfahren bildet daher keine Einschränkung der Allgemeinheit. Damit die Hyperfläche der zu  $x_A$  gleichzeitigen Ereignisse eine raumartige Hyperfläche ist, müssen die Richtungsableitungen der Funktion  $\epsilon(A,B)$  (aufgefasst als eine Funktion von  $B$ ) noch durch eine Konstante (bei unserer Definition der Lichtgeschwindigkeit ist diese Konstante 1) beschränkt sein. Durch geeignete Forderungen an die Funktion  $\epsilon$  erhält man verschiedene Synchronisationsverfahren.

### 3.2 Die Einstein-Synchronisation

Eine sehr allgemeine Forderung an die Gleichzeitigkeit von Ereignissen ist die Symmetrieforderung: Wenn für einen Beobachter  $A$  das Ereignis  $x_B$  auf der Weltlinie eines Beobachters  $B$  gleichzeitig zu dem Ereignis  $x_A$  auf seiner Weltlinie ist, dann soll umgekehrt auch für den Beobachter  $B$  das Ereignis  $x_A$  gleichzeitig zu  $x_B$  sein. Anhand der Abbildung 5(b) kann man sich leicht überzeugen, dass diese Forderung gleichbedeutend mit der Bedingung

$$\epsilon(A,B) = 1 - \epsilon(B,A)$$

ist.

Eine weitere allgemeine Forderung ergibt sich aus der Homogenität des Raumes. Wir verlangen, dass sich die Funktion  $\epsilon(A,B)$  nicht ändert, wenn wir die beiden Beobachter  $A$  und  $B$  um denselben (räumlichen) Vektor verschieben. Symbolisch ausgedrückt:

$$\epsilon(A + \vec{a}, B + \vec{a}) = \epsilon(A, B) .$$

Dies bedeutet, dass  $\epsilon(A,B)$  nur noch von der räumlichen Richtung abhängen kann, unter der der Beobachter  $B$  von  $A$  aus gesehen wird. Innerhalb einer Halbebene, die die Weltlinie von  $A$  als Rand hat, ist  $\epsilon$  konstant. Auch diese Forderung verlangen wir allgemein.

Nun kommen wir zu einer speziellen Forderung, die sich aus der Konstanz der Lichtgeschwindigkeit ergibt. Da die Lichtgeschwindigkeit für jeden Beobachter und für jede Richtung dieselbe sein soll, können wir zusätzlich noch Isotropie unserer Gleichzeitigkeitsdefinition verlangen.  $\epsilon$  soll also auch nicht mehr von der räumlichen Richtung abhängen, unter der  $B$  von  $A$  aus gesehen wird. In diesem Fall gilt

$$\epsilon(A,B) = \text{const.}$$

Zusammen mit der ersten Forderung der Symmetrie folgt sofort:

$$\epsilon = \frac{1}{2}.$$

Die aus diesem Wert für  $\epsilon$  folgende Synchronisation von Uhren bezeichnet man als Einstein-Synchronisation.

Das zweite Axiom der speziellen Relativitätstheorie - die Konstanz der Lichtgeschwindigkeit - führt uns somit schon zur Festlegung der Gleichzeitigkeitsvorschrift. Das erste Axiom - das Relativitätsprinzip - geht indirekt in diese Festlegung ein, da wir die Konstanz der Lichtgeschwindigkeit und somit die Isotropie der Synchronisationsvorschrift für jedes Inertialsystem verlangt haben.

### 3.3 Synchronisation mit der Ätherhypothese

Wir haben aus der Lorentz-Invarianz der universellen Bewegungsgleichungen geschlossen, dass die spezielle Relativitätstheorie in der Formulierung von Einstein äquivalent zur Lorentz-Theorie ist, d.h. zu einer Theorie mit Ätherhypothese und Lorentz-Kontraktion der Längenskalen. Es sollte daher nicht überraschen, wenn es eine Synchronisationsvorschrift gibt, die auf die Lorentz-Theorie führt.

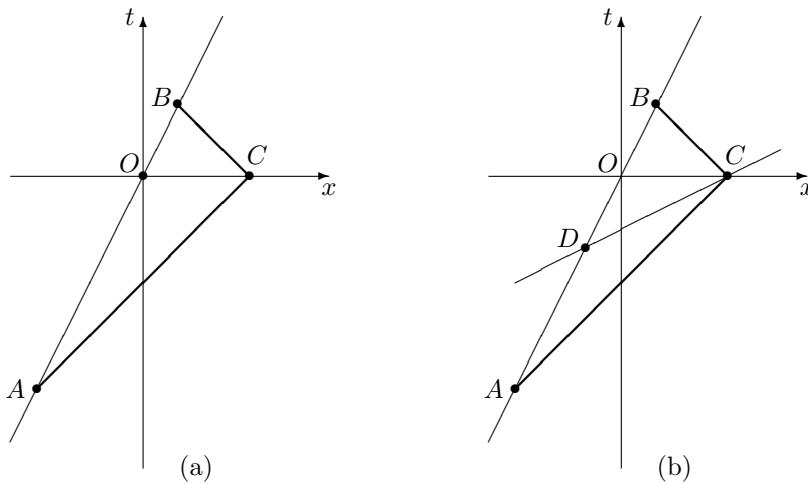


Abbildung 6: Äther- und Einstein-Synchronisation. In Teil (a) rekonstruiert der Beobachter auf der Weltlinie  $AB$  das Ereignis  $O$  als gleichzeitig zum Ereignis  $C$ . Seine Synchronisationsvorschrift hängt von der Geschwindigkeit  $v$  relativ zu dem ausgezeichneten Ruhesystem ab. Dabei gilt für das Verhältnis  $(AO)/(AB) = (1 + v)/2$ . In Abbildung (b) wird nach der Einstein-Synchronisation aus denselben Lichtsignalen das Ereignis  $D$  als zu  $C$  gleichzeitig rekonstruiert.

Auch mit Ätherhypothese verlangen wir von der Gleichzeitigkeit von Ereignissen Symmetrie und Homogenität des Raumes. Was wir jedoch nicht mehr verlangen können, ist die allgemein gültige Isotropie der Synchronisationsvorschrift. Für ein System, das sich relativ zum Äther mit einer Geschwindigkeit  $\vec{v}$  bewegt, wird die Synchronisationsvorschrift von der Richtung relativ zu  $\vec{v}$  abhängen. Allgemein wird also nun gelten:

$$\epsilon(A,B) = \epsilon(\vec{v}; A,B).$$

Lediglich für das ausgezeichnete Inertialsystem, das dem Ruhesystem des Äthers entspricht, ist die Lichtgeschwindigkeit in alle Richtungen dieselbe und somit das Synchronisationsverfahren symmetrisch:

$$\epsilon(\vec{v} = 0; A,B) = \frac{1}{2}.$$

Bewegt sich das Inertialsystem der Beobachter  $A$  und  $B$  relativ zum Äther mit der Geschwindigkeit  $\vec{v}$ , so gilt (vgl. Abb. 6)

$$\epsilon(\vec{v}; A,B) = \frac{1}{2} \left( 1 + \frac{|\vec{v}|}{c} \cos \alpha \right),$$

wobei  $\alpha$  der Relativwinkel zwischen der Richtung von  $\vec{v}$  und der Halbebene ist, die durch die Weltlinie von  $A$  begrenzt wird und die Weltlinie von  $B$  enthält.

Der Vorteil dieses Verfahrens ist, dass Gleichzeitigkeit nun ein absoluter Begriff wird. Sind zwei Ereignisse  $x_A$  und  $x_B$  für einen Beobachter  $A$  gleichzeitig, so sind sie es auch für alle anderen Beobachter, unabhängig von deren Bewegungszustand.

Der Nachteil dieses Verfahrens ist jedoch, dass wir das Ruhesystem des Äthers nicht kennen. Wir müssen also willkürlich ein System auszeichnen, das wir als Ruhesystem definieren, beispielsweise das System des Schwerpunkts der in unserem Universum beobachtbaren Massen oder das Ruhesystem relativ zur Hintergrundstrahlung. (Es ist nicht selbstverständlich, dass diese beiden Systeme identisch sind.) Wegen der Geschwindigkeitsabhängigkeit des Synchronisationsverfahrens gilt das Relativitätsprinzip nicht mehr. Die Lichtgeschwindigkeit ist nur im Ruhesystem des Äthers richtungsunabhängig konstant.

Das Relativitätsprinzip gilt jedoch in einer anderen Form: Wir können jedes beliebige Inertialsystem als Ruhesystem definieren und die Synchronisation der Uhren auf dieses System beziehen. Die physikalischen Gesetze bleiben dieselben. Hier liegt der physikalisch unbefriedigende Aspekt dieses Synchronisierungsverfahrens. Wir brechen die Symmetrie, die sich im Relativitätsprinzip ausdrückt, *per Hand*, indem wir ein Inertialsystem auszeichnen. Umgekehrt trägt die Einstein-Synchronisation dem Relativitätsprinzip Rechnung. Kein Inertialsystem wird durch dieses Verfahren ausgezeichnet. Der Preis ist allerdings die Relativität der Gleichzeitigkeit.

Unter bestimmten Umständen ist es jedoch sinnvoll, eine Synchronisation zu wählen, die der Lorentz'schen Sichtweise näher ist als der Einstein'schen. Beispielsweise könnte man die Lorentz'sche Sichtweise so interpretieren, dass irgendwo in der Mitte unseres Universums eine riesige Uhr steht, und das, was diese Uhr anzeigt, ist die „wahre“ Zeit. Alle anderen Systeme müssen ihre Zeit auf diese wahre Zeit umrechnen. Dieses Verfahren wird in abgewandelter Form beispielsweise bei der Zeitsynchronisation des GPS-Satellitensystems verwendet. Die „riesige Uhr“ steht in der GPS-Zentrale in Colorado. Die Uhren der Satelliten sind so geschaltet, dass ihre Zeit gleich der „Masterzeit“ ist. Insbesondere ist ihr Gang im Vergleich zu einer „richtigen“ Uhr etwas gedrosselt, um den Einfluss der schwächeren Gravitation sowie der Bewegung auszugleichen.

Die Einstein-Synchronisation wie auch die „Äthersynchronisation“ sind Spezialfälle einer Klasse von Synchronisationsvorschriften, bei denen die Konstanz der sogenannten „Zwei-Wege-Lichtgeschwindigkeit“ gefordert wird. Dabei handelt es sich um die Geschwindigkeit  $c$ , die man einem Lichtstrahl zuordnet, der eine Strecke in beide Richtungen - vor und zurück - durchläuft. Ist  $c^+$  die

Geschwindigkeit für eine Richtung und  $c^-$  die Geschwindigkeit für die Rückrichtung, so ist

$$\frac{1}{c} = \frac{1}{2} \left( \frac{1}{c^+} + \frac{1}{c^-} \right).$$

Während jedoch zur Messung von  $c^+$  bzw.  $c^-$  die Uhren an verschiedenen Raumpunkten synchronisiert sein müssen, kann man  $c$  an einem Punkt auswerten, d.h.  $c$  hängt nicht von der Synchronisationsvorschrift ab. Im Michelson-Morley-Versuch beispielsweise wird die Lichtgeschwindigkeit ja gar nicht für verschiedenen Raumrichtungen verglichen, sondern nur die Zwei-Wege-Lichtgeschwindigkeit. Nur von dieser wird gezeigt, dass sie isotrop ist. Die Einstein'sche Forderung der Konstanz der Lichtgeschwindigkeit geht also über das Ergebnis des Michelson-Morley-Versuchs hinaus.

### 3.4 Synchronisation durch langsamen Uhrentransport

Abschließend soll noch gezeigt werden, dass die Synchronisation durch langsamen Uhrentransport zu demselben Ergebnis wie die Einstein-Synchronisation führt. Das bedeutet Folgendes: Synchronisiert man zwei Uhren am selben Ort und bringt dann die eine der beiden Uhren langsam zu einem anderen Ort, ist die so erreichte Synchronisation eine Einstein-Synchronisation.

Zunächst sollten wir etwas genauer definieren, was „Synchronisation durch langsamen Uhrentransport“ bedeutet. Wir stellen uns dazu zwei Beobachter vor, die einen konstanten Abstand  $L$  halten und sich im selben Inertialsystem befinden. Zu einem Zeitpunkt  $t_0 = 0$  werden bei Beobachter 1 zwei Uhren synchronisiert. Anschließend wird eine der beiden Uhren mit der Geschwindigkeit  $v$  zu Beobachter 2 gebracht. Sie benötigt dazu für Beobachter 1 die Zeit  $t = L/v$ , das ist also die Zeit, die Beobachter 1 dem Ereignis „Uhr ist bei Beobachter 2“ zuschreiben wird. Die Anzeige auf der Uhr ist allerdings nach der Relativitätstheorie etwas kürzer, nämlich

$$t_1 = \sqrt{1 - \frac{v^2}{c^2}} t. \quad (11)$$

Wir können nicht einfach argumentieren, dass für  $v/c \ll 1$  die rechte Seite gegen  $t$  geht, denn wenn  $v$  sehr klein ist, wird  $t$  sehr groß, und damit würde zwar  $t_1/t$  gegen 1 gehen, aber  $t_1$  und  $t$  könnten sich möglicherweise selbst im Limes  $v \rightarrow 0$  noch um einen konstanten Term unterscheiden. Es gilt aber

$$t_1 = \left( 1 - \frac{v^2}{c^2} + O((v/c)^4) \right) \frac{L}{v} \approx \frac{L}{v} - \frac{L}{c} \frac{v}{c} + \dots = t - \frac{L}{c} \frac{v}{c} + \dots \quad (12)$$

Da  $L$  und  $c$  Konstanten sind ( $L/c$  ist die Zeit, die das Licht braucht, die Strecke  $L$  zu durchqueren), wird der Korrekturterm für sehr kleine Geschwindigkeiten beliebig klein. Allerdings ist die Korrektur von der Ordnung  $v/c$  und nicht, wie man zunächst nach Gleichung 11 erwarten könnte, von der Ordnung  $(v/c)^2$ .

## Literatur

- [1] G.L. Lamb, Jr.; *Elements of Soliton Theory*; Pure & Applied Mathematics, John Wiley & Sons, 1980.
- [2] Peter Mittelstaedt; *Der Zeitbegriff in der Physik*; BI-Wissenschaftsverlag, 1989.
- [3] Hans Reichenbach; *Philosophie der Raum-Zeit-Lehre*; Hans Reichenbach - Gesammelte Werke Bd. 2; Vieweg-Verlag, Braunschweig; 1977.
- [4] Hans Reichenbach; *Axiomatik der relativistischen Raum-Zeit-Lehre*; in *Die philosophische Bedeutung der Relativitätstheorie*; Hans Reichenbach - Gesammelte Werke Bd. 3; Vieweg-Verlag, Braunschweig, 1977.

# SRT - Effekte

Thomas Filk; Universität Freiburg

In diesem Kapitel sollen einige Effekte beschrieben werden, die sich aus der Lorentz-Invarianz der Minkowski-Raumzeit ergeben. Dazu zählen die Zeitdilatation, die Lorentz-Kontraktion und der relativistische Doppler-Effekt (longitudinal und transversal). Dem sogenannten Zwillings-Paradoxon ist ein eigenes Kapitel gewidmet (Kapitel [Das Zwillingsparadoxon](#)).

Anschließend deuten wir an, wie sich der Lagrange-Formalismus auf die Relativitätstheorie verallgemeinern lässt und was der zum Ortsvektor kanonisch konjugierte Impuls ist. In diesem Zusammenhang definieren wir auch den Begriff der Eigenzeit. Den Abschluss bildet ein von Einstein erdachtes Gedankenexperiment, das eine elegante Herleitung der bekannten Beziehung  $E = mc^2$  erlaubt.

## 1 Zeitdilatation

Wir behandeln zunächst das Phänomen der Zeitdilatation. Schon bei der Pendelkette hatten wir gesehen, dass aus der Lorentz-Invarianz der Feldgleichungen folgt, dass eine relativ zum Äther bewegte Breather-Lösung langsamer schwingt bzw. eine größere Schwingungsperiode hat, als eine Breather-Lösung, die relativ zum Äther ruht. Doch aus dem Relativitätsprinzip sollte folgen, dass ein Beobachter in dem bewegten System umgekehrt ebenfalls den Eindruck hat, dass die Uhren in dem ruhenden System langsamer gehen. Wir wollen nun untersuchen, was genau damit gemeint ist und weshalb dieses Phänomen tatsächlich auftritt.

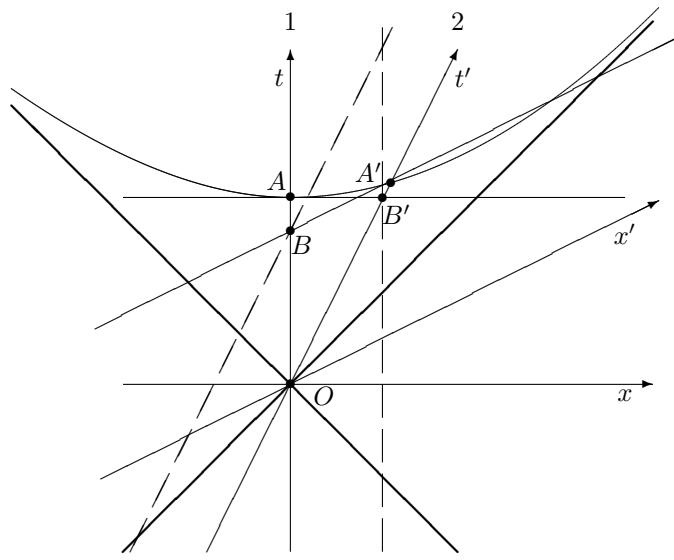


Abbildung 1: Für den ruhenden Beobachter 1 scheint die bewegte Uhr von Beobachter 2 langsamer zu gehen ( $B'$  liegt vor  $A'$ ). Denselben Eindruck hat umgekehrt auch der Beobachter 2 von der Uhr von Beobachter 1 ( $B$  liegt vor  $A$ ). Die gestrichelten Linien beziehen sich auf jeweils einen weiteren Beobachter in System 1 bzw. System 2. Diese Beobachter lesen ihre (mit 1 bzw. 2 synchronisierten) Uhren bei den Ereignissen  $B'$  und  $B$  ab.

Wir betrachten wieder zwei Inertialsysteme mit den jeweiligen Koordinaten  $(x, t)$  und  $(x', t')$  (vgl. Abb. 1). Bei dem Ereignis  $O$  treffen sich die beiden Beobachter im Ursprung ihrer jeweiligen Systeme und synchronisieren ihre Uhren jeweils auf  $t_0 = 0$ . Wir betrachten nun die Situation zunächst aus der Perspektive des Inertialsystems von Beobachter 1. Für diesen Beobachter zeigt die Uhr bei Ereignis  $A$  auf die Zeit  $t$ . In seinem Inertialsystem ist das Ereignis  $B'$  gleichzeitig zu  $A$ , hat also

ebenfalls die Koordinate  $t$ , denn bei diesem Ereignis  $B'$  schneidet die Weltlinie von Beobachter 2 die (in der Zeichnung waagerechte) „Gleichzeitigkeitslinie“ von Beobachter 1 zur Zeitkoordinate  $t$ . Wir wissen jedoch, dass für Beobachter 2 erst im Ereignis  $A'$  dieselbe Zeit vergangen ist wie für Beobachter 1 zum Ereignis  $A$ . Das Ereignis  $A'$  ist aber später als  $B'$ . Das bedeutet, dass die Uhr von Beobachter 2 bei Ereignis  $B'$  noch nicht so viele Zeittakte anzeigt, wie zum Zeitpunkt  $A'$  (nämlich  $t$ ) und damit die Uhr von Beobachter 1 zum Zeitpunkt  $t$ . Beobachter 1 hat also den Eindruck, die Uhr von Beobachter 2 gehe langsamer.

Ich betone hier nochmals, dass Beobachter 1 die Uhr von Beobachter 2 nicht „sieht“ (außer vielleicht im Augenblick, wo sich beide Beobachter treffen, also bei Ereignis  $O$ ). Für den Vergleich der Uhren wählt er für sein Inertialsystem zwei gleichzeitige Ereignisse (z.B.  $A$  und  $B'$ ) und die Zeit auf der Uhr von Beobachter 2 wird von einem anderen Beobachter (dessen Weltlinie parallel zu der von 1 ist aber durch das Ereignis  $B'$  verläuft, dessen Uhr aber mit der von 1 synchronisiert wurde - in Abb. 1 die gestrichelt gezeichnete vertikale Weltlinie) abgelesen. Der Uhrenvergleich erfolgt in Inertialsystem 1 zu vollkommen anderen Ereignissen als in Inertialsystem 2 und der Grund dafür liegt in der unterschiedlichen Zuordnung von Gleichzeitigkeit für Ereignisse.

## 2 Lorentz-Kontraktion

Ähnlich wie im letzten Abschnitt die Zeitdilatation untersuchen wir nun das Phänomen der Lorentz-Kontraktion aus der Sichtweise der verschiedenen Inertialsysteme. Die Länge, beispielsweise eines Lineals, wird dabei als der räumliche Abstand von Anfangs- und Endpunkt des Lineals bestimmt, wobei die Augenblicke der Messung dieses Abstands in den jeweiligen Inertialsystemen gleichzeitig sein sollen.

Wir betrachten zunächst einen Beobachter 1 in dessen Inertialsystem das Lineal ruht (Abb.2 (links)). Die von dem Lineal überstrichene Weltfläche ist in der Abbildung grau unterlegt. Wichtig sind für uns die Weltlinien der beiden Endpunkte des Lineals. Zu einem bestimmten Zeitpunkt in Inertialsystem 1 (beispielsweise  $t = 0$ ) befinden sich die beiden Endpunkte bei den Ereignissen  $O$  und  $A$ . Der räumliche Abstand  $l = OA$  dieser Ereignisse für Beobachter 1 definiert die Länge des Lineals. Im Inertialsystem von Beobachter 2 sind aber die Ereignisse  $O$  (linkes Ende des Lineals) und  $B$  (rechtes Ende) gleichzeitig, in seinem System ist die Länge also durch  $l' = OB$  gegeben. Diese Länge ist aber kürzer als  $l$ , da das Ereignis  $A'$ , das für Inertialsystem 2 von  $O$  denselben räumlichen Abstand hat wie  $A$  für Inertialsystem 1, außerhalb des Lineals liegt.

Wir betrachten nun die umgekehrte Situation: Das Lineal ist in System 2 in Ruhe, seine Anfangs- und Endpunkte bewegen sich also in System 1 mit einer bestimmten Geschwindigkeit (Abb. 2(rechts)). In System 2 wird die Länge des Lineals beispielsweise bei den gleichzeitigen Ereignissen  $O$  und  $A'$  bestimmt, und der zugehörige räumliche Abstand ist  $l = OA'$ . In Inertialsystem 1 wird der Abstand bei den gleichzeitigen Ereignissen  $O$  und  $B'$  gemessen, und deren Abstand ist offensichtlich kleiner als  $l$ .

In beiden Fällen finden wir somit, dass die Länge des Lineals, gemessen von einem bewegten System aus, immer kleiner ist als seine Länge in seinem eigenen Ruhesystem.

## 3 Doppler-Effekte

Der Doppler-Effekt ist schon aus der nicht-relativistischen Mechanik bekannt: Ein Martinshorn klingt höher, wenn das Auto auf uns zukommt, und tiefer, wenn es sich von uns entfernt. Durch die Bewegung des Autos werden die Wellenberge in Fahrtrichtung gestaucht – treffen daher in kürzeren Zeitabständen beim Empfänger ein und klingen höher – und entgegen der Fahrtrichtung gestreckt –

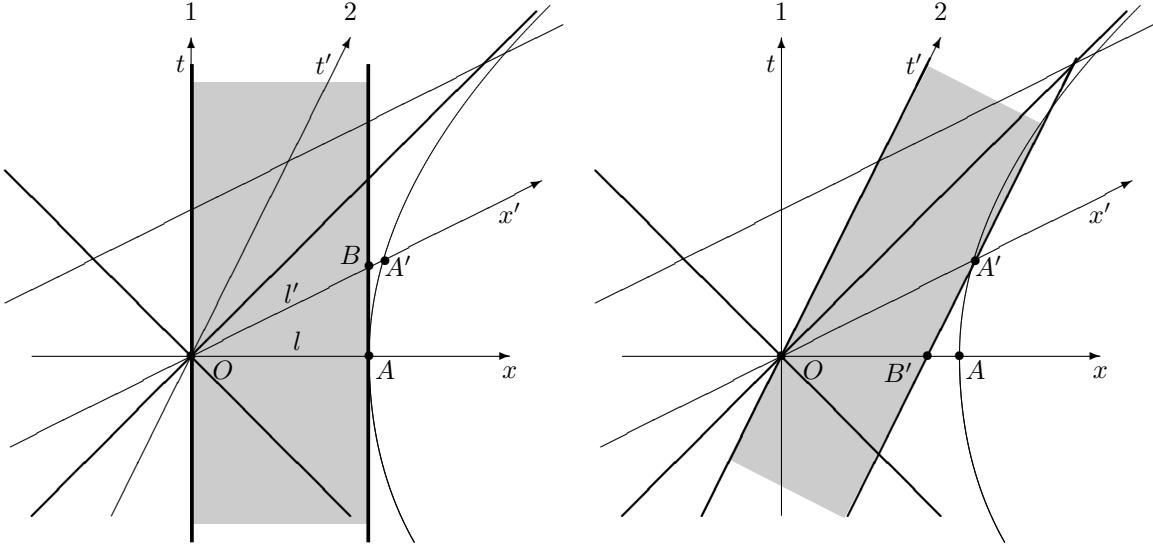


Abbildung 2: Das von einem Lineal überstrichene Raumzeitgebiet (grau unterlegt) erscheint in den verschiedenen Inertialsystemen unterschiedlich breit. (links) Für den bewegten Beobachter 2 scheint das in Inertialsystem 1 ruhende Lineal kürzer. (rechts) Umgekehrt erscheint ein bewegtes Lineal für den ruhenden Beobachter kürzer als für einen Beobachter, der sich mit dem Lineal bewegt.

sie treffen in größeren Zeitabständen beim Empfänger ein und klingen daher tiefer. In der klassischen Mechanik gibt es nur einen *longitudinalen* Doppler-Effekt, d.h. dieser Effekt tritt nur auf, wenn sich der radiale Abstand eines Senders relativ zu einem Empfänger ändert.

In der Relativitätstheorie kommen wegen der Zeitdilatationen bzw. der Lorentz-Kontraktionen in relativ zueinander bewegten Systemen noch weitere Einflüsse hinzu, insbesondere gibt es nun auch den so genannten *transversalen* Doppler-Effekt.

### 3.1 Der Doppler-Effekt in der klassischen Mechanik

Abbildung 3 zeigt eine nicht-relativistische Raumzeit, d.h., die Gleichzeitigkeitslinien sind für alle Beobachter parallel zur  $x$ -Achse und die zeitlichen Abstände zwischen zwei Ereignissen entsprechen den Projektionen auf die  $t$ -Achse. Beobachter E (der Empfänger) sei in Ruhe (bei Schallwellen bedeutet dies im Ruhesystem des Schallträgers, also der Luft), Beobachter S (der Sender) bewegt sich mit konstanter Geschwindigkeit auf E zu, trifft ihn bei Ereignis C und bewegt sich ab dann von E weg. In gleichen Zeitabständen  $\Delta t_S$  sendet Beobachter S bei den Ereignissen A, B, C und D Signale an E. Solange sich der Abstand zwischen dem Sender S und dem Empfänger E mit konstanter Rate verkürzt, empfängt E die Signale in gleichen Zeitabständen  $\Delta t_E$ ; bewegt sich S von E weg, sind die Zeitabstände  $\Delta t'_E$ .

$\Delta t_E$  ist kürzer als  $\Delta t_S$  und zwar um die Zeitdauer, die das Signal braucht, um eine Strecke zurückzulegen, die Beobachter S in der Zeit  $\Delta t_S$  zurücklegt. Um diese Strecke bewegt sich S in der Zeit  $\Delta t_S$  auf E zu, und um diese Strecke ist der Weg für ein Signal kürzer als beim letzten Signal. Die Strecke ist  $\Delta l = v\Delta t_S$ , die Zeit, die das Signal für diese Strecke benötigt, ist  $\Delta T = \Delta l/c$ , also folgt:

$$\Delta t_E = \Delta t_S - \frac{v}{c} \Delta t_S = \left(1 - \frac{v}{c}\right) \Delta t_S. \quad (1)$$

Also ist die Frequenz  $\nu_E$ , mit der E Signale empfängt, um den Faktor  $(1 - \frac{v}{c})^{-1}$  größer als die Frequenz

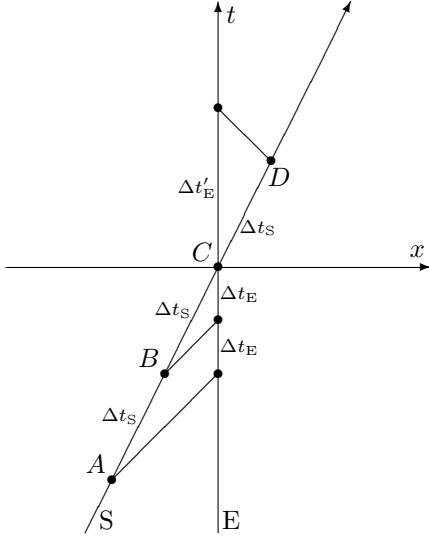


Abbildung 3: Doppler-Effekt in der nicht-relativistischen Mechanik. Die Gleichzeitigkeitslinien sind alle parallel zur  $x$ -Achse. In gleichen Zeitabständen  $\Delta t_S$  (bei den Ereignissen A, B, C, D) sendet Beobachter S Signale an Beobachter E. Solang sich S auf E zubewegt, empfängt E die Signale im Abstand  $\Delta t_E$ , bewegt sich S von E weg, ist der zeitliche Abstand zwischen dem Empfang zweier Signale  $\Delta t'_E$ . Offensichtlich ist  $\Delta t_E$  kürzer als  $\Delta t_S$ , aber  $\Delta t'_E$  länger als  $\Delta t_S$ .

$\nu_S$ , mit der S die Signale abschickt:

$$\nu_E = \frac{1}{(1 - \frac{v}{c})} \nu_S . \quad (2)$$

Entsprechend ist die zugehörige Wellenlänge kürzer:

$$\lambda_E = \left(1 - \frac{v}{c}\right) \lambda_S . \quad (3)$$

Wenn sich S von E entfernt, nimmt der Abstand zwischen den beiden Beobachtern zu und das spätere Signal braucht die Zeit  $\Delta T$  länger als das vorhergehende, um von S zu E zu gelangen:

$$\Delta t'_E = \Delta t_S + \frac{v}{c} \Delta t_S = \left(1 + \frac{v}{c}\right) \Delta t_S . \quad (4)$$

### 3.2 Der longitudinale relativistische Doppler-Effekt

Wie zuvor emittiert ein Sender S in regelmäßigen Zeitabständen  $\Delta\tau_S$  Signale (vgl. Abb. 4). Diese Zeitabstände  $\Delta\tau_S$  beziehen sich nun auf die Eigenzeit des Beobachters. Handelt es sich z.B. dabei um Licht einer bestimmten Frequenz  $\nu_S$ , so bezieht sich diese Frequenz natürlich auf die Eigenzeit in dem System des Senders. Zwei Ereignisse im zeitlichen Abstand  $\Delta\tau_S$  für den Sender haben jedoch im Inertialsystem E des Empfängers eine Zeitdifferenz  $\Delta t_S$ , die um einen Faktor  $\gamma$  größer ist als die Eigenzeit (der „ruhende“ Beobachter sieht die Zeit in einem relativ zu ihm bewegten System langsamer verstreichen):

$$\Delta t_S = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \Delta\tau_S . \quad (5)$$

Abgesehen von diesem Unterschied bleibt die Argumentation dieselbe: In dem System E bewegt sich der Sender im Zeitraum  $\Delta t_S$  um die Strecke  $v \Delta t_S$  und das Lichtsignal benötigt daher bei zwei aufeinanderfolgenden Signalen für das zweite Signal die Zeit  $\Delta T = (\frac{v}{c}) \Delta t_S$  weniger. Insgesamt ergibt sich damit folgende Beziehung:

$$\Delta t_E = \left(1 - \frac{v}{c}\right) \Delta t_S = \frac{\left(1 - \frac{v}{c}\right)}{\sqrt{1 - \frac{v^2}{c^2}}} \Delta\tau_S = \sqrt{\frac{1 - \frac{v}{c}}{1 + \frac{v}{c}}} \Delta\tau_S . \quad (6)$$

Die Vorzeichen für  $v$  drehen sich entsprechend um, wenn sich der Sender vom Empfänger entfernt.

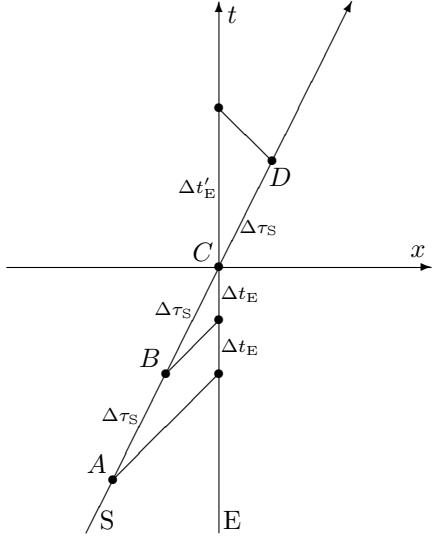


Abbildung 4: Longitudinaler Doppler-Effekt in der relativistischen Mechanik. Die Eigenzeiten  $\Delta\tau_S$  im System des Senders sind nun um einen Faktor  $1/\gamma = \sqrt{1 - \frac{v^2}{c^2}}$  kleiner als der zeitliche Abstand  $\Delta t_S$  derselben Ereignisse im System von Beobachter E. Man beachte, dass die Beziehung der Ereignisse identisch ist, wie im nicht-relativistischen Fall. Geändert hat sich lediglich die Beziehung zwischen der Eigenzeit  $\Delta\tau$  und der entsprechenden Zeit im System des Signalempfängers.

Bewegt sich der Sender auf den Empfänger zu und handelt es sich bei dem ausgetauschten Signal um Licht (was wir beim relativistischen Effekt angenommen haben), so erscheint das Licht für den Empfänger mit einer höheren Frequenz als für den Sender, daher spricht man auch von einer *Blauverschiebung*. Entfernt sich der Sender vom Empfänger, kommt es entsprechend zu einer *Rotverschiebung*.

### 3.3 Der transversale Doppler-Effekt

In der nicht-relativistischen Mechanik gibt es keinen transversalen Doppler-Effekt, da sich der Abstand zwischen Sender und Empfänger nicht ändert und die Zeitdifferenzen für beide Beobachter gleich sind. In der relativistischen Mechanik bleibt bei einer transversalen Bewegung (d.h. der Sender bewegt sich in einem gewissen Abstand senkrecht zum Abstandsvektor) der Abstand ebenfalls konstant, es bleibt aber noch der Faktor der Zeitedilatation. Dieser bewirkt, dass es nun auch zwischen Sender (S) und Empfänger (E) eine Frequenzverschiebung des Lichts gibt:

$$\Delta\nu_S = \sqrt{1 - \frac{v^2}{c^2}} \Delta\nu_S . \quad (7)$$

## 4 „Einparken“

Die Längenkontraktion und die Zeitedilatation geben Anlass zu einer Vielfalt an Scheinparadoxa, die mit der speziellen Relativitätstheorie assoziiert werden. In fast allen Fällen liegt die Ursache der scheinbaren Widersprüche in den unterschiedlichen Ereignismengen, die von den beiden Beobachtern als gleichzeitig empfunden werden.

Das folgende Beispiel bezieht sich auf die Lorentz-Kontraktion und wird in verschiedenen Varianten in der Literatur behandelt. Wir formulieren es hier als ein Problem des Einparkens.

Gegeben sei eine Garage der Länge  $L$  und ein Auto der Länge  $l > L$ . Offensichtlich passt das Auto nicht in die Garage, d.h., wenn die Frontstoßstange die Garagenhinterwand (gegenüber dem Garagentor) berührt, lässt sich das Garagentor nicht schließen. Wenn das Auto aber mit einer genügend großen Geschwindigkeit in die Garage fährt, ist seine Länge kürzer als die Länge der Garage und man sollte das Tor schließen können. Andererseits könnte man sich aber auch in das Inertialsystem des Fahrers versetzen, in dem das Auto in Ruhe ist und sich die Garage mit großer Geschwindigkeit

auf das Auto zubewegt. Nun ist die Garage verkürzt, die Situation ist noch ungünstiger und das Tor sollte sich erst Recht nicht schließen lassen. Ob aber ein Garagentor geschlossen werden kann oder nicht ist eine physikalische Tatsache und kann nicht vom Inertialsystem eines Beobachters abhängen.

Was passiert im Ruhesystem der Garage, wenn das Auto mit hoher Geschwindigkeit hereinfährt? Tatsächlich ist das Auto im Augenblick der Einfahrt kürzer und passt in die Garage – das Garagentor kann geschlossen werden. Doch nun wird das Auto abgebremst und dehnt sich aus, dabei stößt es vorne und hinten gegen die Garagenwand bzw. das Garagentor und wird physikalisch gestaucht.

Wie erfährt der Autofahrer dieselbe Situation? Für ihn ist die Garage wesentlich kürzer als das Auto. Wenn er mit seiner Frontstange gegen die Garagenwand fährt, ist der hintere Teil des Wagens noch weit außerhalb der Garage. Doch wenn der Wagen bezüglich des Systems des Garages vorne und hinten gleichzeitig abgebremst wird, wird er im Ruhesystem des Autofahrers von vorne abgebremst. D.h., der Wagen fährt vorne gegen die Garagenwand und wird dadurch gestaucht, während der hintere Teil des Wagens sich weiter nach vorne bewegt und schließlich ebenfalls ganz in der Garage ist, sodass das Tor geschlossen werden kann. Erst dann erreicht die Stauchung des Wagens auch den hinteren Teil.

In beiden Fällen fährt der Wagen in die Garage und das Tor kann geschlossen werden, aber der Wagen wurde durch das Abbremsen bzw. die Garagenwände derart gestaucht, dass er nicht mehr seine ursprüngliche Länge hat.

Eine wichtige Erkenntnis können wir aus diesem Beispiel festhalten: Es gibt keinen idealen starren Körper! Darunter würde man einen Körper verstehen, der jede Beeinflussung (z.B. Verschiebung) an seinem einen Ende instantan auf den gesamten Körper überträgt. Aufgrund der endlichen Ausbreitungsgeschwindigkeit von Licht kann sich auch in einem Körper kein Signal mit einer größeren Geschwindigkeit ausbreiten. Ein Stoß auf der einen Seite führt notwendigerweise zu einer Stoßwelle, die sich nicht schneller als mit Lichtgeschwindigkeit durch den Körper ausbreitet und daher jede Einwirkung an einem Ende verzögert zu dem anderen Ende überträgt.

## 5 Das Zwillingsparadoxon

Dem Zwillingsparadoxon ist ein eigenes Kapitel gewidmet (siehe Kapitel [Das Zwillingsparadoxon](#)), daher wird es hier nur sehr kurz behandelt. Das Zwillingsparadoxon bezeichnet das folgende Phänomen: Wenn sich zwei Zwillinge (die nach unserer Vorstellung immer dasselbe Alter haben) treffen, und einer von beiden auf der Erde verbleibt während der andere mit großer Geschwindigkeit in den Weltraum hinausfliegt und nach vielen Jahren zurückkommt, dann haben die beiden nicht mehr dasselbe biologische Alter. Der auf der Erde verbliebene Zwilling ist älter als derjenige, der durch den Weltraum gereist ist.

Betrachten wir dazu die Weltlinien der beiden Zwillinge. Bis Ereignis  $A_0 = B_0$  haben beide dieselbe Weltlinie. Dann kommt es zur Trennung. Während Zwilling 1 seinen bisherigen Bewegungszustand beibehält (und die Ereignisse  $A_1, A_2, A_3$  durchläuft), bewegt sich Zwilling 2 sehr rasch zu Ereignis  $B_2$ , dort bremst er ab und beschleunigt in die umgekehrte Richtung, fliegt also wieder auf seinen Zwillingspartner zu. Bei  $A_4 = B_4$  treffen sich die beiden Zwillinge wieder.

Die Zeitspannen lassen sich leicht berechnen: Die Zeitspanne für Zwilling 1 von  $A_0$  bis  $A_1$  ist genauso lang wie die für Zwilling 2 von  $A_0$  bis  $B_2$ . Entsprechend ist die Zeitspanne  $A_3A_4$  dieselbe wie die von Ereignis  $B_2$  bis  $B_4$  für Zwilling 2. Insgesamt hat die Weltlinie von Zwilling 2 von Ereignis  $B_0$  bis  $B_4$  also dieselbe Länge wie die Summe der beiden Abschnitte  $A_0A_1$  und  $A_3A_4$  für Zwilling 1. Die Zwischenzeit - von  $A_1$  bis  $A_3$  - ist die Zeitspanne, um die Zwilling 1 älter ist.

Man könnte auf die Idee kommen, dass das unterschiedliche Alter der beiden Zwillinge darauf

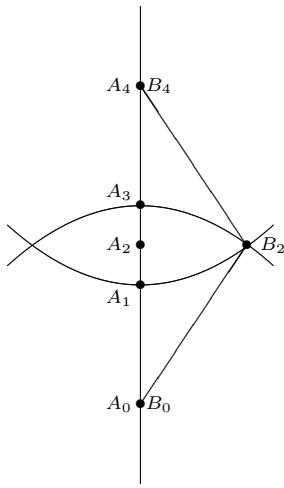


Abbildung 5: Zum Zwillingsparadoxon: Die Weltlinie von Zwilling 1 verläuft entlang der Ereignisse  $A_0 = B_0, A_1, A_2, A_3, A_4 = B_4$ , die von Zwilling 2 entlang  $B_0, B_2, B_4$ . Die Weltlinie von Zwilling 1 ist länger als die von Zwilling 2, d.h., Zwilling 1 ist bei der Wiedervereinigung in Ereignis  $A_4 = B_4$  älter als sein Bruder. Bei  $B_2$  hat Zwilling 2 dasselbe Alter wie Zwilling 1 bei  $A_1$ . Insgesamt ist Zwilling 1 um die Zeitspanne zwischen  $A_1$  und  $A_3$  älter.

zurückzuführen ist, dass Zwilling 2 mehrfach beschleunigt wurde, insbesondere auch bei Ereignis  $B_2$ . Um zu verdeutlichen, dass dieses Argument nicht richtig ist, betrachten wir Abb. 3. Die Weltlinien der drei Drillinge (der erste bleibt in Ruhe, der zweite macht eine kurze Reise über Ereignis  $C_2$ , der dritte macht die große Reise über  $B_2$ ) sind unterschiedlich lang. Drilling 1 ist am meisten gealtert, Drilling 2 etwas weniger und Drilling 3 noch weniger. Insbesondere haben Drilling 2 und Drilling 3 dieselben Beschleunigungsphasen erlebt, sind aber trotzdem bei  $A_4 = B_4$  unterschiedlich alt. Es handelt sich um einen Effekt, der auf der Geometrie der Minkowski-Raumzeit beruht und nicht auf unterschiedlichen Beschleunigungsphasen.

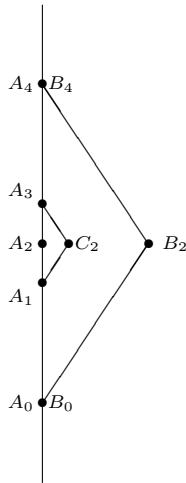


Abbildung 6: Erweiterung des Zwillingsparadoxons für Drillinge. Die drei Weltlinien –  $(A_0 \rightarrow A_1 \rightarrow A_2 \rightarrow A_3 \rightarrow A_4)$  für Drilling 1,  $(A_0 \rightarrow A_1 \rightarrow C_2 \rightarrow A_3 \rightarrow A_4)$  für Drilling 2 und  $(B_0 \rightarrow B_2 \rightarrow B_4)$  für Drilling 3 – sind unterschiedlich lang. Insbesondere ist die Weltlinie von Drilling 3 kürzer als die von Drilling 2, obwohl beide dieselben Beschleunigungsphasen erlebt haben.

## 6 Die Eigenzeit

Wie wir gesehen haben, können ganz allgemein Uhren, die entlang unterschiedlicher Weltlinien transportiert wurden, unterschiedliche Zeitdauern anzeigen, selbst wenn die Weltlinien dieselben Ereignisse verbinden. Im letzten Abschnitt hat es sich um stückweise gerade Weltlinien gehandelt, bei denen wir zur Bestimmen der Zeitdauer die Anteile der einzelnen Teilstücke addiert haben. Dies können wir für

stetige Weltlinien verallgemeinern.

Für zwei zeitartige Ereignisse  $B$  und  $A$  ( $B$  zeitlich nach  $A$ ), mit den Differenzkoordinaten  $\Delta t = t_B - t_A$  und  $\Delta \vec{x} = \vec{x}_B - \vec{x}_A$  in einem beliebigen Inertialsystem, definieren wir die *Eigenzeit*  $\tau_{AB}$  durch

$$\tau_{AB}^2 = (\Delta t)^2 - \frac{1}{c^2}(\Delta \vec{x})^2. \quad (8)$$

$\tau_{AB}$  ist die Zeitdifferenz zwischen den beiden Ereignissen, die von einer Uhr angezeigt wird, die mit konstanter Geschwindigkeit (also in einem Inertialsystem) von Ereignis  $A$  zu Ereignis  $B$  transportiert wird. In dem Inertialsystem dieser Uhr finden beide Ereignisse im räumlichen Koordinatenursprung statt. Für zwei Ereignisse ist  $\tau$  eine Invariante und es spielt keine Rolle, in welchem Inertialsystem  $\tau$  nach Gl. 8 berechnet wird, da die rechte Seite dieser Gleichung invariant unter Lorentz-Transformationen ist (siehe Kapitel [Grundlagen der SRT](#)).

Handelt es sich um infinitesimal benachbartheitbare Ereignisse, deren Raum- und Zeitkoordinaten sich in einem beliebigen Inertialsystem um  $d\vec{x}$  und  $dt$  unterscheiden, erhalten wir

$$d\tau = \sqrt{1 - \frac{1}{c^2} \left( \frac{d\vec{x}}{dt} \right)^2} dt = \sqrt{1 - \frac{v(t)^2}{c^2}} dt \quad (9)$$

als die Eigenzeit zwischen den beiden Ereignissen. Wir können nun einer beliebigen Weltlinie  $\gamma$  (die natürlich an jedem ihrer Punkte eine zeitartige Tangente haben muss) zwischen zwei Ereignissen  $A$  und  $B$  eine Eigenzeit zuordnen:

$$\tau(\gamma) = \int_{t_B; \gamma}^{t_A} \sqrt{1 - \frac{v(t)^2}{c^2}} dt. \quad (10)$$

Diese Zeit wird von einer Uhr angezeigt, die entlang der Weltlinie  $\gamma$  von  $A$  nach  $B$  transportiert wird. Wie schon erwähnt spielt es dabei keine Rolle, in welchem Inertialsystem die momentane Geschwindigkeit  $v(t)$  gemessen und das Integral ausgewertet wird. Die Eigenzeit zwischen zwei Ereignissen entlang eines Weges  $\gamma$  ist am größten, wenn es sich bei  $\gamma$  um eine gerade Verbindungsgeraden handelt. Wir werden später in der Allgemeinen Relativitätstheorie nicht mehr von Geraden sprechen können, wohl aber von geodätischen Verbindungswegen. Diese zeichnen sich durch eine maximale Eigenzeit aus.

## 7 Der kanonische Formalismus

In der klassischen, Newton'schen Mechanik hat sich der Lagrange-Formalismus als sehr nützlich erwiesen. Auch in der speziellen Relativitätstheorie lassen sich eine Lagrange-Funktion und eine zugehörige Wirkung angeben, aus der nicht nur die Bewegungsgleichungen sondern auch die kanonisch konjugierten Variablen folgen.

Die Wirkung ordnet jeder Trajektorie  $t \mapsto \vec{x}(t)$  eine Zahl zu. Ein großer Vorteil des Lagrange-Formalismus ist, dass diese Zahl nicht von der Wahl der Koordinaten abhängt. Daher sollte diese Zahl auch für die relativistische Wirkung eine Invariante sein. Statt diese Invariante abzuleiten, geben wir die Lagrange-Funktion und ihre Wirkung einfach an und zeigen, dass sie im klassischen Grenzfall mit den bekannten Größen übereinstimmen, aber relativistisch invariant sind. (Durch diese beiden Forderungen sind die Größen im Allgemeinen bereits festgelegt.)

Als Lagrange-Funktion eines freien Teilchens definieren wir für ein beliebiges Inertialsystem mit Koordinaten  $(t, \vec{x})$  (und der Geschwindigkeit  $\vec{v}(t) = \dot{\vec{x}} = \frac{d\vec{x}(t)}{dt}$ ):

$$L = -mc^2 \sqrt{1 - \frac{\vec{v}(t)^2}{c^2}}, \quad (11)$$

und die zugehörige Wirkung ist

$$S = -mc^2 \int_{t_0}^{t_1} \sqrt{1 - \frac{\vec{v}(t)^2}{c^2}} dt. \quad (12)$$

Ausgedrückt durch die Eigenzeit entlang einer Bahnkurve ergibt sich für die Wirkung:

$$S = -mc^2 \int_{\tau_0}^{\tau_1} d\tau. \quad (13)$$

Zu integrieren ist jeweils entlang der Bahnkurve. Aus der letzteren Darstellung wird die Invarianz der Wirkung offensichtlich, d.h.  $S$  ist für jedes Inertialsystem gleich.

Entwickeln wir in einem festen Inertialsystem die Quadratwurzel, so erhalten wir für die Lagrange-Funktion in führender Ordnung:

$$L = -mc^2 \left( 1 - \frac{1}{2} \frac{\vec{v}(t)^2}{c^2} + \dots \right) \approx -mc^2 + \frac{1}{2} m\vec{v}(t)^2, \quad (14)$$

also bis auf den konstanten Term  $-mc^2$  (der, wie wir gleich sehen werden, der Ruheenergie eines Teilchens entspricht) die klassische Wirkung.

Die zu  $\vec{x}$  kanonisch konjugierte Variable  $\vec{p}$  ist definiert als

$$p_i(t) = \frac{\partial L}{\partial v_i(t)} = \frac{mv_i(t)}{\sqrt{1 - \frac{\vec{v}(t)^2}{c^2}}} \quad \text{bzw.} \quad \vec{p}(t) = \frac{m\vec{v}(t)}{\sqrt{1 - \frac{\vec{v}(t)^2}{c^2}}}. \quad (15)$$

Entsprechend ist die Energie (die eine Erhaltungsgröße ist, da  $L$  nicht explizit von der Zeit abhängt) definiert als:

$$E = \vec{v} \cdot \vec{p} - L(\vec{v}), \quad (16)$$

woraus wir erhalten:

$$E = \frac{m\vec{v}(t)^2}{\sqrt{1 - \frac{\vec{v}^2}{c^2}}} + \frac{mc^2}{\sqrt{1 - \frac{\vec{v}^2}{c^2}}} \left( 1 - \frac{\vec{v}^2}{c^2} \right) = \frac{mc^2}{\sqrt{1 - \frac{\vec{v}^2}{c^2}}}. \quad (17)$$

In einem kanonischen Formalismus sollte die Energie allerdings als Funktion von Ort und kanonisch konjugiertem Impuls aufgefasst werden. Wir lösen daher Gl. 15 nach  $\vec{v}^2$  auf und erhalten schließlich:

$$E = \sqrt{m^2 c^4 + \vec{p}^2 c^2}. \quad (18)$$

## 8 Die Äquivalenz von Masse und Energie

Eine weitere Folgerung aus der speziellen Relativitätstheorie ist nicht nur besonders bekannt geworden sondern auch von fundamentaler Bedeutung, nämlich die Äquivalenz von träger Masse und Energie.

Allgemein schreibt man die Formel

$$E = mc^2 \quad (19)$$

Einstein zu, der sie 1905 in der Arbeit *Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?* ([1]) für den Sonderfall eines strahlenden Körpers hergeleitet und ihre allgemeine Richtigkeit vermutet hatte. Abgesehen davon, dass die Herleitung dieser Formel als falsch gilt (siehe [3]; Farbtafel XXIV, Zitat aus Jammer), findet man Überlegungen über eine Trägheit von Energie auch schon bei Poincaré 1900.

Die folgende „Elementare Ableitung der Äquivalenz von Masse und Energie“ stammt ebenfalls von Einstein aus dem Jahre 1946, und ist in seinem Buch *Aus meinen späten Jahren* ([2], S. 121) entnommen.

Die vorstehende Ableitung des Äquivalenzgesetzes, die bisher nicht publiziert ist, hat zwei Vorteile. Sie bedient sich zwar des speziellen Relativitätsprinzips, setzt aber die technisch-formalen Hilfsmittel der Theorie nicht voraus, sondern bedient sich nur dreier vorbekannter Gesetzmäßigkeiten:

1. des Satzes von der Erhaltung des Impulses,
2. des Ausdrucks für den Strahlungsdruck beziehungsweise für den Impuls eines in bestimmter Richtung sich ausbreitenden Strahlungs-Komplexes,
3. der wohlbekannte Ausdruck für die Aberration des Lichts (Einfluß der Bewegung der Erde auf den scheinbaren Ort der Fixsterne (Bradley)).

Wir fassen nun folgendes System ins Auge. Bezuglich eines Koordinatensystems  $K_0$  (Abb. 7) schwebt der Körper  $B$  frei im Raum. Zwei Strahlungskomplexe  $S$ ,  $S'$ , je von der Energie  $\frac{E}{2}$  breiten sich längs der positiven bzw. negativen  $x_0$ -Richtung aus und werden dann von  $B$  absorbiert. Bei der Absorption wächst die Energie von  $B$  um  $E$ . Der Körper  $B$  bleibt bei diesem Prozess aus Symmetrie-Gründen in Ruhe.

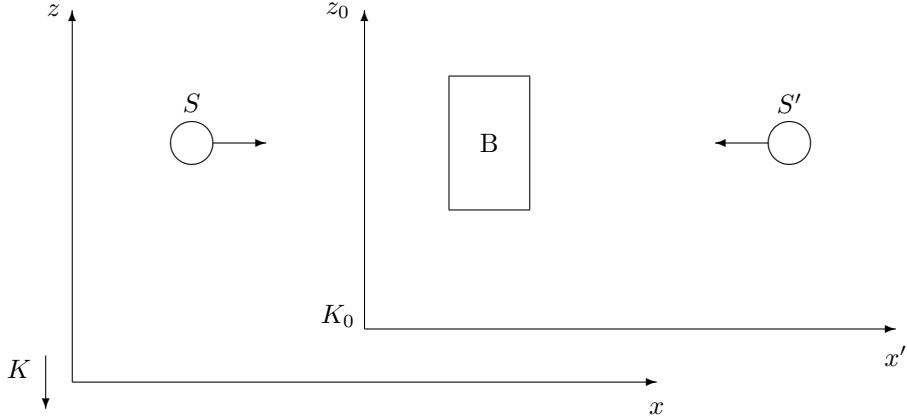


Abbildung 7: Das System  $B$  absorbiert die Energie von  $S$  und  $S'$  - aus dem Inertialsystem von  $B$  aus betrachtet.

Nun betrachten wir diesen selben Prozess von einem System  $K$  aus, welches sich gegenüber  $K_0$  mit der konstanten Geschwindigkeit  $v$  in der negativen  $z_0$ -Richtung bewegt. In bezug auf  $K$  ist dann die Beschreibung des Vorganges [folgende] (siehe Abb. 8).

Der Körper  $B$  bewegt sich in der positiven  $z$ -Richtung mit der Geschwindigkeit  $v$ . Die beiden Lichtkomplexe haben in bezug auf  $K$  eine Fortpflanzungsrichtung, welche einen Winkel  $\alpha$  mit der  $x$ -Achse bildet. Das Aberrationsgesetz besagt, dass in erster Näherung  $\alpha = \frac{v}{c}$  ist, wobei  $c$  die Lichtgeschwindigkeit bedeutet. Aus der Betrachtung in bezug auf  $K_0$  wissen wir, dass die Geschwindigkeit  $v$  von  $B$  durch die Absorption von  $S$  und  $S'$  keine Änderung erfährt.

Nun wenden wir auf den Prozess in bezug auf  $K$  das Gesetz von der Erhaltung des Impulses in bezug auf die Richtung  $z$  des betrachteten Gesamtsystems an.

I. Vor der Absorption sei  $M$  die Masse von  $B$ ;  $Mv$  ist dann der Ausdruck des Impulses von  $B$  (gemäß der klassischen Mechanik). Jeder der Strahlungskomplexe hat die Energie  $\frac{E}{2}$  und deshalb gemäß einer wohlbekannten Folgerung aus Maxwells Theorie den Impuls

$$\frac{E}{2c} . \quad (20)$$

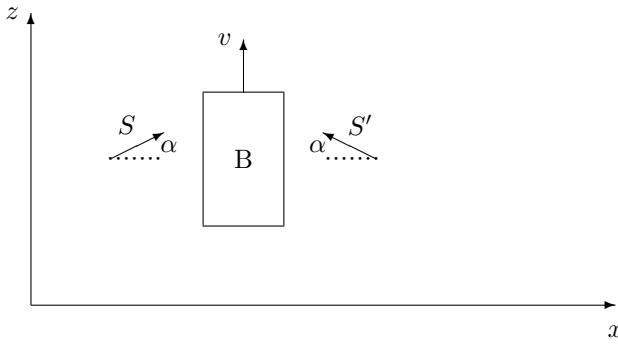


Abbildung 8: Dieselbe Situation wie in Abb. 7, allerdings aus dem bewegten System  $K$  heraus betrachtet.

Dies ist streng genommen zunächst der Impuls von  $S$  in bezug auf  $K_0$ . Wenn aber  $v$  klein ist gegen  $c$ , so muss der Impuls in bezug auf  $K$  bis auf die Größe von zweiter Ordnung  $\frac{v^2}{c^2}$  dieselbe sein. Von diesem Impuls fällt in die  $z$ -Richtung die Komponente  $\frac{E}{2c} \sin \alpha$ , sind aber genügend genau (bis auf Größen höherer Ordnung)  $\frac{E}{2c} \alpha$  oder  $\frac{E}{2} \cdot \frac{v}{c^2}$ .  $S$  und  $S'$  zusammen haben also in der  $z$ -Richtung den Impuls  $E \frac{v}{c^2}$ . Der Gesamtimpuls des Systems vor der Absorption ist also

$$Mv + \frac{E}{c^2}v . \quad (21)$$

II. Nach der Absorption sei  $M'$  die Masse von  $B$ . Wir antizipieren hier die Möglichkeit, dass die Masse bei der Aufnahme der Energie  $E$  eine Zunahme erfahren könnte (dies ist nötig, damit das Endresultat unserer Überlegungen widerspruchsfrei sei). Der Impuls des Systems nach der Absorption ist dann

$$M'v . \quad (22)$$

Nun setzen wir den Satz von der Erhaltung des Impulses als richtig voraus und wenden ihn in bezug auf die  $z$ -Richtung an. Dies ergibt die Gleichung

$$Mv + \frac{E}{c^2}v = M'v \quad (23)$$

oder

$$M' - M = \frac{E}{c^2} . \quad (24)$$

Diese Gleichung drückt den Satz der Äquivalenz von Energie und Masse aus. Der Energiezuwachs  $E$  ist mit dem Massenzuwachs  $\frac{E}{c^2}$  verbunden. Da die Energie ihrer üblichen Definition gemäß eine additive Konstante freilässt, so können wir nach der Wahl der letzteren stattdessen auch kürzer schreiben

$$E = Mc^2 . \quad (25)$$

## Literatur

- [1] Albert Einstein; *Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?* (Ann. Phys., Leipzig, 18 (1905) 639.
- [2] Albert Einstein; *Aus meinen späten Jahren*; Ullstein Sachbuch, Verlag Ullstein, Frankfurt, Berlin, 1993.
- [3] Károly Simonyi; *Kulturgeschichte der Physik*; Verlag Harri Deutsch, Thun, Frankfurt am Main, 1990.

# Das Zwillingsparadoxon

Thomas Filk; Universität Freiburg

Das Zwillingsparadoxon gehört zu den bekanntesten Gedankenexperimenten der Physik, wobei dieser Effekt in abgewandelter Form auch experimentell überprüft bzw. beobachtet wurde. Die Idee: Eine Person bleibt hier auf der Erde (die wir näherungsweise als Inertialsystem ansehen), eine zweite, gleich alte Person - daher betrachtet man gerne Zwillinge - macht mit einem Raumschiff eine ausgedehnte Weltraumreise mit nahezu Lichtgeschwindigkeit. Nach einiger Zeit treffen die beiden Personen wieder zusammen und es stellt sich heraus, dass die Person, die auf der Erde geblieben ist, deutlich mehr gealtert ist als die Person, die den Weltraumflug gemacht hat.

An dieser Situation sind zwei Dinge paradox: (1) Wie können Zwillinge sich in ihrem Alter um viele Jahre unterscheiden? - hier liefert die Relativitätstheorie eine unmittelbare Erklärung - und (2) Was unterscheidet das Bezugssystem der Person, die auf der Erde geblieben ist, von dem Bezugssystem der Person, die die Weltraumreise gemacht hat? Es muss ja einen physikalischen Grund geben, welche der beiden Personen jünger geblieben ist. Befänden sich z.B. beide in einem Inertialsystem, müsste die Physik auch für beide Personen dieselbe sein. Mindestens eine der beiden Personen befand sich also nicht in einem Inertialsystem; dies ist die Person, welche die Weltraumreise gemacht hat und irgendwann mindestens einmal beschleunigen musste, um wieder an den Ausgangsort zurückkommen zu können. Gelegentlich heißt es deshalb, die Beschleunigung sei dafür verantwortlich, dass die eine Person jünger geblieben ist. Diese Aussage - insbesondere die Sprechweise „dafür verantwortlich sein“ - ist irreführend und führt zu vielen Fehlvorstellungen: Die Beschleunigung hat keinen kausalen Einfluss auf den Gang von Uhren, aber sie unterscheidet ein Nicht-Inertialsystem von einem Inertialsystem.

In diesem Kapitel klären wir zunächst den Begriff der Eigenzeit und wie man die Eigenzeit entlang einer Weltlinie berechnen kann. Als Weltlinie bezeichnen wir eine Trajektorie  $\mathbf{x}(t)$  in einer vierdimensionalen Raumzeit bzw. den Graph der Abbildung  $t \mapsto \mathbf{x}(t)$ . Anschließend wird das Zwillingsparadoxon nochmals genauer beschrieben und erörtert. Begriffe wie Koordinatensystem, Bezugssystem, Inertialsystem sowie auch das Konzept der Einstein-Synchronisation von Uhren werden vorausgesetzt.

## 1 Die Eigenzeit entlang einer Weltlinie

Eines der Grundpostulate der Speziellen Relativitätstheorie ist die Aussage, dass die Lichtgeschwindigkeit für einen inertialen Beobachter unabhängig vom Bewegungszustand der Lichtquelle immer denselben Wert hat. Seien  $A$  und  $B$  zwei Ereignisse, die durch einen Lichtstrahl verbunden werden können (z.B. ( $A$ ) der Moment, in dem ein Lichtstrahl durch eine Blende tritt und ( $B$ ) der Moment, in dem dieser Lichtstrahl an einem Spiegel reflektiert wird). Zwei solche Ereignisse bezeichnet man als *lichtartig*. Dann gilt für jeden inertialen Beobachter:

$$\frac{|\Delta\mathbf{x}|}{\Delta t} = c \quad \text{oder} \quad (\Delta\mathbf{x})^2 - c^2(\Delta t)^2 = 0. \quad (1)$$

Hierbei sind  $\Delta\mathbf{x}$  bzw.  $\Delta t$  der in einem inertialen Bezugssystem gemessene Abstand (genauer der Differenzvektor, dessen Betrag der Abstand ist) bzw. die Zeitdauer zwischen den beiden Ereignissen  $A$  und  $B$ .  $c$  bezeichnet die Lichtgeschwindigkeit im Vakuum und ist somit eine universelle Konstante. Aus dieser Gleichung wird oft gefolgert, dass

$$(\Delta\mathbf{x})^2 - c^2(\Delta t)^2 = (\Delta\mathbf{x}')^2 - c^2(\Delta t')^2, \quad (2)$$

wobei sich nun  $\Delta\mathbf{x}$  bzw.  $\Delta t$  auf die räumliche Differenz und die Zeitspanne zwischen zwei beliebigen Ereignissen  $A$  und  $B$  bezieht, also nicht unbedingt nur lichtartige Ereignisse.  $\Delta\mathbf{x}'$  und  $\Delta t'$  bezeichnen den entsprechenden Differenzvektor bzw. die Zeitspanne zwischen denselben Ereignissen in einem anderen inertialen Bezugssystem.

Dieser Schluss ist allerdings zu begründen, denn ganz allgemein kann man natürlich aus der Gleichheit zweier Ausdrücke an der Stelle 0 nicht auf die Gleichheit allgemein schließen und was für lichtartige Ereignisse gilt, muss nicht unbedingt für beliebige Ereignisse gelten. Die wesentlichen weiteren Annahmen, die hier eingehen sind: (1) der Übergang für räumliche und zeitliche Abstände von einem inertialen Bezugssystem in ein anderes inertiales Bezugssystem wird durch eine lineare Transformation beschrieben, und (2) unser Raum ist homogen und isotrop.

Die erste Annahme, die Linearität der Transformation, bedeutet, dass sich die linke und die rechte Seite von Gleichung 2 nur um einen Faktor unterscheiden können. Die Annahme begründet sich daraus, dass unter den Transformationen Inertialsysteme wieder in Inertialsysteme übergehen sollen, sodass in einem Raum-Zeit-Diagramm gerade Linien wieder in gerade Linien transformiert werden. Aus der Isotropie und Homogenität des Raumes folgt dann, dass dieser Faktor 1 sein muss.<sup>1</sup>

Seien nun  $A$  und  $B$  zwei Ereignisse, sodass das Ereignis  $B$  durch das Ereignis  $A$  kausal beeinflusst werden kann. Das bedeutet:

$$c^2(\Delta t)^2 > (\Delta\mathbf{x})^2. \quad (3)$$

Gilt diese Eigenschaft für zwei Ereignisse in einem Inertialsystem, dann gilt sie in allen Inertialsystemen. Man bezeichnet die Ereignisse  $A$  und  $B$  in diesem Fall als *zeitartig* bzw. genauer als *relativ zeitartig zueinander*. (Die Begriffe lichtartig, zeitartig und raumartig (s.u.) kennzeichnen immer eine Relation zwischen zwei Ereignissen; für ein einzelnes Ereignis sind diese Begriffe sinnlos.) Zu zwei zeitartigen Ereignissen gibt es immer ein Inertialsystem, sodass diese beiden Ereignisse am Ort  $x = 0$  (oder allgemeiner am selben Ort) stattfinden. Für dieses Inertialsystem ist  $\Delta\mathbf{x} = 0$  und somit wird der Abstand zwischen diesen beiden Ereignissen ausschließlich durch die Zeitspanne  $\Delta\tau$  zwischen ihnen in diesem ausgezeichneten Inertialsystem charakterisiert. Diese spezielle Zeitspanne bezeichnet man als die *Eigenzeit* des Streckenabschnitts der inertialen Weltlinie, die die beiden Ereignisse verbindet.

Man beachte, dass die Eigenzeit nicht einfach zwei Ereignissen, sondern einer Weltlinie, die die Ereignisse verbindet, zugeschrieben wird. Handelt es sich bei dieser Weltlinie um die Weltlinie des Ursprungs eines Inertialsystems, wie oben angenommen, bezeichnet man diese Eigenzeit auch als Abstand der beiden Ereignisse. Handelt es sich bei den Ereignissen um *raumartige* Ereignisse, gilt  $c^2(\Delta t)^2 < (\Delta\mathbf{x})^2$ . In diesem Fall gibt es immer ein Inertialsystem, in dem die beiden Ereignisse gleichzeitig stattfinden, d.h. in diesem Inertialsystem ist  $\Delta t = 0$  und der räumliche Abstand der Ereignisse ist  $|\Delta\mathbf{x}|$ . Man beachte, dass diese Zuordnungen - Eigenzeiten zu zwei zeitartigen Ereignissen und Abstand zu raumartigen Ereignissen - ein spezielles Inertialsystem voraussetzen. Andererseits ist beispielsweise die Eigenzeit zwischen zwei Ereignissen  $A$  und  $B$  eine Lorentz-Invariante: Alle Beobachter sind sich darin einig, dass eine (ideale) Uhr, die sich im Ursprung eines Inertialsystems befindet und deren Weltlinie die Ereignisse  $A$  und  $B$  verbindet, diese Eigenzeit anzeigt.

Wir können den Begriff der Eigenzeit von einer geraden Strecke in einem Inertialsystem zu einer beliebigen Weltlinie verallgemeinern. Sei  $\mathbf{x}(t)$  die Weltlinie eines Beobachters. Wir können diese Weltlinie in kleine Abschnitte unterteilen, die geradlinig verbunden werden. Die Weltlinie wird also

---

<sup>1</sup>Die inverse Transformation muss wegen der Isotropie denselben Faktor haben wie die Transformation selbst, daher kommt für den Faktor nur noch  $\pm 1$  in Frage. Da man aber (in mehr als einer Raumdimension) eine Transformation stetig in ihr Inverses überführen kann, muss der Faktor  $+1$  sein.

durch einen Polygonzug angenähert. Für jede infinitesimale Teilstrecke ist

$$(\Delta\tau(t))^2 = (\Delta t)^2 - \frac{1}{c^2}(\Delta\mathbf{x})^2 \quad \text{bzw.} \quad \Delta\tau(t) = \sqrt{1 - \frac{1}{c^2} \left( \frac{\Delta\mathbf{x}}{\Delta t} \right)^2} \Delta t \quad (4)$$

die Eigenzeit. Hierbei sind  $\Delta t$  und  $\Delta\mathbf{x}$  die Zeitdifferenz und die räumliche Differenz in einem beliebigen Inertialsystem. Im Sinne eines Riemann'schen Integrals können wir schreiben

$$\tau(\gamma) = \int_{\gamma} d\tau = \int_{t_0}^{t_1} \sqrt{1 - \frac{1}{c^2} \left( \frac{d\mathbf{x}}{dt} \right)^2} dt. \quad (5)$$

$\gamma$  ist hierbei eine Weltlinie, die in einem beliebigen Inertialsystem durch die Zeit  $t$  in diesem System parametrisiert und durch  $\mathbf{x}(t)$  beschrieben wird.  $\tau(\gamma)$  ist die Eigenzeit zwischen den beiden Ereignissen  $\mathbf{x}(t_0)$  und  $\mathbf{x}(t_1)$  entlang der Weltlinie  $\gamma$ .

An dieser Stelle wurde eine wesentliche Annahme gemacht: Durch die Beschleunigung in nicht-inertialen Systemen wird der Gang einer Uhr nicht zusätzlich beeinflusst. Es trägt nur die Eigenzeit bei, die sich aus einer Summe der geraden (inertialen) Abschnitte der durch einen Polygonzug angenäherten Weltlinie ergibt. Die „Knicke“ geben keinen zusätzlichen Beitrag.

Außerdem beachte man, dass die direkte Verbindungsweltlinie zwischen zwei Ereignissen, also die Verbindung in einem Inertialsystem, die größte Eigenzeit hat - alle anderen Weltlinien, die dieselben beiden Ereignisse verbinden, haben kürzere Eigenzeiten. Dies ist umgekehrt zur euklidischen Geometrie, wo die direkte Verbindung die kürzeste Verbindung ist. Der Unterschied beruht auf dem Minuszeichen in Gl. 5: Je schneller ein Weg durchlaufen wird, umso kürzer ist die Eigenzeit; und jeder Weg, der zwei Ereignisse nicht direkt verbindet, muss insgesamt schneller durchlaufen werden.

## 2 Die Euklid- und Minkowski-Schablone

Wenn wir in der euklidischen Ebene den Abstand von zwei Punkten  $A$  und  $B$  bestimmen möchten, legen wir ein Lineal (mit Abstandsmarkierungen) an, sodass die beiden Punkte von dem Lineal verbunden werden, und bilden die Differenz zwischen den Abstandsmarkierungen der beiden Punkte. Das bedeutet einerseits, dass wir die Länge der kürzesten Verbindungsstrecke zwischen den beiden Punkten - einer Geraden - ausmessen und nicht die Länge einer anderen Verbindungsstrecke, und andererseits setzen wir voraus, dass die Länge eines Lineals nicht davon abhängt, an welchem Punkt und unter welcher Orientierung wir das Lineal anlegen.

In einem kartesischen Koordinatensystem ist der euklidische Abstand  $d(A,B)$  zwischen zwei Punkten  $A$  und  $B$  durch

$$d(A,B) = \sqrt{(\Delta x)^2 + (\Delta y)^2} \quad (6)$$

gegeben. Dieses Abstandsmaß ist offensichtlich invariant unter Transformationen, für die

$$(\Delta x)^2 + (\Delta y)^2 = (\Delta x')^2 + (\Delta y')^2 \quad (7)$$

gilt, wobei sich  $\Delta x'$  und  $\Delta y'$  auf die Koordinatendifferenzen zwischen den beiden Punkten in einem anderen kartesischen Koordinatensystem beziehen. Dieses andere Koordinatensystem kann dabei durch eine Verschiebung, Drehung oder Spiegelung aus dem ursprünglichen Koordinatensystem hervorgegangen sein. Und dies sind genau die Transformationen, für die der Ausdruck  $(\Delta x)^2 + (\Delta y)^2$  eine Invariante ist.

Wir können den Abstand zwischen zwei Punkten in der euklidischen Ebene aber auch anders bestimmen: Wir betrachten eine Schablone (siehe Abb. 1, links) aus konzentrischen Kreisen mit äquidistanten Radien. Legen wir den Mittelpunkt dieser Scheibe auf einen Punkt  $A$ , können wir die

Distanz zu  $B$  an dem Radius des Kreises ablesen, auf dem  $B$  liegt. Hierbei nutzen wir aus, dass die Punkte auf einem Kreis alle denselben Abstand von einem Zentralpunkt haben (das ist eine mögliche Definition eines Kreises).

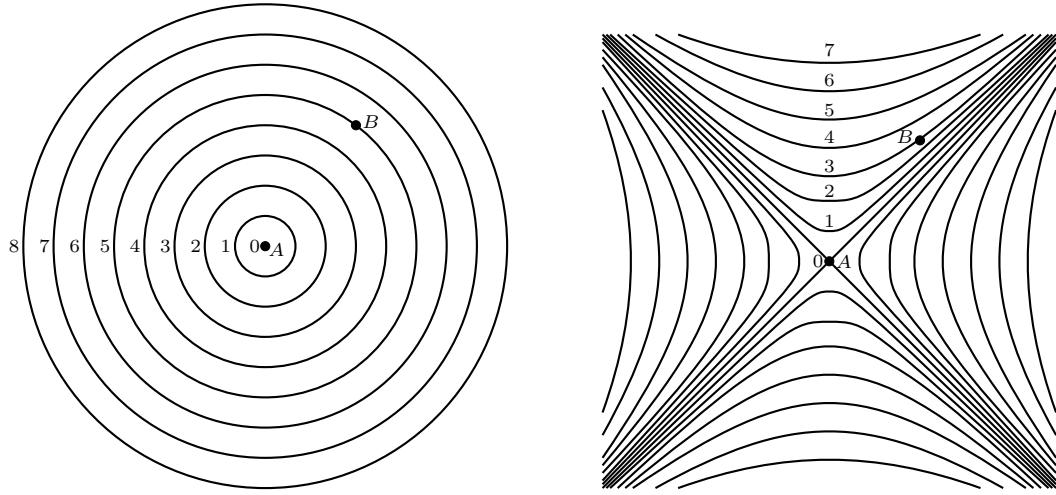


Abbildung 1: Die Euklid-Schablone (links) und die Minkowski-Schablone (rechts). Die Punkte  $A$  und  $B$  im euklidischen Raum (links) haben einen Abstand 5:  $A$  liegt im Zentrum und  $B$  auf dem Kreis mit der Markierung 5. Die Punkte  $A$  und  $B$  auf der rechten Seite haben den Abstand 3, da  $B$  auf der Hyperbel mit der Markierung 3 liegt. In beiden Fällen bezieht sich „Abstand“ immer auf die Länge einer geraden Verbindungslinie zwischen beiden Punkten.

In ähnlicher Weise können wir auch den Abstand zweier Punkte in einem Minkowski-Diagramm bestimmen. Hierbei nutzen wir aus, dass  $(\Delta t)^2 - (\Delta x)^2$  eine Invariante ist und die Wurzel aus diesem Ausdruck die Länge (Eigenzeit) einer geraden Verbindung zwischen zwei Punkten ist (vgl. Gl. 4).<sup>2</sup> Dazu verwenden wir eine Minkowski-Schablone (dies ist ebenso wie Euklid-Schablone kein gängiger Fachausdruck sondern eine von mir gewählte Bezeichnung), bei der Hyperbeln die Punkte konstanten Abstands vom Ursprung anzeigen (siehe Abb. 1, rechts).

Um den Abstand zwischen zwei Ereignissen in einem Raum-Zeit-Diagramm zu ermitteln, legen wir die Schablone mit ihrem Zentrum auf eines der Ereignisse und können nun ablesen, auf welcher Hyperbel das andere Ereignis liegt (in Abb. 1, rechts, hat Ereignis  $B$  den Abstand 3 von Ereignis  $A$ ). Man beachte, dass „Abstand“ wieder die Länge einer geraden Verbindung zwischen den Ereignissen bezeichnet. Das ist bei zeitartigen Ereignissen die Eigenzeit in dem ausgezeichneten Inertialsystem, in dem die beiden Ereignisse  $A$  und  $B$  am selben Raumpunkt stattfinden. Bei raumartigen Ereignissen ist das der euklidische Abstand in einem Inertialsystem, in dem die Ereignisse gleichzeitig stattfinden. Lichtartige Ereignisse haben immer den Abstand 0. Da der Lichtkegel für alle Beobachter derselbe ist, kann man die Schablone nun nicht in der Raum-Zeit-Ebene drehen. Andererseits spielt es keine Rolle, welches Inertialsystem man als Ursprung wählt, d.h. welches Inertialsystem die senkrechte Zeitachse nach oben auszeichnet.

Wie schon erwähnt, ist die Euklid-Schablone invariant unter Verschiebungen, Spiegelungen und Drehungen. Entsprechend ist die Minkowski-Schablone invariant unter Verschiebungen, Spiegelungen an der Zeit- oder Raumachse und unter Lorentz-Transformationen. (Spezielle) Lorentz-Transformationen lassen den Ursprung fest, ebenso die Lichtkegel. Sie transformierten Geraden wieder in Graden (zeitartige Graden in zeitartige Graden und raumartige in raumartige), sodass jeder

<sup>2</sup>Hier und in den folgenden Gleichungen verwenden wir Einheiten, in denen  $c$  den Wert 1 annimmt.

Punkt auf seiner Hyperbel bleibt.

Diese Konzepte lassen sich auf mehr Raumdimensionen verallgemeinern: Die Euklid-Schablone wird in drei Dimensionen zu konzentrischen Kugelschalen und die Minkowski-Schablone erhält man, indem man die obige Konstruktion in höheren Dimensionen um die Zeitachse dreht. In 2+1 Raum-Zeit-Dimensionen wird der Lichtkegel zu einem richtigen (Doppel-)Kegelmantel, die zeitartigen Hyperbeln erhalten die Form von Schalen und die raumartigen Hyperbeln die Form von Reifenfelgen.

### 3 Das Zwillingsparadoxon

Wir sind nun in der Lage, das Zwillingsparadoxon der speziellen Relativitätstheorie zu erläutern. Dazu betrachten wir zwei Personen (Zwillinge), die unterschiedliche Weltlinien durchlaufen (siehe Abb. 2). Person A befindet sich in einem Inertialsystem, d.h. ihre Weltlinie ist eine Gerade. Diese Gerade durchläuft die Ereignisse  $A_0$  bis  $A_4$ . Person B durchläuft eine Weltlinie, die sich bei Ereignis  $B_0 = A_0$  von Person A trennt und sich von A mit großer Geschwindigkeit entfernt. Bei  $B_2$  bremst die Person B ab und beschleunigt anschließend zurück - dies wird in Abb. 2 als Knick dargestellt, könnte aber auch durch eine glatte Kurve beschrieben werden. Bei dem Ereignis  $B_4 = A_4$  treffen die beiden Personen wieder zusammen.

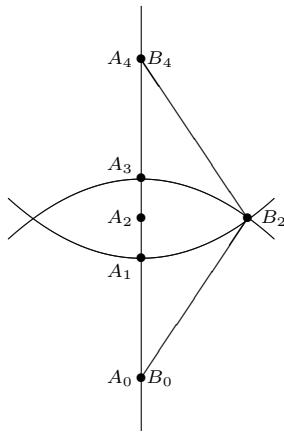


Abbildung 2: Zum Zwillingsparadoxon: Die Weltlinie von Zwilling A verläuft entlang der Ereignisse  $A_0 = B_0, A_1, A_2, A_3, A_4 = B_4$ , die von Zwilling B entlang  $B_0, B_2, B_4$ . Die Weltlinie von Zwilling A ist länger als die von Zwilling B, d.h., Zwilling A ist bei der Wiedervereinigung in Ereignis  $A_4 = B_4$  älter als sein Zwillingspartner. Bei  $B_2$  hat Zwilling B dasselbe Alter wie Zwilling A bei  $A_1$ . Andererseits ist die Eigenzeit zwischen  $A_3$  und  $A_4$  dieselbe wie die zwischen  $B_2$  und  $B_4$ . Insgesamt ist Zwilling A um die Zeitspanne zwischen  $A_1$  und  $A_3$  älter.

Legen wir nun unsere Minkowski-Schablone in den Punkt  $A_0$  so erkennen wir, dass die Eigenzeit von  $A_0$  bis  $A_1$  für Person A genauso lang ist wie die Eigenzeit von Person B von  $B_0 (= A_0)$  bis zum Ereignis  $B_2$ . Diese beiden Punkte liegen auf einer zeitartigen Hyperbel. Umgekehrt können wir unsere Schablone in den Punkt  $A_4$  legen und erkennen, dass die Eigenzeit von Ereignis  $A_3$  bis  $A_4$  für Person A genauso lang ist, wie die Eigenzeit von  $B_2$  bis  $B_4$  für Person B. Für Person B haben wir aber damit die gesamte Weltlinie ausgemessen: Sie hat dieselbe Länge (Eigenzeit), wie die beiden Abschnitte der Weltlinie von Person A von  $A_0$  bis  $A_1$  plus  $A_3$  bis  $A_4$ . Für Person A kommt aber noch die Eigenzeit von Abschnitt  $A_1$  bis  $A_3$  hinzu. Um diese Eigenzeit ist Person A beim abschließenden Zusammentreffen älter als Person B.

### 4 Die Rolle der Beschleunigung

In seinen berühmten „Feynman Lectures on Physics“ beschreibt Feynman in Kapitel 16-2 auch das Zwillingsparadoxon. Er schreibt dort ... *the man who has felt the accelerations ... is the one who*

*would be younger.* Diese (und ähnliche Bemerkungen in anderen Lehrbüchern) haben gelegentlich dazu geführt, dass die Beschleunigung als Ursache dafür angesehen wird, dass der eine Zwilling jünger bleibt. Andererseits haben wir oben betont, dass die Beschleunigung keinen Einfluss auf den Gang einer (idealen) Uhr haben soll und die Länge einer Weltlinie nicht davon abhängt, wie viele Beschleunigungsphasen auftreten, sondern nur davon, wie lang die Summe der infinitesimalen inertialen Abschnitte ist, durch die wir die Weltlinie immer besser annähern können (also das Riemann'sche Integral in Gl. 5; in das nur die Geschwindigkeit aber keine Beschleunigung eingeht).

Tatsächlich ist auch die Beschleunigung nicht die Ursache dafür, dass eine Person jünger geblieben ist, aber das behauptet Feynman auch nicht. Die Beschleunigungsphase ist lediglich die physikalisch nachweisbare Eigenschaft, um die sich das Bezugssystem von B von dem Bezugssystem von A unterscheidet (damit ist gemeint, dass man in einem lokalen, abgeschlossenen Labor feststellen kann, dass eine Beschleunigung vorliegt). Wären die beiden Bezugssysteme physikalisch gleichwertig (würde es sich beispielsweise bei beiden Bezugssystemen um Inertialsysteme handeln), würde aus dem Relativitätsprinzip folgen, dass auch die Physik in beiden Systemen die gleiche sein muss und damit kann nicht in dem einen System mehr und dem anderen weniger Eigenzeit vergangen sein. Die Tatsache, dass die Beschleunigungsphase die beiden Bezugssysteme unterscheidet, bedeutet nicht, dass die Beschleunigung auch die unmittelbare Ursache dafür ist, dass diese Person jünger geblieben ist.

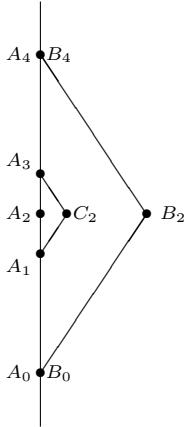


Abbildung 3: Erweiterung des Zwillingsparadoxons für Drillinge. Die drei Weltlinien –  $(A_0 \rightarrow A_1 \rightarrow A_2 \rightarrow A_3 \rightarrow A_4)$  für Drilling A,  $(A_0 \rightarrow A_1 \rightarrow C_2 \rightarrow A_3 \rightarrow A_4)$  für Drilling C und  $(B_0 \rightarrow B_2 \rightarrow B_4)$  für Drilling B – sind unterschiedlich lang. Insbesondere ist die Weltlinie von Drilling B kürzer als die von Drilling C, obwohl beide dieselben Beschleunigungsphasen erlebt haben.

Um dieses Argument zu untermauern, betrachten wir in Abb. 3 eine etwas erweiterte Situation. Nun sind drei Personen (Drillinge) gleichen Alters gegeben: A, B und C. Die Personen A und B durchlaufen dieselben Weltlinien wie oben. Die Person C verbleibt jedoch bei Person A bis zum Ereignis  $A_1$ . Während A weiterhin in einem Inertialsystem verbleibt, beschleunigt C von A weg bis zum Ereignis  $C_2$ , kehrt dort um und trifft bei  $A_3$  wieder mit A zusammen. Wir haben es nun also mit drei Weltlinien zu tun. Wie man in Abb. 3 erkennen kann, sind die Beschleunigungsphasen von B und C identisch: Eine Beschleunigung von A weg (bei  $B_0$  und  $A_1$ ) eine Beschleunigung für die Wende (bei  $C_2$  und  $B_2$ ) und ein Abbremsen in das Inertialsystem von A (bei  $A_3$  und  $B_4$ ). Die Beschleunigungen sind auch gleich groß. Doch obwohl Person C dieselben Beschleunigungsphasen mitgemacht hat wie B, ist die Weltlinie von B deutlich kürzer und Person B ist am Ende die Jüngste. Person C ist älter als B (trotz derselben Beschleunigungphasen) aber jünger als A. Der wichtige Unterschied in den Bezugssystemen von C und B besteht nicht in den Beschleunigungsphasen, sondern in den Zeitdauern zwischen den Beschleunigungsphasen bzw. den in diesen Zeitdauern zurückgelegten unterschiedlichen Weltlinien.

Betrachten wir dazu noch ein Beispiel aus der euklidischen Geometrie (siehe Abb. 4). Dort

gilt die Dreiecksungleichung:  $d(A,C) \leq d(A,B) + d(B,C)$ . Feynmans Bemerkung könnte man auf diesen Fall übertragen: Der Weg von  $A$  nach  $C$ , der einen Knick hat, ist der längere. Oder, wenn wir es physikalischer formulieren wollen: Der Weg von  $A$  nach  $C$ , bei dem man irgendwann beschleunigen muss, ist der längere. Diese Aussage ist sicherlich richtig. Aber es wäre irreführend zu sagen, der Knick sei die Ursache dafür, dass der Weg über den Punkt  $B$  der längere sei.

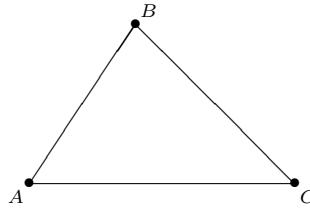


Abbildung 4: Ein Dreieck in der Euklidischen Ebene. Der Weg von Punkt  $A$  nach Punkt  $C$  über den Punkt  $B$  ist länger als der direkte Weg. Trotzdem würde man den Knick bei  $B$  nicht als Ursache dafür ansehen, dass dieser Weg länger ist, obwohl jeder Weg, der länger als die direkte (gerade) Verbindungslinie ist, einen Knick (oder Bogen) haben muss.

Damit erhebt sich die Frage, welche Rolle die Beschleunigung für das unterschiedliche Alter der Zillinge denn nun wirklich spielt. Man könnte das Argument von Feynman ja auch folgendermaßen formulieren: In allen Inertialsystemen ist nach dem Relativitätsprinzip die Physik dieselbe, und da es sich bei den Abschnitten  $B_0$  nach  $B_2$  einerseits und  $B_2$  nach  $B_4$  andererseits um inertiale Weltlinien handelt, ebenso wie für Beobachter A die gesamte Weltlinie von  $A_0$  bis  $A_4$ , kann der Unterschied in den beiden Weltlinien nur von dem „Knick“ bei  $B_2$ , also der Beschleunigungsphase, herrühren.

Vielleicht sollte man hier einen Unterschied machen zwischen „Ursache für etwas sein“ und „Indiz für etwas sein“. Nach den Postulaten der Relativitätstheorie gibt es keine gesonderten Beiträge zur Eigenzeit, wenn ein System beschleunigt wird. Uhren laufen in solchen Phasen bzw. Momenten nicht plötzlich schneller oder langsamer oder machen Sprünge. Aber eine Beschleunigungsphase zu haben ist eine notwendige und hinreichende Bedingung dafür, dass ein Bezugssystem kein Inertialsystem ist, und damit gilt für ein solches System auch das Relativitätsprinzip nicht mehr. Unter den unendlich vielen Weltlinien, die zwei zeitartige Ereignisse miteinander verbinden, gibt es nur eine, die den Ursprung eines Inertialsystems definiert, und diese Weltlinie hat keine Beschleunigungsphase. Alle anderen Weltlinien haben eine kürzere Eigenzeit und notwendigerweise eine Beschleunigungsphase. Allerdings hängt die Eigenzeit einer Weltlinie nicht davon ab, wie viele Beschleunigungsphasen auftreten oder wie groß die Beschleunigungen sind. Die folgenden Beispiele sollen dies noch einmal verdeutlichen.

In Abbildung 5 sind verschiedene Abwandlung des Zwillingsparadoxons dargestellt. In Abb. 5 (a) ist nochmals das Drillingsparadoxon aus Abb. 3 wiedergegeben, allerdings hat Person B nun vergleichsweise schwache Beschleunigungsphasen wohingegen Person C stärkeren Beschleunigungen unterliegt. Allein durch Variation der Abstände zwischen den Ereignissen (z.B. den Ereignissen  $A_1$  und  $A_3$ , bei denen Person C zur Weltreise ansetzt) kann man erreichen, dass entweder die Weltlinie von C kürzer ist als die von B oder aber länger. Die Weltlinie von Person A, die keine Beschleunigungen erfährt, bleibt natürlich die längste. Dieses Beispiel zeigt, dass die Stärke der Beschleunigungen nicht darüber entscheidet, welche Weltlinie länger und welche kürzer ist.

In Abb. 5 (b) hat Person C mehrere Beschleunigungsphasen, die der Beschleunigung von Person B im Punkt  $B_2$  entsprechen. Person C fliegt kürzere Strecken im Zickzack und beschleunigt bei  $C_1, A_1, C_2, A_3$  und  $C_3$  (abgesehen von den Beschleunigungsphasen bei  $B_0$  und  $B_4$ , die Person C mit Person B gemeinsam hat). Trotzdem sind die Weltlinien von C und B gleich lang. Das zeigt, dass die Anzahl der Beschleunigungsphasen nicht darüber entscheidet, welche Weltlinie kürzer oder länger ist.

Schließlich haben wir es in Abb. 5 (c) mit drei intertialem Weltlinien A, B und C zu tun. Hier finden überhaupt keine Beschleunigungen statt, allerdings werden bei den Ereignissen  $A$  und  $B$

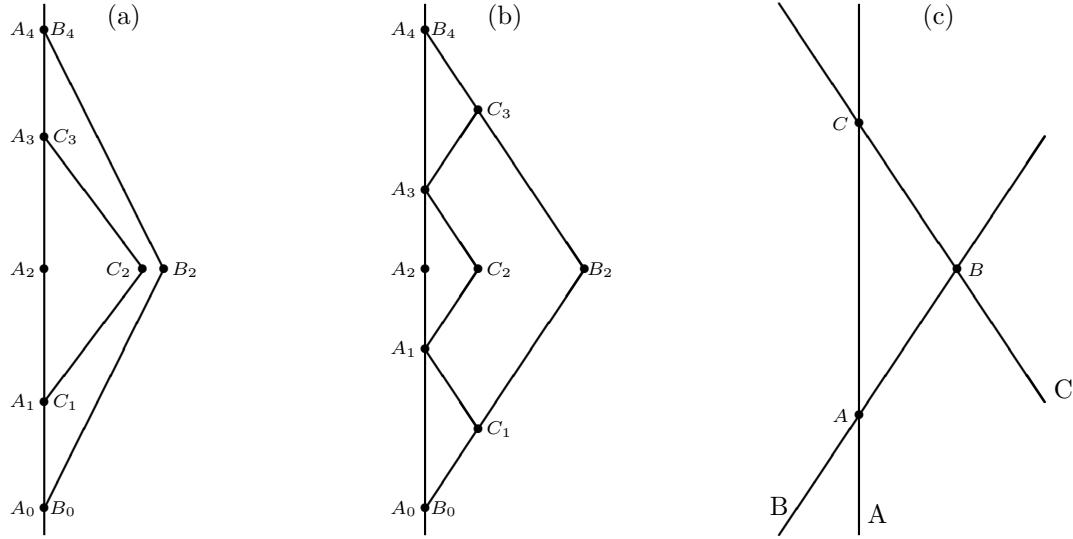


Abbildung 5: Verschiedene Beschleunigungsphasen im Zwillings- bzw. Drillingsparadoxon. (a) Eine Abwandlung des Drillingsparadoxons: Person C hat stärkere Beschleunigungsphasen als Person B. Durch geeignete Verschiebung der Abstände - ohne Veränderung der Beschleunigungen - kann man erreichen, dass die Weltlinie von C länger oder kürzer ist als die von B. (b) Person C beschleunigt nun mehrfach und bewegt sich entlang einer Zickzacklinie, wohingegen Person B nur einmal beschleunigt (in Ereignis  $B_2$ ). Trotzdem sind B und C beim Wiedersehen in Ereignis  $A_4$  gleich alt. (c) Die drei Intertialsysteme (A, B, C) treffen sich paarweise in den Ereignissen A, B und C und synchronisieren in diesen Momenten ihre Uhren. Es finden keine Beschleunigungen statt. In C treffen A und C zusammen und vergleichen ihre Uhren. Die Uhr von A zeigt mehr verflossene Zeit an als die Uhr von C.

Uhren synchronisiert. Bei Ereignis A synchronisierten Person A und B ihre Uhren, bei Ereignis B synchronisieren nochmals Person B und C ihre Uhren und zwar derart, dass C seine Uhr auf die Uhr von B einstellt.

Es werden also verschiedene Uhren entlang der Weltlinie von A über B nach C so synchronisiert, dass die jeweilige Uhr die Eigenzeit entlang dieser Weltlinie anzeigt. Andererseits zeigt die Uhr von Person A die Eigenzeit der Weltlinie A an. Wenn sich bei C die Personen A und C mit ihren Uhren treffen, zeigt die Uhr von C weniger Zeit an als die Uhr von A und zwar in demselben Maß, in dem ein Zwilling entlang der Weltlinie über B jünger geblieben wäre.

Dieses letzte Beispiel ist gleichzeitig ein Beweis, dass bei einer Beschleunigung keine zusätzliche Beeinflussung einer Uhr und damit der Eigenzeit entlang einer Weltlinie stattfindet. Die Uhren selbst werden nicht beschleunigt, sondern lediglich an den Treffpunkten, wo auch keine Laufzeitverzögerungen der Signalübertragung berücksichtigt werden müssen, synchronisiert. Die Uhr von Person C, die letztendlich bei Ereignis C wieder mit A zusammentrifft, zeigt dieselbe Zeit an, die auch eine Uhr angezeigt hätte, die bei den Ereignissen A und B beschleunigt worden wäre.

## 5 Vergleich der Bezugssysteme

Nachdem wir in den vergangenen Abschnitten festgestellt haben, dass die Beschleunigung keinen unmittelbaren Einfluss auf den Gang einer Uhr hat und somit nicht die Ursache für die unterschiedlichen Eigenzeiten entlang der verschiedenen Weltlinien ist, kommen wir nochmals auf die Frage zurück, welche Rolle die Beschleunigung bei dem Zwillingsparadoxon spielt. Insbesondere interessiert in diesem

Zusammenhang, was der Zwilling in einem Bezugssystem von den Ereignissen auf der Weltlinie des Zwillings in dem jeweils anderen Bezugssystem beobachtet. Es zeigt sich, dass die Beschleunigung in diesem Fall eine sehr große Rolle spielt.

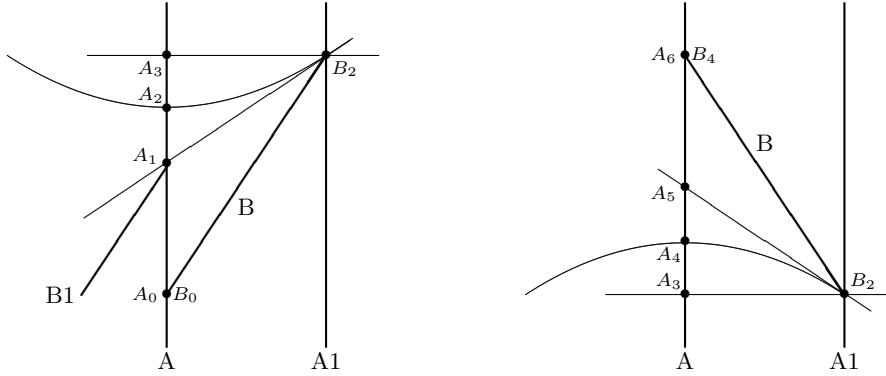


Abbildung 6: Die verschiedenen Phasen des Zwillingsparadoxons. (links) Der erste Teil der Reise von Zwilling B bis kurz vor dem Umkehrpunkt  $B_2$ . (rechts) Der zweite Teil der Reise nach dem Umkehrpunkt  $B_2$ . (Erläuterungen siehe Text.)

Betrachten wir zunächst den ersten Teil der Reise, bis Zwilling B das Ereignis  $B_2$  erreicht. In Abb. 6, links, sind die Weltlinien von vier Beobachtern dargestellt: (A) die Weltlinie von Zwilling A, ( $A_1$ ) eine zweite Weltlinie in dem Bezugssystem von A (also parallel zur Weltlinie von A), allerdings geht diese Weltlinie durch das Ereignis  $B_2$ ; (B) die Weltlinie von B und ( $B_1$ ) die Weltlinie eines zweiten Beobachters in dem Bezugssystem von B. Außerdem ist eine waagerechte Linie durch die Ereignisse  $A_3$  und  $B_2$  dargestellt: Sie repräsentiert alle Ereignisse, die in dem Bezugssystem von A zum selben Zeitpunkt wie das Ereignis  $B_2$  stattfinden. Eine weitere Gleichzeigkeitslinie durch die Punkte  $A_1$  und  $B_2$  repräsentiert alle Ereignisse, die im Bezugssystem von B gleichzeitig zum Ereignis  $B_2$  sind.

Man erkennt nun Folgendes: Rein objektiv, ohne auf die unterschiedlichen globalen Gleichzeigkeitsdefinitionen von A und B Bezug zu nehmen, zeigen die Uhren von A und B in den Ereignissen  $A_2$  und  $B_2$  dieselbe Zeit an - sie liegen auf derselben Hyperbel der Minkowski-Schablone. Wenn jedoch Zwilling B das Ereignis  $B_2$  erreicht, hat Zwilling A bezüglich seiner Gleichzeigkeitsdefinition das Ereignis  $A_3$  erreicht. Auf der Uhr von A ist aber bei diesem Ereignis mehr Zeit vergangen, als auf der Uhr von B. Daher hat A den Eindruck, die Uhr von B gehe langsamer (entsprechend der bekannten Zeidilatation in der Speziellen Relativitätstheorie). Dies wird allerdings nicht direkt von A gemessen, sondern in seinem Bezugssystem von  $A_1$ , dessen Uhren mit A synchronisiert sind. Für den Beobachter B bzw. in seinem Bezugssystem hat A aber erst das Ereignis  $A_1$  erreicht, wenn B bei  $B_2$  ankommt. Von dem Beobachter  $B_1$ , der sich im Bezugssystem von B befindet und dessen Uhr mit der von B synchronisiert ist, wird registriert, dass bei diesem Ereignis auf der Uhr von A weniger Zeit vergangen ist. Insofern hat man in dem Bezugssystem von B den Eindruck, die Uhren in dem System A gingen langsamer.

Abb. 6, rechts, zeigt die gleiche Situation für den zweiten Teil der Reise, nachdem Beobachter B bei  $B_2$  beschleunigt hat und sich nun wieder auf A zubewegt. Die Ereignisse  $A_3$  und  $B_2$  sind für A gleichzeitig (wie vorher), nun sind für B aber die Ereignisse  $A_5$  und  $B_2 = B_4$  gleichzeitig. Die Eigenzeit, angezeigt von der Uhr von B, zwischen den Ereignissen  $B_2$  und  $A_6 = A_4$ , ist dieselbe, wie die Eigenzeit in dem System von A zwischen den Ereignissen  $A_4$  und  $A_6$ . Insgesamt kommen wir wieder zu dem Ergebnis, dass die Gesamtzeit, die im Bezugssystem von B vergangen ist, gleich den beiden Zeidauern

$A_0$  bis  $A_2$  plus  $A_4$  bis  $A_6$  im Bezugssystem von A ist, und in diesem Bezugssystem die Zeitdifferenz zwischen  $A_2$  und  $A_4$  zusätzlich vergangen ist.

Wir erkennen jetzt die besondere Bedeutung der Beschleunigung in  $B_2$ : Sie verändert in diesem „Moment“ die Gleichzeitigkeitslinien von Bezugssystem B und zwar derart, dass kurz vor dem Ereignis  $B_2$  für Beobachter B das Ereignis  $A_1$  gleichzeitig ist, und unmittelbar nach dem Ereignis  $B_2$  ist es das Ereignis  $A_5$ . Durch die Beschleunigung verpasst Beobachter B also alle Ereignisse zwischen  $A_1$  und  $A_5$  (bzw., da jede Beschleunigung eine endliche Zeitdauer benötigt, werden diese Ereignisse in dem Bezugssystem von B in einem beliebig kurzen Zeitraum „erlebt“). Man könnte etwas übertrieben sagen, dass für Zwilling B die Ereignisse zwischen  $A_2$  und  $A_4$  im Bezugssystem von A aufgrund der „unendlichen“ Beschleunigung keine Zeitzuordnung haben.

In diesem Zusammenhang ist anzumerken, dass die Konstruktion von globalen Gleichzeitigkeitslinien (bzw. in drei Raumdimensionen „Gleichzeitigkeitsräumen“) eine Besonderheit der Speziellen Relativitätstheorie ist. Diese Konstruktion ist nur sinnvoll, solange man es mit Inertialsystemen zu tun hat, deren Weltlinien Geraden sind. Rein operational setzt sie voraus, dass sich zwei Beobachter im selben Bezugssystem für die Zeit, während der sie im Sinne der Einstein-Synchronisation ihre Signale austauschen, auf geraden Weltlinien bewegen. Sobald ein Bezugssystem eine Beschleunigung erfährt, machen solche globalen Gleichzeitigkeitslinien keinen Sinn mehr: In manchen Bereichen läuft die Zeit rückwärts, in anderen läuft sie beliebig schnell vorwärts; und operational lässt sich eine Einstein-Synchronisation in diesen Fällen nicht sinnvoll durchführen (sie würde verschiedene Gleichzeitigkeitsdefinitionen für Beobachter im selben Bezugssystem ergeben). Daher betrachtet man in der Allgemeinen Relativitätstheorie auch lieber das, was ein Beobachter von den Ereignissen wirklich sieht, d.h., man berücksichtigt die Laufzeitverzögerungen durch die endliche Ausbreitungsgeschwindigkeit von Licht. In der Speziellen Relativitätstheorie sollte man dies bei beschleunigten Bezugssystemen ebenfalls tun.

## 6 Kuriositäten

### 6.1 Das Zwillingsparadoxon in einem periodischen Universum

In einem räumlich periodischen Universum können zwei verschiedene inertiale Weltlinien dieselben zwei Ereignisse verbinden. Ein solches periodisches Universum kann lokal dem flachen Minkowski-Raum entsprechen und ist somit eine Lösung der Einstein-Gleichungen der allgemeinen Relativitätstheorie. Die Einstein-Gleichungen legen keine globalen topologischen Eigenschaften der Raum-Zeit fest.

Wir betrachten wieder zwei Bezugssysteme (Beobachter) A und B. Bezugssystem A ist „in Ruhe“ und die zugehörige Weltline verbindet die beiden Ereignisse A und B direkt. Bezugssystem B hat bezüglich A eine bestimmte Geschwindigkeit. Die beiden Bezugssysteme treffen sich bei Ereignis A. Bezugssystem B bewegt sich nun mit seiner Geschwindigkeit weiter, windet sich einmal um das periodische Universum und trifft bei Ereignis B wieder mit Bezugssystem A zusammen (siehe Abb. 7).

Legen wir nun wieder unsere Minkowski-Schablone an die Weltlinien, stellen wir fest, dass die Weltlinie zwischen den beiden Ereignissen A und B von Bezugssystem B kürzer ist als die Weltlinie von A. Doch in diesem Fall sind beide Bezugssysteme Inertialsysteme. Wie kann das sein?

Die Antwort auf diesen scheinbaren Widerspruch lautet: Das Relativitätsprinzip gilt nicht mehr. Wir haben oben schon davon gesprochen, dass sich das Bezugssystem A „in Ruhe“ befindet. Dieser Ausdruck ist in diesem Fall sinnvoll: Es gibt nur ein Bezugssystem, für das die Einstein-Synchronisation von Uhren global konsistent ist. Das bedeutet Folgendes: Wir können in einem

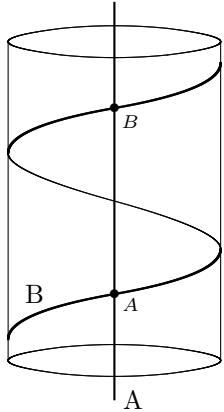


Abbildung 7: Das Zwillingsparadoxon in einem räumlich periodischen Universum. Eine Person bleibt an ihrem Ort, die andere bewegt sich mit konstanter Geschwindigkeit einmal um „das Universum“ herum. Beide trennen sich bei Ereignis A, wo sie gleich alt sind bzw. ihre Uhren synchronisiert haben, und treffen bei Ereignis B wieder aufeinander. Person B ist jünger als Person A. Beide befinden sich während der gesamten Zeit in einem Inertialsystem. In diesem Fall gilt jedoch das Relativitätsprinzip nicht.

räumlich periodischen Universum die Synchronisation von Uhren an verschiedenen Punkten in einem Bezugssystem auf zwei Weisen durchführen. Wir können die beiden Punkte wegen der Periodizität des Raums auf verschiedene Weisen verbinden (einmal links um den Torus und einmal rechts um den Torus in Abb. 7) und die Einstein-Synchronisation entlang beider Richtungen durchführen. Es gibt nur ein System - und dies bezeichnen wir als das Ruhesystem - bei dem diese beiden Synchronisationsvorschriften dieselbe Gleichzeitigkeitszuordnung liefern. Wenn aber ein absolutes Ruhesystem ausgezeichnet und physikalisch bestimmbar ist, gilt die Lorentz-Invarianz und damit auch das Relativitätsprinzip nicht mehr.

# Beschleunigte Systeme und das Rindler-Universum

Thomas Filk; Universität Freiburg

Das Äquivalenzprinzip besagt im Wesentlichen, dass wir in einem lokalen Bezugssystem nicht zwischen einer konstanten Beschleunigung und einem konstanten Gravitationsfeld unterscheiden können. Wir werden dieses Prinzip in den nächsten Abschnitten ausgiebig nutzen, um den Einfluss von Gravitationsfeldern zu untersuchen und damit ersten Schritte zur Allgemeinen Relativitätstheorie zu gehen.

In diesem Kapitel geht es konkret um einen konstant beschleunigten Beobachter in der Speziellen Relativitätstheorie. Viele der Effekte lassen sich dann über das Äquivalenzprinzip auf die Allgemeine Relativitätstheorie übertragen.

## 1 Die konstante Beschleunigung

Schon allein die Frage, was genau unter einer konstanten Beschleunigung zu verstehen ist, bedarf in der Speziellen Relativitätstheorie einer eingehenderen Analyse. Wir können an einem ausgedehnten Körper nicht einfach eine Kraft angreifen lassen, da kein Körper wirklich starr ist - dies würde der Speziellen Relativitätstheorie widersprechen - und sich die Wirkung jeder angreifenden Kraft erst über eine Stoßwelle auf den Körper ausdehnt. Der Einfachheit halber betrachten wir daher zunächst nur einen idealisierten Massepunkt, der konstant beschleunigt werden soll. Doch auch hier ist das Konzept einer konstanten Beschleunigung nicht trivial.

Einerseits befindet sich der beschleunigte Gegenstand zu jedem Zeitpunkt in einem anderen Inertialsystem, andererseits hängt die naheliegende Antwort – eine konstante Beschleunigung bedeutet einen pro Zeiteinheit konstanten Geschwindigkeitszuwachs – vom Bezugssystem ab. Eine invariante Definition des Konzepts einer konstanten Beschleunigung, die wir auch in diesem Kapitel verwenden werden, ist folgende: *Im jeweiligen momentanen Inertialsystem des beschleunigten Massepunktes ist die Beschleunigung konstant.* Damit ist gemeint, dass es in jedem Augenblick ein Inertialsystem gibt, das dieselbe Geschwindigkeit wie der beschleunigte Beobachter hat (natürlich ändert sich dieses Inertialsystem ständig); von diesem „momentanen Inertialsystem“ aus betrachtet erfährt der beschleunigte Beobachter einen konstanten Geschwindigkeitszuwachs.

In diesem Abschnitt betrachten wir die Bewegung eines konstant beschleunigten Massepunktes von einem Inertialsystem aus, in dem der Massepunkt zum Zeitpunkt  $t_0 = 0$  ruht. Zunächst leiten wir die Differentialgleichung für die Geschwindigkeit – gemessen in diesem Inertialsystem – her, anschließend lösen wir diese Gleichung und bestimmten die Bahnkurve  $x(t)$  für diesen Massepunkt.

### 1.1 Herleitung der Differentialgleichung für die Geschwindigkeit

Die Differentialgleichung für die Geschwindigkeit im Inertialsystem des zu Beginn ruhenden Teilchens lautet:

$$\frac{dv(t)}{dt} = g \left(1 - \frac{v(t)^2}{c^2}\right)^{3/2}. \quad (1)$$

Diese Gleichung werden wir im nächsten Abschnitt lösen, doch zunächst wollen wir sie auf zwei verschiedene Weisen ableiten.

In dem Ruhesystem des Massepunktes zu einem bestimmten Zeitpunkt ( $t$  im Inertialsystem des anfänglichen Ruhezustands,  $\tau$  im Bezugssystem des beschleunigten Massepunktes) soll eine Kraft

wirken, die ihn in der infinitesimalen Eigenzeit  $d\tau$  immer auf dieselbe infinitesimale Geschwindigkeit  $dv$  beschleunigt:

$$dv = g d\tau, \quad (2)$$

wobei  $g$  ein Maß für die konstante Beschleunigung ist.

Im anfänglichen Ruhesystem, d.h. bezüglich der Zeit  $t$ , nimmt die Geschwindigkeit also zu einem Zeitpunkt  $t$  um

$$dv = g \sqrt{1 - \frac{v(t)^2}{c^2}} dt \quad (3)$$

zu. Hierbei wurde die Beziehung  $d\tau = \sqrt{1 - v^2/c^2} dt$  zwischen der Eigenzeit  $\tau$  und der Zeit  $t$  in einem (momentanen) Inertialsystem verwendet.

Hat für den ruhenden Beobachter das beschleunigte System zum Zeitpunkt  $t$  die Geschwindigkeit  $v(t)$ , so hat es zum Zeitpunkt  $t + dt$  nach dem Geschwindigkeitsadditionstheorem (vgl. Gl. 21) die Geschwindigkeit:

$$v(t + dt) = \frac{v(t) + dv}{1 + \frac{v(t) dv}{c^2}} \approx v(t) + \left(1 - \frac{v(t)^2}{c^2}\right) dv + \dots \quad (4)$$

$$= v(t) + g \left(1 - \frac{v(t)^2}{c^2}\right)^{3/2} dt + \dots \quad (5)$$

Daraus erhalten wir unmittelbar Gl. 1.

Die zweite Herleitung der Differentialgleichung geht von der allgemeinen Beziehung für das Transformationsgesetz der Beschleunigung zwischen zwei Bezugssystemen aus. Wenn die Beschleunigung in dieselbe Richtung wie die Transformation erfolgt, gilt:

$$a' = \gamma^3 a. \quad (6)$$

Mit  $a' = g$  und  $a = \frac{dv}{dt}$  folgt obige Differentialgleichung.

## 1.2 Bestimmung der Bahnkurve

Wir können die Differentialgleichung 1 beispielsweise durch Trennung der Variablen lösen:

$$\frac{dv}{\left(1 - \frac{v^2}{c^2}\right)^{3/2}} = g dt. \quad (7)$$

Da

$$\frac{d}{dv} \left( \frac{v}{\sqrt{1 - \frac{v^2}{c^2}}} \right) = \frac{1}{\left(1 - \frac{v^2}{c^2}\right)^{3/2}} \quad (8)$$

und wir für  $t_0 = 0$  die Anfangsbedingung  $v = 0$  gesetzt haben, erhalten wir

$$\frac{v(t)}{\sqrt{1 - \frac{v(t)^2}{c^2}}} = gt. \quad (9)$$

Diese Gleichung können wir nach  $v(t)$  auflösen und finden:

$$v(t) = \frac{gt}{\sqrt{1 + \frac{(gt)^2}{c^2}}}. \quad (10)$$

Für  $t \ll c/g$  nimmt  $v(t)$  offensichtlich linear mit  $t$  zu, wie es in der nicht-relativistischen Mechanik für die konstante Beschleunigung gelten muss, für  $t \gg c/g$  nähert sich  $v(t)$  der Lichtgeschwindigkeit  $c$  als der Grenzgeschwindigkeit.<sup>1</sup>

Wir können diese Gleichung nochmals integrieren und erhalten die Trajektorie  $x(t)$ :

$$x(t) = \int_0^t \frac{gt'}{\sqrt{1 + \frac{(gt')^2}{c^2}}} dt' = \frac{c^2}{g} \sqrt{1 + \frac{(gt')^2}{c^2}} \Big|_0^t \quad (11)$$

$$= \frac{c^2}{g} \left( \sqrt{1 + \frac{(gt)^2}{c^2}} - 1 \right). \quad (12)$$

Im nächsten Abschnitt werden wir die Lösungen für  $v(t)$  und  $x(t)$  etwas genauer untersuchen. Zum Abschluss dieses Abschnitts möchte ich noch eine kurze Anmerkung zu der relativistischen Bewegungsgleichung machen:

Multiplizieren wir Gleichung 9 auf beiden Seiten mit der Konstanten  $m$  (der Ruhemasse des Teilchens) und bilden die Ableitung nach  $t$ , so erhalten wir:

$$\frac{d}{dt} \left( \frac{mv(t)}{\sqrt{1 - \frac{v(t)^2}{c^2}}} \right) = mg, \quad (13)$$

was wir mit dem relativistischen Impuls

$$p = \frac{mv}{\sqrt{1 - \frac{v^2}{c^2}}} \quad (14)$$

auch in der Form

$$\frac{dp}{dt} = mg = F \quad (15)$$

schreiben können. Dies ist die relativistische Bewegungsgleichung (auch für eine allgemeine Kraft  $F$ ) und aus ihr hätten wir die Bewegungsgleichung für die konstante Beschleunigung durch Umkehrung der obigen Schritte sofort ableiten können. Es mag allerdings zunächst überraschen, dass auf der linken Seite der Gleichung die Ableitung nach  $t$  und nicht nach der Eigenzeit  $\tau$  im beschleunigten System steht. Es sieht daher zunächst so aus, als ob diese Gleichung nicht invariant sei. Doch die relativistische Kraft ist eigentlich nicht  $F$  sondern  $\gamma F$  und die invariante Gleichung lautet

$$\frac{dp}{d\tau} = \gamma F. \quad (16)$$

Da

$$\frac{dp}{d\tau} = \frac{dp}{dt} \frac{dt}{d\tau} = \frac{dp}{dt} \gamma, \quad (17)$$

hebt sich der  $\gamma$ -Faktor auf beiden Seiten weg.

### 1.3 Erste Analyse der konstant beschleunigten Bewegung

Wir untersuchen zunächst die Trajektorie aus der Sichtweise des ruhenden Beobachters (im Ruhesystem der Ausgangslage des beschleunigten Systems). Anschließend betrachten wir die Situation aus der Sichtweise eines Beobachters in dem konstant beschleunigten System (beispielsweise einer konstant beschleunigten Rakete).

---

<sup>1</sup>Für die Erdbeschleunigung  $g = 9,81 \text{ m/s}^2$  entspricht die Zeitskala  $c/g \approx 354$  Tage bzw. fast einem Jahr. Ab dieser zeitlichen Größenordnung lohnt sich eine Weltraumreise mit konstanter Beschleunigung.

Die Trajektorie der konstanten relativistischen Beschleunigung im Inertialsystem der anfänglichen Ruhelage beschreibt einen Hyperbelbelast (siehe Abb. 1). Dies sieht man leicht, wenn man Gl. 12 in folgende Form bringt:

$$\left( x(t) + \frac{c^2}{g} \right)^2 - \frac{(gt)^2}{c^2} = 1. \quad (18)$$

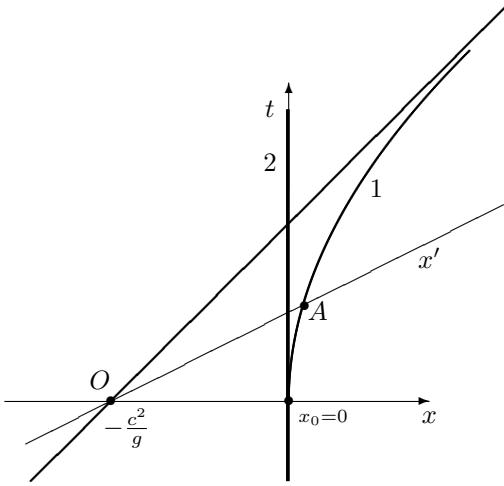


Abbildung 1: Die konstante Beschleunigung. Die Trajektorie eines konstant beschleunigten Massepunktes (1) beschreibt im Inertialsystem eines ruhenden Beobachters (2) eine Hyperbel. Bei dem Ereignis  $O$  mit den Koordinaten  $x = -\frac{c^2}{g}$  und  $t_0 = 0$  schneiden sich alle Gleichzeitzitigkeitslinien der Trajektorie einschließlich der Weltlinie des Lichtstrahls, dem sich die Trajektorie asymptotisch nähert.

Für kleine Werte von  $t$  (genauer  $t \ll \frac{c}{g}$ ) beschreibt die Trajektorie die zu erwartende Parabel der Newton'schen Mechanik. Dazu entwickeln wir die Lösung nach kleinen Werten von  $(tg)/c$ :

$$x(t) \approx \frac{c^2}{g} \left( 1 + \frac{1}{2} \frac{g^2 t^2}{c^2} - \frac{1}{8} \frac{g^4 t^4}{c^4} + \dots - 1 \right) = \frac{1}{2} g t^2 + \dots \quad (19)$$

Für sehr große Werte von  $t$  nähert sich die Trajektorie immer mehr dem Lichtstrahl  $x(t) = ct$ . Dieses Verhalten zeigt sich auch in der Geschwindigkeit (Gl. 10), die für kleine Werte von  $t$  linear zunimmt,  $v(t) \approx gt + \dots$ , und sich für große Werte von  $t$  der Lichtgeschwindigkeit nähert.

Interessant ist, dass sich alle Gleichzeitzitigkeitslinien zu der Trajektorie in einem Ereignis  $O$  bei  $t = 0$  und  $x = -\frac{c^2}{g}$  schneiden (in Abb. 1 ist die Gleichzeitzitigkeitslinie zum Ereignis  $A$  angegeben). Dies ist gleichzeitig das Ereignis, bei dem ein ausgesandter Lichtstrahl die Asymptote der Hyperbel bildet. Der Abstand, gemessen in einem augenblicklichen Inertialsystem, zwischen einem Punkt auf der Hyperbelbahn (z.B. dem Ereignis  $A$ ) und diesem Ereignis  $O$  bleibt konstant.

Wir betrachten nun dieselbe Situation, allerdings aus der Sichtweise eines Beobachters in dem konstant beschleunigten System (man denke beispielsweise an eine Rakete, die konstant beschleunigt wird). Zunächst müssen wir die Zeit  $t$  in die Eigenzeit  $\tau$  des beschleunigten Beobachters umrechnen. Dazu verwenden wir die allgemeine Beziehung

$$d\tau = \sqrt{1 - \frac{v(t)^2}{c^2}} dt \quad (20)$$

und nutzen nun aus, dass

$$\sqrt{1 - \frac{v(t)^2}{c^2}} = \frac{1}{\sqrt{1 + \frac{(gt)^2}{c^2}}}. \quad (21)$$

Diese Beziehung folgt unmittelbar aus den beiden Gleichungen 9 und 10. Wir erhalten somit

$$\tau = \int_0^t \frac{1}{\sqrt{1 + \frac{(gt')^2}{c^2}}} dt' = \frac{c}{g} \sinh^{-1} \frac{gt}{c} \quad (22)$$

oder, aufgelöst nach  $t$ :

$$t = \frac{c}{g} \sinh \frac{g\tau}{c}. \quad (23)$$

Zwischen der verstrichenen Zeit  $t$  eines Beobachters im anfänglichen Ruhesystem (beispielsweise auf der Erde) und der Zeit  $\tau$  für einen Beobachter in dem konstant beschleunigten System besteht also für große Zeiten eine exponentielle Beziehung. Für die relativ zum anfänglichen Ruhesystem zurückgelegte Strecke als Funktion der Eigenzeit einer Person in dem beschleunigten System erhalten wir

$$x(\tau) = \frac{c^2}{g} \left( \cosh \frac{g\tau}{c} - 1 \right). \quad (24)$$

Diese Beziehung scheint zunächst den Möglichkeiten eines bemannten Raumflugs sehr entgegen zu kommen: Für die Reise zum rund 2 Millionen Lichtjahre entfernten Andromeda-Nebel (der nächsten großen Galaxie außerhalb der Milchstraße) würden bei einer konstanten Beschleunigung von  $g = 9,81 \text{ m/s}^2$  (der Erdbeschleunigung) für einen Astronauten in seinem Bezugssystem nur rund 28 Jahre vergehen.<sup>2</sup>

Dies widerspricht nicht der Relativitätstheorie: Die 2 Millionen Lichtjahre erfahren für den Beobachter in der Rakete eine (Lorentz-)Kontraktion auf weniger als 28 Lichtjahre. Das bedeutet aber, dieser Beobachter „sieht“ den Andromeda-Nebel mit Überlichtgeschwindigkeit auf sich zukommen.

Einen ähnlich erstaunlichen Effekt sieht der Beobachter auch, wenn er zurückblickt. Das Ereignis  $O$  bei  $(t = 0, x = -\frac{c^2}{g})$  bleibt für immer auf seiner Gleichzeitzigkeitslinie. Es wird zu einem Augenblick, der nie vergeht. Auch der Abstand zwischen ihm und diesem Ereignis bleibt in seinem Bezugssystem immer konstant. Allerdings sollte nochmals betont werden, dass sich diese Effekte auf ein augenblickliches globales Inertialsystem beziehen, und dies lässt sich für einen beschleunigten Beobachter nicht operational realisieren.

In Abschnitt 3 kommen wir nochmals auf den konstant beschleunigten Beobachter zu sprechen und beschreiben dort, was der Beobachter wirklich „sieht“.

## 2 Zwei konstant beschleunigte Systeme

Die bekannte Schriftensammlung „Speakable and unspeakable in quantum mechanics“ von John Bell ([2], Kapitel 9) enthält auch einen Artikel, der nichts mit Quantentheorie zu tun hat. Er trägt den Titel „How to teach special relativity“, und hier plädiert Bell dafür, Studierende der Physik auch mit der „alten“ Version der Speziellen Relativitätstheorie vertraut zu machen, wie sie von Larmor, Lorentz und Poincaré vertreten wurde und wie sie in Kapitel [Philosophischer Hintergrund der SRT](#) angedeutet wurde. Er beschreibt dort eine einfache Situation aus der Speziellen Relativitätstheorie, die seiner Ansicht nach in der „alten“ Sichtweise leichter nachvollzogen werden kann als in der Form, in der die Relativitätstheorie heute meist gelehrt wird.

Zwei gleichartige Raketen befinden sich zunächst in Ruhe und haben einen Abstand  $L$  (in diesem Ruhesystem haben sie die Koordinaten  $x_1 = 0$  und  $x_2 = L$ ). Sie seien durch ein Seil der Länge  $L$  miteinander verbunden. Ab dem Zeitpunkt  $t = 0$  erfahren beide Raketen dieselbe konstante Beschleunigung  $g$  in Richtung ihres Abstandsvektors (also die  $x$ -Richtung). Ihre Weltlinien sind somit Hyperbeln, deren Abstand im anfänglichen Ruhesystem (mit den Koordinaten  $(t, x)$ ) konstant bleibt (vgl. Abb. 2).

Bell stellt nun die Frage, ob das Seil zwischen den beiden Raketen irgendwann reißt. Ein solches Ereignis ist eine physikalische Tatsache und hängt daher nicht vom Bezugssystem ab. Anschei-

---

<sup>2</sup>Auf der Internetseite von John Baez [1] findet man eine sehr schöne Beschreibung der seltsamen Effekte einer konstant beschleunigten relativistischen Rakete.

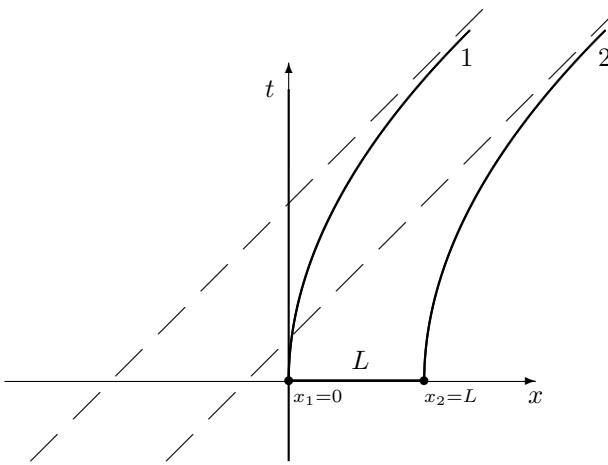


Abbildung 2: Zwei Raketen (1 und 2) erfahren dieselbe konstante Beschleunigung. Sie seien durch ein Seil miteinander verbunden, dessen Länge gerade dem anfänglichen Abstand  $L$  der Raketen entspricht. Wird das Seil reißen oder nicht?

nend hat er diese Frage in den 70er Jahren mehreren Physikern am CERN gestellt und sehr unterschiedliche Antworten erhalten (viele scheinen behauptet zu haben, das Seil reiße nicht). Tatsächlich reißt das Seil. Wir betrachten nun diese Situation aus allen drei Bezugssystemen – dem Inertialsystem, in dem die Raketen anfänglich in Ruhe sind, sowie den beiden Bezugssystemen der Raketen.

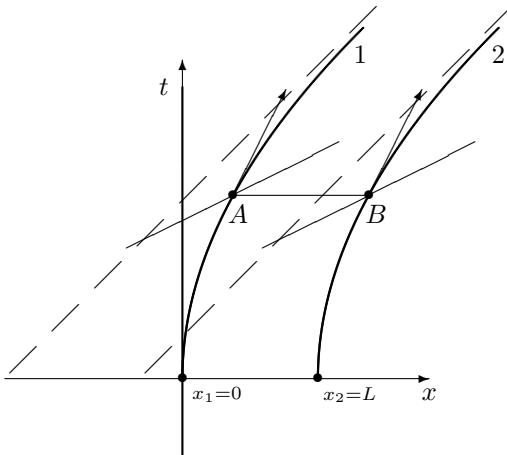


Abbildung 3: Die beiden Ereignisse  $A$  und  $B$  sind im ruhenden Inertialsystem gleichzeitig. Auch die Eigenzeiten der beiden Raketen sind bei diesen Ereignissen gleich. Der räumliche Abstand der Ereignisse ist  $L$ .

Wir beginnen unsere Betrachtungen im ruhenden Inertialsystem mit den Koordinaten  $(t,x)$ . Zwei in diesem System gleichzeitige Positionen der Raketen (z.B.  $A$  und  $B$ ) haben immer noch den räumlichen Abstand  $L$  (vgl. Abb. 3). Trotzdem bewegt sich das Seil mit einer bestimmten Geschwindigkeit relativ zu dem Ruhesystem (angedeutet durch die Vektorpfeile in Abb. 3). Damit sind die Ereignisse  $A$  und  $B$  für das Seil auch nicht gleichzeitig (auch die Gleichzeitigkeitslinien zu den beiden Ereignissen sind in der Abbildung angedeutet). Da sich das Seil bewegt, kommt es zu einer Lorentz-Kontraktion und das Seil wird reißen.

Bell bemerkte, dass damals viele Physiker in der Lorentz-Kontraktion nur eine scheinbare Verkürzung von Abständen sahen, weil man eine Länge aus verschiedenen Systemen mit unterschiedlichen Gleichzeitigkeitsvorstellungen ausmisst. Doch das Reißen des Seils ist eine Tatsache, kein „Scheineffekt“. Hier, so argumentiert er, gibt die Lorentz'sche Vorstellung einer tatsächlichen Verkürzung ein besseres Bild: Die Reichweite der elektromagnetischen Kräfte, die das Seil zusammenhalten, wird (ähnlich wie die Solitonen bei unserer linearen, gekoppelten Kette) kürzer, doch die

Atome müssen, da sie zwischen den Raketen eingespannt sind, ihren Abstand behalten. Irgendwann wird die Reichweite der Kräfte so klein, dass die Atome nicht mehr zusammengehalten werden können und das Seil reißt.

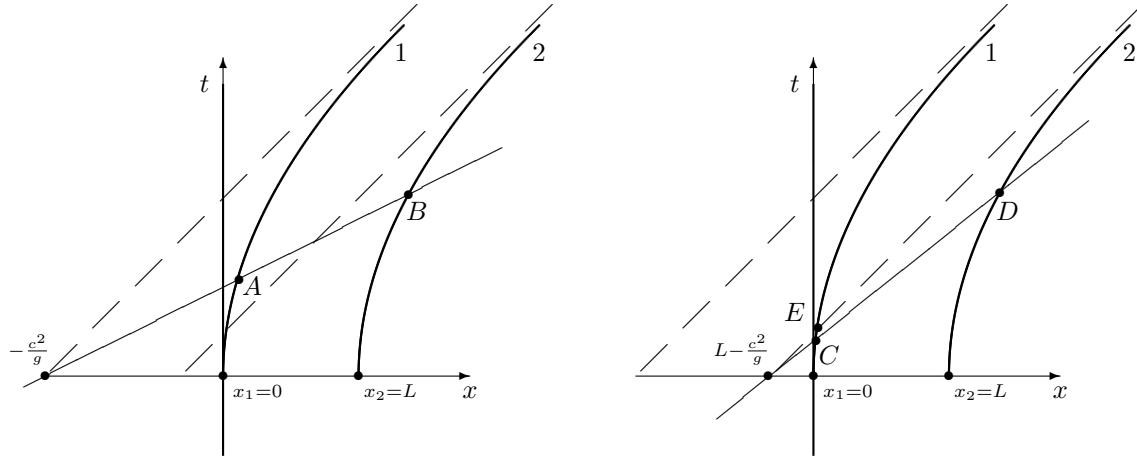


Abbildung 4: Die Gleichzeitsigkeitslinien für die beiden Raketen. (Links) Ereignis  $A$  befindet sich auf der Weltlinie von Rakete 1. In dem augenblicklichen Inertialsystem ist Ereignis  $B$  auf der Weltlinie von Rakete 2 gleichzeitig zu  $A$ . Doch bei  $B$  hat Rakete 2 bereits eine größere Geschwindigkeit als Rakete 1 bei Ereignis  $A$ , sodass der Abstand von Rakete 2 aus Sicht von Rakete 1 zunimmt. (Rechts)  $D$  ist ein Ereignis auf der Weltlinie von Rakete 2 und in dem augenblicklichen Inertialsystem ist Ereignis  $C$  auf der Weltlinie von Rakete 1 gleichzeitig zu  $D$ . Doch nicht nur ist die Geschwindigkeit von Rakete 1 bei  $C$  sehr viel langsamer als die von Rakete 2 bei  $D$ , sodass der Abstand zwischen den beiden Raketen aus der Sicht von Rakete 2 zunimmt, sondern Rakete 2 sieht Rakete 1 auch nie zu dem Ereignis  $E$  gelangen. Für Rakete 2 endet die Weltlinie von Rakete 1 an diesem Punkt.

Wir betrachten nun die Situation aus dem Bezugssystem von Rakete 1, also der hinteren Rakete bezüglich der Beschleunigungsrichtung (Abb. 4 (links)). Zu einem beliebigen Ereignis  $A$  auf der Weltlinie dieser Rakete kann der Beobachter in der Rakete zumindest mathematisch seine augenblickliche Gleichzeitsigkeitslinie konstruieren (er kann sie nicht im Sinne einer Einstein-Synchronisation operational realisieren): Das sind alle Ereignisse zu Vektoren, die bezüglich der Minkowski-Metrik senkrecht auf dem augenblicklichen 4-Vektor der Geschwindigkeit bzw. der augenblicklichen Tangente an die Weltlinie stehen. Danach ist Ereignis  $B$  auf der Weltlinie von Rakete 2 gleichzeitig zu Ereignis  $A$  (für Rakete 1). Doch bei  $B$  bewegt sich Rakete 2 bereits wesentlich schneller als Rakete 1 bei  $A$ . Das bedeutet, für Rakete 1 nimmt der Abstand zu Rakete 2 ständig zu. Somit wird das Seil auch irgendwann reißen.

Abschließend betrachten wir dieselbe Situation noch aus dem Bezugssystem von Rakete 2 (Abb. 4 (rechts)).  $D$  sei ein beliebiges Ereignis auf dieser Weltlinie, und ähnlich wie zuvor konstruiert der Beobachter in dieser Rakete die Gleichzeitsigkeitslinie zu diesem Ereignis. Er findet, dass in diesem augenblicklichen Inertialsystem Ereignis  $C$  auf der Weltlinie von Rakete 1 gleichzeitig zu  $D$  ist. Doch bei  $C$  bewegt sich Rakete 1 wesentlich langsamer als Rakete 2 bei  $D$ , daher nimmt auch in seinem Bezugssystem der Abstand zu Rakete 1 zu und das Seil wird reißen.

Wir beobachten hier aber noch etwas Weiteres: Alle Gleichzeitsigkeitslinien zur Weltlinie von Rakete 2 liegen (bei  $x$ -Koordinaten größer als  $L - \frac{c^2}{g}$ ) unterhalb des Lichtstrahls, der zur Asymptote der Bahnkurve von 2 wird. Das bedeutet, für Rakete 2 wird Rakete 1 niemals weiter als bis zu dem Ereignis  $E$  auf diesem Lichtstrahl gelangen, Rakete 1 wird dieses Ereignis *aus der Sichtweise von*

*Rakete 2* noch nicht einmal erreichen.

Im nächsten Abschnitt gehen wir auf diesen letztgenannten Punkt nochmals ein. Jedenfalls sind sich alle drei Beobachter darin einig, dass das Seil nach den physikalischen Gesetzen in ihrem Bezugssystem reißen muss. Man sollte aber in jedem Fall berücksichtigen, dass die Konstruktion einer „augenblicklichen Gleichzeitigkeitsfläche“ bei beschleunigten Weltlinien rein mathematische ist und sich physikalisch nicht realisieren lässt. Bei einem „ewigen“ Inertialsystem ist eine solche Konstruktion zumindest im Prinzip operational möglich.

### 3 Das Rindler-Universum

Wir haben gesehen, wie sich im Rahmen der Speziellen Relativitätstheorie bereits einige sehr interessante Effekte in der Physik eines konstant beschleunigten Beobachters untersuchen lassen. Dabei haben wir allerdings von globalen Gleichzeitigkeitslinien Gebrauch gemacht, die für einen Beobachter entlang einer Weltlinie nicht unbedingt von Relevanz sind. Beispielsweise kann bei beschleunigten Systemen die Folge solcher Gleichzeitigkeitslinien, selbst wenn sie entlang der Weltlinie in kausaler Reihenfolge konstruiert werden, Ereignisse in großem Abstand (hier definiert  $d = \frac{c^2}{g}$  die Skala) in kausal rückläufiger Reihenfolge überstreichen (man betrachte beispielsweise Ereignisse, die in Abb. 1 links von Ereignis  $O$  liegen). Aus diesem Grunde verwendet man auch in der Allgemeinen Relativitätstheorie solche globalen Gleichzeitigkeitslinien – in  $2+1$  Raumzeit-Dimensionen sind es natürlich Flächen und in  $3+1$  Raumzeit-Dimensionen Volumina – nur selten.

In diesem Abschnitt soll nochmals der konstant beschleunigte Beobachter betrachtet werden, allerdings mit den Methoden, die wir später bei verallgemeinerten Geometrien in der Allgemeinen Relativitätstheorie verwenden werden: (1) durch Angabe der kausalen Beziehungen und (2) durch die Untersuchung von Signalen, die zwischen Beobachtern auf verschiedenen Weltlinien ausgetauscht werden können. Die Raumzeit eines solchen konstant beschleunigten Beobachters bezeichnet man auch als Rindler-Universum. Aufgrund des Äquivalenzprinzips lassen sich viele der beobachteten Effekte qualitativ (und in manchen Einzelheiten sogar quantitativ) auf einen Beobachter in der Nähe eines schwarzen Loches übertragen. Von besonderer Bedeutung ist in diesem Zusammenhang das Konzept eines „Horizonts“.

#### 3.1 Kausale Beziehungen

Wie wir in Abschnitt 1.2 gezeigt haben, lässt sich im Raumzeit-Diagramm eines inertialen Beobachters  $B$  die Weltlinie eines konstant beschleunigten Beobachters  $A$  als Hyperbel darstellen. In diesem Abschnitt betrachten wir eine Weltlinie zu einem System, das seit „ewigen Zeiten“ einer konstanten Beschleunigung unterlag und auch für ewige Zeiten dieser Beschleunigung unterliegen wird (vgl. Abb. 5). Das System kommt also aus dem Unendlichen auf den inertialen Beobachter  $B$  zu und wird dabei konstant abgebremst bis es schließlich für einen Augenblick relativ zu dem inertialen Beobachter in Ruhe ist und sich nun mit derselben Beschleunigung wieder von dem inertialen Beobachter entfernt.

Der Minkowski-Raum des inertialen Beobachters  $B$  lässt sich durch die kausale Relationen der Ereignisse zu dem beschleunigten Beobachter  $A$  in vier Klassen einteilen:

- I Dieser Bereich enthält alle Ereignisse, die irgendwann einmal in der kausalen Zukunft des Beobachters  $A$  lagen und gleichzeitig irgendwann einmal in der kausalen Vergangenheit von  $A$  sein werden. Jedes der Ereignisse konnte von  $A$  einmal beeinflusst werden und kann umgekehrt einmal einen Einfluss auf  $A$  haben. Dieser Bereich entspricht also im üblichen Sinne der kausalen Raumzeit für Beobachter  $A$ .

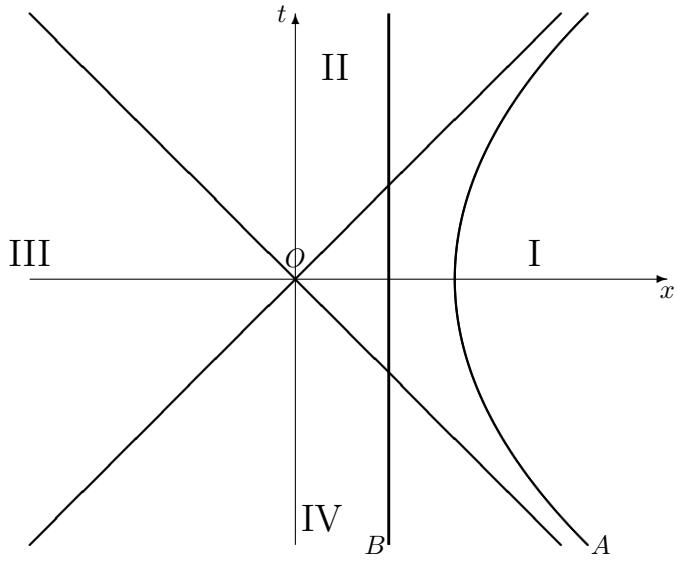


Abbildung 5: Rindler-Universum.  
Dargestellt sind die Weltlinien eines konstant beschleunigten Beobachters  $A$  und eines inertialen Beobachters  $B$ . Die Hyperbelbahn von Beobachter  $A$  definiert die angegebenen Lichtstrahlen sowie das Ereignis  $O$ . Die Quadranten I—IV stehen jeweils in einer besonderen kausalen Beziehung zu Beobachter  $A$ .

- II Dieser Bereich enthält alle Ereignisse, die in der kausalen Zukunft von Ereignissen auf der Weltlinie von  $A$  liegen, aber nicht in der kausalen Vergangenheit irgendeines Ereignisses auf der Weltlinie von  $A$ . Der Beobachter  $A$  kann diesen Bereich also nicht „einsehen“ bzw. er kann von den Ereignissen in diesem Bereich nie kausal beeinflusst werden, er kann aber umgekehrt die Ereignisse in diesem Bereich kausal beeinflussen.
- III Die Ereignisse in diesem Bereich haben keinen kausalen Zusammenhang – weder in der Zukunft noch in der Vergangenheit – zu irgendeinem Ereignis auf der Weltlinie von Beobachter  $A$ .
- IV Alle Ereignisse in diesem Bereich liegen irgendwann einmal in der kausalen Vergangenheit von  $A$ , waren aber niemals in seiner kausalen Zukunft.  $A$  kann von diesen Ereignissen also kausal beeinflusst werden, hat aber umgekehrt keinen Einfluss auf sie.

Hinsichtlich der Kausalbeziehungen ist für den beschleunigten Beobachter  $A$  die Ereignismenge in Bereich I so, wie für einen inertialen Beobachter die Ereignismenge der Minkowski-Raum-Zeit: Zu jedem Ereignis gibt es auf seiner Weltlinie einen Zeitpunkt in der Vergangenheit, *vor* dem dieses Ereignis in seiner kausalen Zukunft lag, es kann also durch diesen Beobachter beeinflusst werden. Ebenso gibt es zu jedem Ereignis einen Zeitpunkt, *ab* dem der Beobachter in der kausalen Zukunft des Ereignisses liegt, d.h. dieses Ereignis wahrnehmen bzw. von ihm Kenntnis erlangen kann.

Alle anderen Bereiche haben für einen inertialen Beobachter in einer Minkowski-Raum-Zeit kein Gegenstück. Die Ereignisse in den Bereichen III und IV liegen beispielsweise niemals in der kausalen Zukunft von  $A$ . Der beschleunigte Beobachter hat somit auch keine Möglichkeit, diese Ereignisse jemals zu beeinflussen. Allerdings kann er die Ereignisse aus Bereich IV wahrnehmen bzw. kausal von ihnen beeinflusst werden, da er sich irgendwann in der kausalen Zukunft von diesen Ereignissen befinden wird. Der Bereich III gehört zu einem Teil des Universums, der mit  $A$  überhaupt keine kausale Verbindung hat, weder in der Zukunft noch in der Vergangenheit. In gewisser Hinsicht existiert dieser Bereich für den beschleunigten Beobachter  $A$  gar nicht. Die Ereignisse von Bereich II können zwar von  $A$  beeinflusst werden, allerdings kann  $A$  diesen Bereich ebenfalls nie einsehen.

### 3.2 Was „sehen“ die Beobachter voneinander?

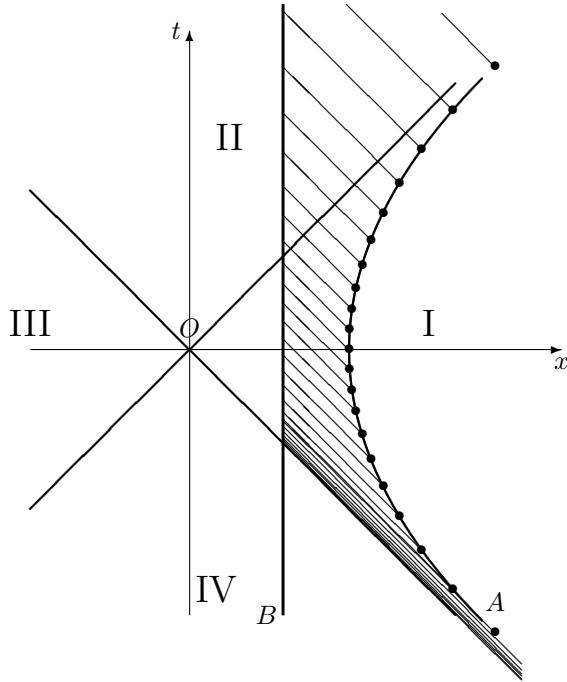


Abbildung 6: Was sieht der inertielle Beobachter  $B$  von dem beschleunigten Beobachter  $A$ ? In regelmäßigen Eigenzeitabständen (Ereignisse auf der Weltlinie von  $A$ , gekennzeichnet durch „•“) sendet der beschleunigte Beobachter Signale aus.  $B$  empfängt diese Signale in seinem System in unterschiedlichen Zeitabständen.

Wir überlegen uns zunächst, was der inertielle Beobachter  $B$  von dem beschleunigten Beobachter  $A$  „sieht“. In Abbildung 6 sind beide Weltlinien dargestellt, zusätzlich sind in gleichmäßigen Eigenzeitabständen von  $A$  Ereignisse markiert, bei denen  $A$  ein Lichtsignal zu Beobachter  $B$  aussendet.

Solange  $B$  sich in Bereich IV befindet, hat er keinerlei Kenntnisse von  $A$ . Erst beim Überscheiten der Grenze zwischen Bereich IV zu Bereich I „erfährt“ Beobachter  $B$  von  $A$ . Das geschieht allerdings gleich sehr heftig: Innerhalb einer beliebig kurzen Zeit nimmt Beobachter  $B$  eine unendliche Zeitspanne in der Vergangenheit von Beobachter  $A$  wahr.

Hinsichtlich seiner Wahrnehmung sieht Beobachter  $B$  den beschleunigten Beobachter  $A$  in einer beliebig kurzen Zeit eine unendliche Strecke auf ihn zukommen. Diese scheinbare Wahrnehmung widerspricht natürlich nicht der Aussage, dass die Lichtegeschwindigkeit eine Grenzgeschwindigkeit darstellt. Da Beobachter  $B$  in beliebig kurzer Zeit eine unendliche Vergangenheit von  $A$  wahrnimmt, sind die eintreffenden Lichtwellen auch unendlich blau-verschoben. In gewisser Hinsicht ist dieses Ereignis für Beobachter  $B$  wie eine Singularität.

Solange sich der inertielle Beobachter  $B$  im Bereich I befindet, kann er mit dem beschleunigten Beobachter  $A$  Information austauschen. Für Beobachter  $B$  ändert sich auch nicht viel, wenn er in den Bereich II tritt. Für ihn hat die Grenze zwischen Bereich I und II keinerlei Bedeutung und er hat an dieser Grenze auch keine besondere Wahrnehmung. Er kann den beschleunigten Beobachter  $A$  bis in eine beliebige Zukunft weiter beobachten. Allerdings werden die Zeitabstände zwischen Signalen, die  $A$  in gleichen Eigenzeitabständen aussendet, für  $B$  immer größer. Der inertielle Beobachter  $B$  sieht also den beschleunigten Beobachter  $A$  immer stärker rot-verschoben. Diese Rotverschiebung entspricht im Wesentlichen dem Doppler-Effekt eines sich zunehmend rasch entfernenden Senders.

Nun untersuchen wir, was der beschleunigte Beobachter  $A$  von dem inertiellen Beobachter  $B$  sieht. In Abb. 7 sind wieder die beiden Weltlinien dargestellt, diesmal sendet aber  $B$  in regelmäßigen Abständen Lichtsignale aus. Auf der Weltlinie von  $A$  sind in äquidistanten Eigenzeitabständen Er-

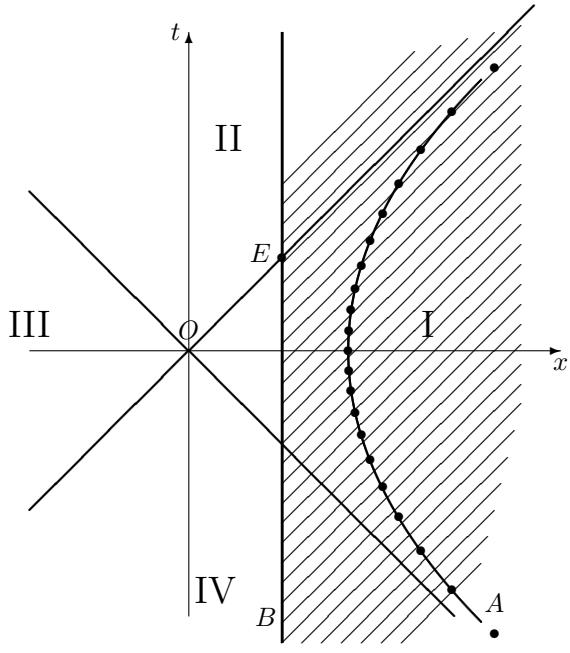


Abbildung 7: Was sieht der beschleunigte Beobachter  $A$  von dem inertialen Beobachter  $B$ ? Nun sendet  $B$  in regelmäßigen Eigenzeitabständen Signale aus, die  $A$  empfängt. Wegen der unterschiedlichen relativen Geschwindigkeit sowie zusätzlich den unterschiedlichen Skalen für die Eigenzeiten von  $A$  relativ zu  $B$  (Ereignisse in gleichen Eigenzeitabständen wurden auf der Weltlinie von  $A$  wieder gekennzeichnet) empfängt auch  $A$  die Signale in unterschiedlichen Zeitabständen.

eignispunkte eingetragen – der Eigenzeitabstand entspricht dem zeitlichen Abstand, mit dem  $B$  die Lichtsignale abschickt.

In beliebig ferner Vergangenheit kann  $A$  den Beobachter  $B$  schon wahrnehmen, allerdings treffen die Lichtsignale bezüglich seiner Eigenzeit sehr viel rascher bei ihm ein, sodass er Beobachter  $B$  blau-verschoben wahrnimmt, wiederum wie bei einem Doppler-Effekt. Je weiter man in die Vergangenheit von  $A$  zurückgeht, umso blauverschobener sieht  $A$  den inertialen Beobachter  $B$ . Sobald  $B$  auch in den Bereich I eingedrungen ist, können  $A$  und  $B$  Signale austauschen und sich gegenseitig verstündigen.

Wenn die beiden Weltlinien von  $A$  und  $B$  für einen Augenblick parallel sind, sieht  $A$  die Signale von  $B$  mit derselben Frequenz, wie  $B$  sie aussendet. Ab dann empfängt  $A$  die Signale seltener, d.h.  $A$  sieht  $B$  rotverschoben.

Offenbar kann  $A$  keinerlei Signale mehr aus dem Bereich II empfangen. Das bedeutet, kein Signal, das  $B$  nach dem Ereignis  $E$  verschickt, wird  $A$  jemals erreichen. Für  $A$  endet die Weltlinie von  $B$  an dem Ereignis  $E$ . Der Lichtstrahl, der die Bereiche I und II trennt, ist für Beobachter  $A$  ein Horizont, hinter den er nicht blicken kann. Man bezeichnet diesen Horizont auch als *Ereignishorizont*.

Der Beobachter  $A$  sieht den Beobachter  $B$  aber nicht einfach hinter dem Horizont verschwinden. Im Gegenteil: Er kann für alle Zukunft den Beobachter  $B$  wahrnehmen, wie er sich immer mehr dem Horizont bzw. dem Ereignis  $E$  nähert. Die Abstände, mit denen  $A$  aber von  $B$  die in gleichen Zeitabständen ausgesandten Signale erhält, werden immer größer.  $A$  nimmt die Zeit im  $B$ -System immer langsamer wahr. Damit ist eine Rotverschiebung der Strahlung verbunden.  $B$  verschwindet also nicht hinter dem Horizont, sondern  $B$  verschwindet an der Oberfläche des Horizonts im langwelligen Bereich des Spektrums.

Die Grenzen zwischen den Bereichen I und IV einerseits und den Bereichen I und II andererseits verhalten sich also in gewisser Hinsicht symmetrisch:  $A$  kann die Ereignisse in Bereich IV wahrnehmen, nicht aber die Ereignisse in Bereich II. Umgekehrt hat  $B$  keine Kenntnis von  $A$ , solange er sich in Bereich IV befindet, er nimmt das Schicksal von  $A$  aber durchaus wahr, wenn er sich in Bereich II befindet.

Der Bereich III gehört zu einem Teil des Universums, der mit  $A$  überhaupt keine kausale Verbindung hat, weder in der Zukunft noch in der Vergangenheit. In gewisser Hinsicht existiert dieser Bereich für den beschleunigten Beobachter  $A$  gar nicht. Trotzdem ist dieser Bereich für den inertialen Beobachter  $B$  ein ganz normaler Teil seines Universums. Da andererseits  $B$  von diesem Bereich auch erst erfährt, nachdem er den Horizont zwischen I und II durchschritten hat, kann er  $A$  keine Mitteilung davon machen.

## Literatur

- [1] Baez, John; [math.ucr.edu/home/baez](http://math.ucr.edu/home/baez), und speziell für Physik [math.ucr.edu/home/baez/physics](http://math.ucr.edu/home/baez/physics)
- [2] John Bell; *Speakable and Unspeakable in Quantum Physics*, 2. edition, Cambridge University Press (2004).