

Romanos Physikkurs I

Geometrische Optik
Mechanik
Thermodynamik
Elektrizität

Romano A. Rupp
Universität Wien
Wintersemester 2017

September 12, 2018

© Romano A. Rupp. Alle Rechte vorbehalten, auch das des auszugsweisen Nachdrucks, der auszugsweisen oder vollständigen Wiedergabe, der Speicherung in Datenverarbeitungsanlagen und der Übersetzung. Für Nutzungsanfragen wenden Sie sich an Romano A. Rupp (romano.rupp@univie.ac.at).

Vorwort

Die mathematischen Anforderungen sind zu Anfang dieses Physikkurses bewusst gering gehalten. Dadurch wird Studienanfängern Zeit gegeben, Lücken in ihren mathematischen Kenntnissen durch die meist parallelen mathematischen Vorlesungen des ersten Semesters zu schließen. Durch dieses Konzept hoffe ich, der typischen Frustration jener Physikanfänger vorzubeugen, die von Schultypen herkommen, an denen die mathematische Ausbildung nicht im Vordergrund stand. Mit dem Fortschreiten des physikalischen Stoffes werden zunehmend mathematische Kenntnisse benötigt, die das Schulwissen übersteigen. Daher wurden einige erforderliche Kenntnisse in den mit M# (wobei # eine fortlaufende Nummer ist) gekennzeichneten mathematischen Ergänzungskapitel knapp zusammengefasst.

Der Einführungskurs legt von Anfang an Wert darauf, die großen physikalischen Konzepte logisch herzuleiten und ihre Grundlagen transparent darzulegen. Dieses Ziel bringt es mit sich, dass die Reihenfolge des Stoffes von derjenigen vieler Physiklehrbücher abweicht. Üblicherweise stellen diese die Newtonsche Mechanik an den Anfang. Stattdessen beginne ich mit der Geometrischen Optik, denn sie setzt als mathematisches Rüstzeug nur Kenntnisse der Geometrie auf dem Niveau der Mittelschule voraus. Sie eignet sich hervorragend, um die grundlegenden Methoden der Physik (z.B. Hypothesenbildung, Erstellung abstrakter Modelle, Umgang mit Messunsicherheiten) zu veranschaulichen. Auch aus der historischen Perspektive heraus ist es richtig, mit der Geometrischen Optik zu beginnen. Die hier erworbenen Methoden der geometrischen Argumentation und der linearen bzw. projektiven Transformation legen den Grundstein für Teil II, der in die Theorie von Raum und Zeit bzw. die Spezielle Relativitätstheorie einführt. Sie wird nicht aus den Einsteinschen Postulaten abgeleitet, sondern aus drei grundlegenden Axiomen über Raum und Zeit.

Der dritte und vierte Teil des Buches befasst sich mit der Mechanik. Aus Gründen der Vereinfachung und Fokussierung auf die wesentlichen Aspekte, wird sie eindimensional dargeboten, dafür aber von Anfang an relativistisch. Sie wird analog zur traditionellen Reihenfolge mit der Kinematik eröffnet. Grundlage für die daran anschließende Dynamik ist der Begriff der Erhaltungsgröße. Die zentralen Erhaltungsgrößen der Mechanik, Energie und Impuls, werden auch für die Ausarbeitung der nichtrelativistischen Mechanik konsequent zugrunde gelegt. Damit kommt dem Kraftbegriff auch nicht mehr die zentrale Rolle zu, die er in der Newtonschen Mechanik einnimmt.

In Teil V werden einige ausgewählte Themen aus der Kontinuumsphysik der Materie mit dem Ziel besprochen, den Boden für die nachfolgende Einführung der Thermodynamik zu bereiten. Dem Physikanfänger werden die für die Thermodynamik typische Betrachtungsweise und theoretische Methodik anhand der ihm nunmehr vertrauten Mechanik

präsentiert. Das zentrale Problem, mit dem sich die Thermodynamik beschäftigt, ist die Beschreibung der in makroskopischen Systemen beobachteten Irreversibilität bestimmter Prozesse. In Teil VI wird mit der thermischen Energie erstmals eine nicht-mechanische Energieform eingeführt und an diesem Beispiel das Muster gelehrt, wie neue Energieformen theoretisch aufgestellt werden. Auch hier weiche ich vom traditionellen Weg ab, der zunächst Temperatur und Wärme bespricht. Stattdessen führe ich von Anfang an die Entropie ein, weil diese extensive Zustandsvariable der thermischen Energieform der Schlüsselbegriff der Thermodynamik ist. Im Anschluss werden mit der chemischen und elektrischen Energie zwei weitere Beispiele für nicht-mechanische Energieformen vorgestellt. In Teil VII wird die Erweiterung der Mechanik auf zwei bzw. drei Raumdimensionen sowie die Einführung vektorieller Größen vorgenommen. Dabei werden wesentliche Aspekte herausgearbeitet, ohne die man die in Teil VIII dargestellten Transportprozesse und die Theorie klassischer Felder nicht verstehen kann.

Aus Kostengründen ist diese Druckversion in schwarz-weiß-rot gehalten. Eine Farbversion steht zum Herunterladen zur Verfügung. Ich verwende durchgehend den von der ETH Zürich auch für den deutschen Sprachraum empfohlenen *Dezimalpunkt* und nicht das nur in Deutschland und Österreich übliche Dezimalkomma. Alle auftretenden Gattungs- oder Berufsbezeichnungen sind grundsätzlich geschlechtsneutral zu verstehen.

Wien, den 6. November 2017

Romano Rupp

Danksagung

An vorderster Stelle möchte ich der Fakultät für Physik an der Universität Wien danken, die den Mut hatte, mir die Freiheit zu geben, diesen unkonventionellen Kurs den Physikanfängern der ersten beiden Semester vorzutragen und damit auch die Möglichkeit, ihn aus der lehrenden Praxis heraus zu entwickeln. In den ersten Jahren, in denen ich die Vorlesung gehalten habe, war der Kurs eine permanente Baustelle, denn ich hatte kein Vorbild und kein Lehrbuch, auf das ich mich stützen konnte (von der Richard-Feynman-Vorlesung, die mir ein Vorbild war, einmal abgesehen), und so möchte ich auch meinen Studenten danken, welche die Geduld aufgebracht haben, meinen nicht immer ausgereiften Ideen und den vielen Änderungen zu folgen.

Ich danke Bogdan Pammer, Herbert Hartl, Angelika Ambrusch und Ewa Partyka-Jankowska, die mich bei der Anfertigung der Bilder unterstützt haben. Allen, die dazu beigetragen haben Tippfehler, Schlampereien oder sachliche Fehler auszumerzen, möchte ich an dieser Stelle ebenfalls danken, insbesondere Dido Massimo, Heinz Kabelka, Gerhard Krexner, Lukas Perner, Franz Embacher und Wolfgang Lucha.

Inhaltsverzeichnis

I.	Geometrische Optik	1
1.	Die physikalische Methode	1
1.1.	Bildung von Hypothesen	1
1.2.	Mathematisches Schließen	4
1.3.	Quantifizierung	7
1.4.	Näherung	11
1.5.	Messung	13
1.5.1.	Präparation	13
1.5.2.	Vermessung der Welt in der Antike	14
1.5.3.	Vermessung der Welt in der frühen Neuzeit	15
1.5.4.	Messunsicherheit	17
1.6.	Präzisierung	22
1.6.1.	Bezugselemente und Konventionen für Winkel	24
1.6.2.	Reflexionsgesetz	24
1.6.3.	Brechungsgesetz	25
1.6.4.	Brechung am Prisma	26
1.6.5.	Das Snellius'sche Brechungsgesetz	27
2.	Bilder	31
2.1.	Das physikalische Bild	31
2.2.	Schattenwurf und Lochkamera	33
2.3.	Planspiegel	35
2.4.	Theorie der perfekten strahlenoptischen Abbildung	37
3.	Sphärische optische Elemente	42
3.1.	Sphärische Spiegel	43
3.2.	Linsen	45
4.	Optische Instrumente	47
4.1.	Theorie linearer optischer Systeme	48
4.1.1.	Systemcharakterisierung	49
4.1.2.	Linsensysteme	50
4.2.	Auge	52
4.3.	Brille und Lupe	53
4.4.	Fernrohr und Mikroskop	54
4.5.	Petzval's Revolution der Geometrischen Optik	56
II.	Kinematik	58
5.	Raum	58
5.1.	Homogenität	58
5.2.	Isotropie	59
5.3.	Bezugssysteme	59
5.4.	Koordinatensysteme	59
5.4.1.	Kreis- und Kugelkoordinaten	60
5.4.2.	Händigkeit	61
5.5.	Lineare Koordinatentransformationen	62

6.	Zeit	64
6.1.	Lokale Zeit	65
6.2.	Geschwindigkeit und Uhrensynchronisation	66
6.3.	Orts-Zeit-Diagramme	67
7.	Äquivalenzprinzipien	68
7.1.	Äquivalenzprinzip für invertierte Bezugssysteme	68
7.2.	Äquivalenzprinzip für gedrehte Bezugssysteme	69
7.3.	Galileisches Äquivalenzprinzip	71
8.	Spezielle Relativitätstheorie	72
8.1.	Die Axiome einer Theorie von Raum und Zeit	73
8.2.	Koordinatentransformationen zwischen Galileisystemen	73
8.3.	Die Raum/Zeit-Konstante c	77
8.4.	Definition der Längeneinheit Meter	79
8.5.	Invarianz der Ereignisintervalle	79
8.6.	Zeitdilatation	80
9.	Analysis (M1)	82
9.1.	Ortsableitung	82
9.2.	Zeitableitung	83
9.3.	Stammfunktion	83
10.	Geschwindigkeit und Beschleunigung	84
10.1.	Momentangeschwindigkeit	84
10.2.	Geradlinige beschleunigte Bewegung	84
10.3.	Relativistische Transformation der Beschleunigung	85
III.	Dynamik	87
11.	Grundbegriffe der Dynamik	87
11.1.	Extensive Zustandsgrößen und Erhaltungsgrößen	90
11.2.	Konstruktion der grundlegenden Erhaltungsgröße der Mechanik	94
11.3.	Einheitenkonversionen mit Naturkonstanten	95
11.4.	Wie misst man die Masse?	96
11.5.	Inertialsysteme	98
11.6.	Die Axiome der Dynamik	99
12.	Mechanische Energien	102
12.1.	Kinetische Energie	102
12.2.	Stoßprozesse	103
12.2.1.	Elastischer Stoß	103
12.2.2.	Inelastischer Stoß	104
12.2.3.	Zerfall eines physikalischen Systems	105
12.3.	Die mechanischen Grenzfälle	106
12.3.1.	Relativistischer Grenzfall	106
12.3.2.	Newtonscher Grenzfall	106
12.4.	Potentielle Energie	107
12.4.1.	Das Hookesche Gesetz	109
12.4.2.	Das Schweregesetz	111
12.5.	Dissipationskinetik	113

12.6.	Leistung und Energiefluss bei offenen Systemen	116
12.7.	Platons Höhlengleichnis und die Phänomenologie	116
13.	Differentialgleichungen (M2)	118
14.	Bewegungsgleichungen mechanischer Systeme	119
14.1.	Bewegungsverlauf bei Dissipation	121
14.2.	Freier Fall	121
14.3.	Schwingungen	124
14.3.1.	Idealer harmonischer Oszillator	124
14.3.2.	Die Schwingungsgleichung	127
14.3.3.	Gedämpfter harmonischer Oszillator	128
14.3.4.	Energiezufuhr von außen	129
14.3.5.	Gekoppelte Oszillatoren	132
15.	Kraft	133
15.1.	Newtonsche Bewegungsgleichung	133
15.2.	Actio und reactio	137
16.	Arbeit	138
17.	Das Prinzip des Minimums der mechanischen Energie	139
18.	Kontinuumsmechanik der Materie	141
18.1.	Materiemodelle	141
18.2.	Stoffmenge und Dichte	142
18.3.	Auftrieb	143
18.4.	Elastizität	143
18.4.1.	Dehnung	145
18.4.2.	Reine Formänderungen	145
18.5.	Elastische Volumendeformationen	146
18.5.1.	Druck	146
18.5.2.	Druckausgleich	147
18.5.3.	Weg-Kraft-Transformation	148
18.5.4.	Kompressibilität	149
18.6.	Enthalpie	150
18.7.	Grenzflächenphänomene	152
18.7.1.	Grenzflächenenergie	152
18.7.2.	Haftung	155
18.7.3.	Dissipation an Grenzflächen	156
19.	Funktionen mehrerer Variabler (M3)	158
19.1.	Partielle Ableitung	158
19.2.	Potential	159
20.	Systeme mit mehreren Freiheitsgraden	159
20.1.	Zustandsdiagramme	160
20.2.	Extensive und intensive Größen	161
20.3.	Fundamentalfom der inneren Energie	162
20.4.	Konjugierte Variable	163
20.5.	Warum und wie werden neue Zustandsvariable eingeführt?	165

IV. Phänomenologische Thermodynamik	167
21. Die Hauptsätze der Thermodynamik	167
22. Physiker und ihre Verbotstafeln	171
23. Arbeit und Wärme	173
24. Das Standardmodell der Thermodynamik	175
25. Thermisches Gleichgewicht	177
26. Thermometrie	179
26.1. Praktische Thermometrie.	179
26.2. Definition der Temperatureinheit	181
27. Revision von Begriffen aus der Mechanik	184
27.1. Thermodynamisches Druckgleichgewicht	184
27.2. Enthalpie im Standardmodell	184
28. Thermische Materialeigenschaften	186
28.1. Nomenklatur thermodynamischer Prozesse	186
28.2. Thermische Ausdehnung	186
28.3. Expansionskoeffizient und Wärmekapazitäten	187
29. Umwandlung von Arbeit in Wärme	190
30. Allgemeine Zustandsgleichung	191
30.1. Van-der-Waals Zustandsgleichung für Fluide	191
30.2. Die Zustandsgleichung verdünnter Gase	192
31. Adiabatische Prozesse	192
32. Wärmepumpen und Wärmekraftmaschinen	194
33. Gibbs-Energie	198
34. Chemische Energie	199
34.1. Heizwert	200
34.2. Grundgesetze der Chemie	200
34.3. Der chemische Stoffmengenbegriff	201
34.4. Das ideale Gas	202
34.5. Chemisches Potential	203
34.6. Phasenumwandlungen und chemische Reaktionen	205
35. Elektrische Energie	208
35.1. Elektrizität	208
35.2. Elektrische Energie	210
35.3. Ladungserhaltung	211
35.4. Thermoelektrische Energieumwandlungen	212
35.4.1. Richtungsvorgabe für den Ladungstransport	212
35.4.2. Potentialausgleich	213
35.4.3. Elektrische Heizung	214
35.4.4. Thermoelektrische Generatoren und Thermoelemente	215
35.5. Elektrochemische Energieumwandlungen	216
35.5.1. Galvanische Zelle	216
35.5.2. Elektrolyse	217
35.6. Elektrische Maßeinheiten	218

35.7. Kapazität	220
35.7.1. Metallkugel	221
35.7.2. Kondensator	222
35.7.3. Korrektur elektrischer Messungen.	224
35.7.4. Elektrische Energie eines Kondensators	225
36. Strom	226
37. Widerstand	229
37.1. Messung elektrischer Widerstände.	232
37.2. Messbereichserweiterung	234
V. Mechanik im euklidischen Raum	235
38. Lineare Algebra (M1)	235
38.1. Vektoren	235
38.2. Skalarprodukt	236
38.3. Lineare Transformationen	237
38.4. Vektoren im euklidischen Raum	237
38.4.1. Die Richtungsbasis	237
38.4.2. Drehungen	238
38.4.3. Orientierung bzw. Händigkeit	239
38.4.4. Kreuzprodukt	240
38.4.5. Ableitung von Vektorfunktionen	241
39. Relativitätstheorie im dreidimensionalen Raum	241
39.1. Beschränkung aller Geschwindigkeiten durch c	241
39.2. Ulks	242
39.3. Warum ist der isotrope homogene Raum euklidisch?	243
39.4. Relativistische Kinematik und Dynamik in drei Dimensionen	243
40. Bewegung im Raum	244
40.1. Bewegung im zweidimensionalen Raum	244
40.2. Bewegung im dreidimensionalen Raum	245
41. Dreidimensionale Newtonsche Dynamik	245
41.1. Newtonscher Impuls	246
41.2. Newtonscher Drehimpuls und Drehmoment	247
42. Winkelgeschwindigkeit und -beschleunigung	248
42.1. Gleichförmige Kreisbewegung in zwei Dimensionen	248
42.2. Kreisbewegung in drei Dimensionen	250
42.3. Pseudoskalare und polare Vektoren	251
42.4. Verhalten physikalischer Größen unter Drehung und Inversion	252
42.5. Die Vektoren der Winkelgeschwindigkeit und -beschleunigung	253
43. Gravitations- und Trägheitskräfte in der Newtonschen Physik	253
43.1. Das Gravitationsgesetz	253
43.2. Alle Körper fallen gleich schnell	255
43.3. Trägheitskräfte	256
43.4. Trägheitskräfte bei Drehbewegungen	257
43.5. Das Äquivalenzprinzip	259
43.6. Symmetrien	260

44. Nichtlineare Koordinatentransformationen (M4)	260
45. Geometrie eines Körpers	264
46. Drehung makroskopischer Körper	266
46.1. Schwerpunktsystem: Schwerpunkt und Eigendrehimpuls	266
46.2. Rotationsenergie	267
VI. Anhang	270
A. Newtons Hypothese der absoluten Zeit	270
B. Probleme des intergalaktischen Handels und deren künftige Lösung	271
C. Das Zwillingsparadoxon	273
D. Relativistische Federn	276
E. Legendre-Transformation	277
F. Hamilton- und Lagrangefunktion	279

Teil I.

Geometrische Optik

1. Die physikalische Methode

1.1. Bildung von Hypothesen

Das Phänomen Licht hat Menschen schon seit jeher beschäftigt. Die Thora, Grundlage des jüdischen Glaubens, geht bereits im dritten Satz auf das Licht ein:

וַיֹּאמֶר אֱלֹהִים יְהִי אֹר וַיְהִי־אֹר:

Gott sprach, es werde Licht, und es wurde Licht

וַיֵּרָא אֱלֹהִים אֶת־הָאֹר כִּי־טוֹב וַיַּבְדֵּל אֱלֹהִים בֵּין הָאֹר וּבֵין הַחֹשֶׁךְ:

Gott sah, dass das Licht gut war,
und Gott schied das Licht von der Finsternis

וַיִּקְרָא אֱלֹהִים | לְאֹר יוֹם וְלַחֹשֶׁךְ קָרָא לַיְלָה וַיְהִי־עֶרֶב וַיְהִי־בֹקֶר יוֹם אֶחָד:

Gott nannte das Licht Tag und die Finsternis Nacht,
und so wurde Morgen und wurde Abend: der erste Tag

Die Sätze enthalten Definitionen und Behauptungen. Sie setzen voraus, dass die in ihnen auftretenden Begriffe (wie *Gott*, *Licht*, *gut* usw.) für den Rezipienten der Botschaft selbstverständlich und klar sind. Es wird Menschen geben, die den Glaubenssätzen einer religiösen Gemeinschaft vertrauen. Dann sind sie Gläubige dieser Religion. Im anderen Fall werden sie als Ungläubige angesehen und wenn sie gar den Glauben an einen Gott als Basisannahme nicht akzeptieren, als Gottesleugner. Manche Religionen verlangen von ihren Gläubigen eine bedingungslose Unterwerfung unter die Glaubensaussagen und -traditionen und ahnden den Abfall vom Glauben mit Strafen bis hin zum Ausschluss aus der sozialen Gemeinschaft. Religiöse Aussagen werden als unverrückbar und wahr angenommen. Eine Erwägung anderer Aussagen und erst recht ein Akzeptieren anderer Aussagen steht nicht zur Disposition. Manche Religionen erheben einen Alleinvertretungsanspruch der Wahrheit, womit alle anderen Überzeugungen – egal welche – von vorne herein als falsch gebrandmarkt sind.

Die Grundhaltung jener Menschen, die sich der Wissenschaft widmen wollen, ist eine andere. Zunächst einmal das Gemeinsame: Sowohl den Glaubensaussagen als auch den als HYPOTHESEN bezeichneten wissenschaftlichen Annahmen und Behauptungen liegen i.A. Naturerfahrungen und Beobachtungen zugrunde. Hypothesen stehen jedoch grundsätzlich unter skeptischem Vorbehalt. Damit meint man, dass sie so lange als gültig angesehen werden, wie ihnen nichts Überzeugendes widerspricht. Einer Hypothese wird Offenheit zugebilligt und Tragfähigkeit abverlangt. Offenheit bedeutet, dass es jederzeit möglich ist, auch andere Hypothesen zu erwägen. Es ist jederzeit möglich, eine Hypothese zu verwerfen und einen anderen Standpunkt einzunehmen. Man vertraut darauf, dass sich nützliche und tragfähige Hypothesen im wissenschaftlichen Diskurs durchsetzen werden, dass sich andere Wissenschaftler von ihnen durch Argumente und Einsicht überzeugen lassen. Tragfähigkeit bedeutet, dass eine Hypothese weitreichende Schlüsse ermöglicht, die empirisch überprüfbar sind, vielfach überprüft wurden und sich dabei bewährt haben. Die Grundlage einer wissenschaftlichen THEORIE besteht i.A. aus mehreren Hypothesen. Sie bilden die Postulate bzw. Axiome der Theorie und dürfen sich nicht widersprechen. Auch die daraus gezogenen Folgerungen dürfen ebenfalls nicht in einem logischen Widerspruch zueinander stehen. Es kann sein, dass eines der Postulate einer Theorie oder einer der aus den Postulaten folgenden Schlüsse durch erweiterte empirische Erfahrung falsifiziert wird. Auch wenn nur ein einziges Postulat empirisch widerlegt ist, muss die darauf basierende Theorie verworfen werden. In der Regel wird man die alte Theorie dabei nicht in Bausch und Bogen verdammen, weil sich die ursprüngliche empirische Erfahrung, auf denen ihre Hypothesen beruhten, ja nicht geändert haben. Im Rahmen der durch die neue Theorie geklärten Voraussetzungen, unter denen die ursprüngliche Erfahrung möglich war, bestehen auch ältere Theorien fort, wenn sie für gewisse Einsichten nützlich bleiben.

Die Erfahrung des Menschen mit Quellen des Lichtes beschränkte sich ursprünglich auf Sonne, Mond, Sterne und das Feuer. Mit dieser Erfahrung eng verbunden war die Erfahrung von Dunkelheit: Wo Licht ist, ist auch Schatten, und tritt man in den Schatten der Erde ein, so wird es Nacht. Heute stehen uns viel bequemere Lichtquellen zum Experimentieren zur Verfügung: z. B. Glühlampen, Leuchtdioden und Laser. Mit ihnen kann man drei Ur-Erfahrungen der Menschheit leicht nachvollziehen:

1. Lichtwege sind Geraden.
2. Licht ist instantan da.
3. Licht kann Licht folgenlos durchdringen.

1. gerade

Die wichtigste Hypothese über das Licht ist jene von der Geradlinigkeit der Lichtausbreitung. Sie ist die Grundlage der Vermessungstechnik (Abb. 1), und auch wenn ein Jäger über Kimme und Korn zielt, dann geht er davon aus, dass das Licht, das da zu ihm kommt, einen geraden Weg gegangen ist. Nicht zuletzt wird beim Deuten

von Himmelserscheinungen angenommen, dass der Lichtweg zwischen Erscheinung und Beobachter ein gerader ist.

2. instantan

Ob Licht eine endliche Zeitspanne benötigt, um von einem Ort zum anderen zu gelangen, kann man nicht ohne weiteres feststellen. Licht scheint Räume instantan zu erfüllen. Genauso wie das Licht, erfüllt auch die Dunkelheit den Raum scheinbar instantan: Unsere Beobachtungen lassen nicht erkennen, ob sich die Dunkelheit nach dem Abschalten einer Lichtquelle mit einer endlichen Geschwindigkeit im Raum „ausbreitet“ oder nicht.

Die Berechtigung, dennoch von einer LICHTAUSBREITUNG zu sprechen, die von Lichtquellen ausgeht, beruht auf der Feststellung einer Ursache-Folge-Beziehung: Ändert man den Schattenwurf einer Hand, indem man die Lage der Hand ändert, dann ist ursächlich der Willensimpuls da, die Lage der Hand zu ändern. Dem folgt dann die Änderung des Schattens. Das legte bereits in der Antike die Vermutung nahe, dass Licht sich nicht instantan ausbreiten kann. Galileo Galilei hat versucht, die Hypothese einer instantanen Lichtausbreitung experimentell zu falsifizieren. Er blendete eine Laterne auf, und ein weit von ihm entfernter Gehilfe sollte seine Laterne aufblenden, sobald er das Licht von Galileis Laterne erkannte, und so das Lichtsignal zu ihm zurückschicken. Galilei maß die Zeitspanne zwischen Absenden und Wiedereintreffen des Lichtsignals. Unter Berücksichtigung der Reaktionszeit des Gehilfen blieb keine Zeitspanne mehr übrig, die für eine Ausbreitung mit einer endlichen Geschwindigkeit hätte sprechen können. René Descartes entschied sich daher 1620 für die Hypothese, dass sich Licht unendlich schnell ausbreitet, also *instantan* von einem Ort zum anderen gelangt. Das ist auch die einfachste Hypothese, die mit dem experimentellen Resultat von Galilei im Einklang ist. Da sie bis 1676 nicht falsifiziert werden konnte, wollen wir sie hier vorerst akzeptieren. Dadurch wird unsere vorläufige Theorie des Lichts reine Geometrie: GEOMETRISCHE OPTIK. Sie ist eine statische Theorie, in welcher die Begriffe Zeit und Geschwindigkeit nicht auftreten.

3. durchdringbar

Die Erfahrung, dass ein Objekt durch ein anderes hindurchgehen kann, ohne dass man irgendeine Änderung beobachtet, kennt man aus anderen Bereichen der Alltagserfahrung nicht. Ein Wasserstrahl durchdringt beispielsweise nicht einfach folgenlos einen anderen Wasserstrahl. Lichtstrahlen können hingegen einander kreuzen, ohne dass eine gegenseitige Beeinflussung erkennbar ist.

Ausgehend von Naturerfahrungen formuliert die Physik Hypothesen und versucht die Realität daraus zu deuten. Um ein Verständnis für die physikalische Vorgangsweise zu entwickeln, soll dies anhand der genannten Ur-Erfahrungen über das Licht nachvollzogen werden. Man stellt die Naturerfahrungen also als Hypothesen an den Beginn der Überlegungen. Streng genommen ist aus heutiger Sicht keine davon richtig: Im Schwerfeld verläuft Licht „gekrümmt“, das Licht hat eine endliche Ausbreitungsgeschwindigkeit und es wechselwirkt auch ein wenig mit sich selbst. Im Rahmen der Geometrischen Optik, die unsere erste Annäherung an eine Theorie des Lichtes darstellt, wird dies ignoriert.



Abbildung 1: (a) Vermessung mit einem Theodoliten. Grundlage ist die Annahme, dass sich Licht geradlinig ausbreitet. (b) Entfernungsmessung mit dem Feldzirkel.

1.2. Mathematisches Schließen

Die Physik setzt Begriffe nicht als selbstverständlich voraus, sondern fordert eine Begriffsklärung. Für die in den obigen Axiomen der Geometrischen Optik auftretenden Begriffe der Geradlinigkeit bzw. Geraden wird das durch die Mathematik geleistet. Die im 4. Jh. v. u. Z. von Euklid entwickelte Mathematik, die Euklidische Geometrie, ist eine Theorie über einen abstrakten mathematischen Raum, den euklidischen Raum, und klärt u.a. die Begriffe Gerade, Winkel und Strecke. Die Gerade wird beispielsweise als kürzeste Verbindungslinie zwischen zwei Punkten definiert. Ferner zieht die Euklidische Geometrie logische Schlüsse aus nur wenigen im voraus gesetzten Grundannahmen, den Voraussetzungen bzw. Axiomen. Diesen Schlüssen kann man das Attribut „wahr“ zuordnen. Ihnen widersprechende Schlüsse bezeichnet man als „falsch“. „Wahr“ und „falsch“ sind Attribute, die nur an der Fakultät für Mathematik und an der Fakultät für Theologie sinnvoll sind, wenn auch aus verschiedenen Gründen.

Im Rahmen der Euklidischen Geometrie sind u.a. folgende Aussagen gültig:

- Die Winkelsumme im Dreieck ist 180° .
- Das Verhältnis von Kreisumfang zu Kreisdurchmesser ergibt die Zahl π .
- Es gilt der Lehrsatz des Pythagoras (6. Jh. v. u. Z.).

Am Beispiel der Geometrischen Optik stellt sich das Verhältnis von Mathematik, einer „reinen“ Geisteswissenschaft, und der „dreckigen“ Naturwissenschaft nun folgendermaßen dar: Die Geometrische Optik (und damit die Physik) geht von der Hypothese aus, dass für den realen Raum, d.h. für den Raum gemäß unserer Erfahrung bzw. für den Raum, den wir mit unseren Sinnesorganen und Instrumenten erkunden, die Axiome des euklidischen Raums erfüllt sind. Das ist keine mathematische Aussage mehr, sondern

eine PHYSIKALISCHE HYPOTHESE. Der Euklidische Raum wird zum Modell des physikalischen Raums. Damit setzt man zugleich voraus, dass alle über den Euklidischen Raum logisch ableitbaren Schlüsse auch für den physikalischen Raum wahr sind. Die Aussage, dass Lichtwege Geraden sind, hat nun vor dem Hintergrund der Begriffe der euklidischen Geometrie einen Sinn: Wenn zwei beliebige Punkte durch einen Lichtstrahl miteinander verbunden sind, dann ist das die kürzeste Verbindungslinie zwischen diesen beiden Punkten. Aufgabe der experimentellen Physik ist es, die dem Modell entspringenden Schlüsse einer Prüfung durch das Experiment zu unterziehen. Physikalische Hypothesen können nämlich prinzipiell durch Naturbeobachtungen falsifiziert werden, während mathematische Hypothesen (Axiome) in diesem Sinne nicht widerlegbar bzw. falsifizierbar sind. Sie müssen nur widerspruchsfrei sein. Ansonsten können sie als freie Produkte des phantasiebegabten menschlichen Geistes einfach a priori gesetzt werden, wie es dem Mathematiker beliebt.

Diese Identifikation des physikalischen Raums mit dem euklidischen Raum ermöglicht umgekehrt, die euklidische Geometrie im Schulunterricht als eine graphische bzw. anschauliche mathematische Theorie darzustellen: Geometrische Konstruktionen lassen sich allein mit Zirkel und Lineal durchführen. Keine der damit konstruierten „Kreise“ und „Geraden“ sind Kreise bzw. Geraden im streng mathematischen Sinne. Doch die Anschaulichkeit der visuellen Erfahrung erleichtert das Auffinden bzw. die Akzeptanz der mathematischen Sätze. Ob sie wahr sind, lässt sich jedoch niemals durch solche empirisch-anschaulichen Konstruktionen beantworten. Das Etikett, wahr zu sein, kann grundsätzlich nie einer Erkenntnis zukommen, die aus einer empirischen Erfahrung gewonnen wurde und somit a posteriori, d.h. im Nachhinein. Die Allgemeingültigkeit bzw. das apodiktische Wahrheitsattribut ergibt sich allein aus der mathematischen Analyse bzw. logischen Herleitung aus den fünf Axiomen Euklids. Wenn diese gesetzt sind, dann haben alle daraus logisch gefolgerten Sätze einen Wahrheitsgehalt a priori. Damit sind sie auch für den physikalischen Raum wahr. Das gilt so lange, wie man keine Zweifel an der Identifikation des physikalischen Raums mit dem euklidischen Raum hat.

Was nun, wenn eines Tages eine Naturbeobachtung gemacht wird, die irgendeiner der mathematischen und somit denotwendigen Aussagen über den physikalischen Raum widerspricht? Welche Konsequenzen hätte es, wenn Messungen eines Tages zeigen, dass es ein Dreieck gibt, dessen Winkelsumme zweifelsfrei und unbestreitbar größer als 180° ist? Da jeder der mathematischen Schlüsse aus den Axiomen der euklidischen Geometrie unbezweifelbar wahr ist, folgt daraus, dass die physikalische Ausgangshypothese falsch sein muss, d.h. dass der physikalische Raum nicht euklidisch sein kann. Es genügt bereits eine einzige empirische Beobachtung, die irgendeiner der mathematisch erschlossenen Aussagen über den physikalischen Raum widerspricht, um die Hypothese der Euklidizität des physikalischen Raumes zu Fall zu bringen. Angesichts der möglicherweise ungeheuer vielen vorherigen empirischen Beobachtungen, welche die Hypothese der Euklidizität bestätigten, mag dieses harte Verdikt erstaunen. Aber aus der Tatsache, dass sehr viele empirische Beobachtungen zuvor mit dieser Hypothese im Einklang waren, lässt sich nur schließen, dass im Rahmen der zuvor angewandten Beobachtungs- und Messmethoden die Hypothese der Euklidizität des Raumes eine ausreichende Näherung war.

Die Mathematik beschäftigt sich mit dem Problemkreis des widerspruchsfrei Denkbar und des durch formale Beweise logisch voneinander Ableitbar. Ihre philosophische Heimat ist der Rationalismus. Die Grundlage der Physik ist das sinnlich Erfahrbare und das daraus (durch Nachdenken) Erschließbare. Ihre philosophische Heimat ist der Empirismus. Die Grenzen der Erkenntnis der Mathematik werden durch die Grenzen des menschlichen Denkens gesetzt. Die Grenzen der physikalischen Erkenntnis sind einerseits die Grenzen der menschlichen Sinnlichkeit, weil die Physik aus dem Empirischen schöpft. Sie heißt nicht zuletzt Physik, weil sie sich auf Physisches bezieht bzw. eine real vorhandene Natur (Realismus). Andererseits sind die Grenzen durch die Mathematik gesetzt, weil physikalische Erkenntnisse über die Zusammenhänge verschiedener empirischer Befunde aus theoretischen Denkprozessen hervorgehen.

Euklidizität des Raumes. Wenn ein Raum mehr als eine Dimension hat, stellt sich die Frage, welche Verbindung zwischen den Dimensionen des Raumes besteht, also die Frage nach der Geometrie des Raumes. Nimmt man an, dass der physikalische Raum euklidisch ist, d.h. dass er die Eigenschaften erfüllt, die Euklid als Axiome seiner Geometrie eines mathematischen Raumes formuliert hatte, so impliziert das, dass er auch homogen und isotrop ist (s. Kap. 5). Bis Anfang des 20. Jahrhunderts schien die Euklidizitätshypothese unumstritten und hatte sich in allen Experimenten bewährt. In seinem Werk „Prinzipien der Reinen Vernunft“ setzte der Philosoph Immanuel Kant sogar voraus, dass der euklidische Raum der einzig denkbare und damit der einzig mögliche Raum sei und folglich a priori gegeben wäre. Carl Friedrich Gauß zeigte jedoch, dass sich auch nicht-euklidische Räume mathematisch widerspruchsfrei denken ließen. Seine Überlegungen gipfeln in seinem Theorema Egregium („herausragender Lehrsatz“) der Differentialgeometrie, welche einen Zusammenhang zwischen dem metrischen Tensor eines gekrümmten Raumes und dem Krümmungsradius herstellt. Daher war bereits gegen Ende des 19. Jahrhunderts unter Physikern Allgemeingut, dass die alleinige Denkmöglichkeit eines euklidischen Raumes, wie das von Philosophen der Kantschen Schule, beispielsweise von Arthur Schopenhauer, behauptet wurde, falsch war. Der Physiker und Philosoph Ludwig Boltzmann merkte daher kritisch an [8]:

„Überhaupt war Schopenhauer in dem, was er als apriorisch bezeichnete, nicht besonders glücklich. So bezeichnet er als apriorisch klar, dass der Raum drei Ausdehnungen hat. Heute wissen die Forscher, dass „a priori“ ein mehr als dreidimensionaler Raum denkbar, dass auch ein nicht Euklidischer Raum nicht undenkbar ist. Natürlich handelt es sich nicht darum, ob der erfahrungsgemäße Raum ein Euklidischer ist oder nicht, es handelt sich vielmehr darum, was a priori evident, was bloßer Erfahrungssatz ist.“

Gauß war auch der erste, der in Erwägung zog, dass der physikalische Raum gekrümmt sein könnte. Er unterzog die Euklidizitätshypothese einem experimentellen Test, indem er durch präzise Vermessung eines großräumigen Dreiecks nachzuweisen versuchte, dass man in der Natur Abweichungen von den Sätzen der Euklidischen Geometrie finden kann [?]. Im Rahmen seiner Messunsicherheit ergab sich jedoch nichts. Daher hatte Boltzmann auch keine Veranlassung zu bezweifeln, dass der erfahrungsgemäße Raum ein Euklidischer war. Erst die von Albert Einstein im Jahr 1915 entwickelte Allgemeine Relativitätstheorie

zeigte die Möglichkeit auf, dass Gravitation eine Raumkrümmung bewirken kann. Die theoretischen Schlüsse aus seiner Theorie haben sich empirisch bewährt. Für irdische Verhältnisse ist der Effekt der Raumkrümmung jedoch so winzig, dass er hier erst einmal außer Acht gelassen werden soll, d.h. in den ersten beiden Bänden dieses einführenden Buchs soll der physikalische Raum durch das einfachere Modell des euklidischen Raumes beschrieben werden.

Kartesische Koordinaten. Für Positionsbeschreibungen im euklidischen Raum werden im Folgenden meist kartesische Koordinaten verwendet. Als Bezugssystem für das kartesische Koordinatensystem wählt man drei orthogonal zueinander stehende Bezugsachsen, die man im Dreidimensionalen als x -, y - und z -Achse bezeichnet, und die sich in einem Bezugspunkt (Nullpunkt) schneiden. Ein Zahlentripel (x, y, z) gibt dann die Koordinaten des Punktes $P(x, y, z)$ bezüglich des Bezugssystem an. Im Zweidimensionalen genügt die Angabe zweier Koordinaten (z.B. die x - und y -Koordinate).

1.3. Quantifizierung

Ein weiterer Charakterzug der Naturwissenschaft ist, dass sie bestrebt ist, die Natur nicht nur qualitativ zu beschreiben, sondern auch quantitativ. Beispielsweise kann man mit einem Zirkel Längen abtragen und miteinander vergleichen. Im Gelände kann eine zurückgelegte Entfernung durch Abzählen der Schritte bestimmt werden. Die Beine entsprechen den Zirkelschenkeln. Der entscheidende Vorgang ist hierbei das Vergleichen: Ein Merkmal eines Objektes ist das *soundso* Vielfache des Merkmals eines anderen Objektes. Das setzt selbstverständlich erst einmal eine Vergleichbarkeit der Merkmale voraus. Der Vergleich muss sich auf die gleiche PHYSIKALISCHE GRÖSSE beziehen.

Physikalische Messwerte werden durch Vergleich mit einer Einheit ermittelt.

Längen

Die Strecke zwischen zwei Ortspunkten wird durch Längenmessung bestimmt. Die größenordnungsmäßige Richtvorstellung für unsere heutige Maßeinheit der Distanz, das METER, ist die Schrittweite des Menschen. Die Armeen der antiken Welt hatten Schrittzähl-Spezialisten. Es ist erstaunlich, mit welcher Genauigkeit beispielsweise die Schrittzähler von Alexander dem Großen die Etappenstrecken zwischen Makedonien und Indien gemessen haben. Nun ist die Schrittweite bei Menschen unterschiedlich und eine Distanzangabe wird sicher besser, wenn man sie objektiviert und auf eine Einheitslänge bezieht, beispielsweise indem man Entfernungen mit einem Feldzirkel misst (Abb. 1b), der mit einem wohldefinierten Maß für die Länge geeicht worden ist.

Historisch wurden sehr unterschiedliche und oft nur regional gebräuchliche Längenmaße von den jeweiligen Eichbehörden festgelegt. Heute verwendet man das METER als INTERNATIONALE LÄNGENEINHEIT. Wie das Meter als Eichnormal definiert ist, wird in Kap. 8.4 dargestellt. Unabhängig davon nach welchem Verfahren die Streckenmessung

Tabelle 1: Präfixe für die Bezeichnung der Zehnerpotenzen

Präfix	Potenz	Präfix	Potenz	Präfix	Potenz	Präfix	Potenz
-	1	-	1	mikro	10^{-6}	mega	10^6
dezi	10^{-1}	deka	10^1	nano	10^{-9}	giga	10^9
centi	10^{-2}	hekto	10^2	pico	10^{-12}	tera	10^{12}
milli	10^{-3}	kilo	10^3	femto	10^{-15}	peta	10^{15}

tatsächlich geschieht, wollen wir im Folgenden alle Messinstrumente für gerade Strecken als LINEALE bezeichnen. Lineale kann man sich gerade Maßstäbe vorstellen, die mit einer Längenskala versehen sind. Mit dem Zirkel kann man Längen abtragen und durch Vergleich mit einem Lineal bestimmen.

Wenn man also eine Länge mit einem Zirkel abnimmt und feststellt, dass man den Zirkel zehn Mal abtragen muss, bis man die Längeneinheit von einem Meter erreicht, dann entspricht jeder dieser zehn Abschnitte einem Dezimeter. „Dezi-“ ist eines der gebräuchlichen Präfixe, die man einer Einheit voranstellt, wenn man Zehnerpotenzen einer Einheit angeben will (s. Tab. 1). Die Entwicklung des Maßes der Länge erfolgte über mehrere Zwischenschritte, die historisch für andere physikalische Größen in ähnlicher Weise erfolgte:

1. Der erste Schritt ist der Übergang von einem individuellen oder subjektiven Maß zu einem objektivierten Maß, etwa durch eine objektivierte Messvorschrift.
2. Der zweite Verbesserungsschritt besteht in der Erstellung von Übereinkünften, also von Konventionen. Diese wurden zunächst auf der Ebene von Regionen umgesetzt und mündeten schließlich in der Entwicklung einer internationalen Einheitenkonventionen, dem *Système international d'unités* (kurz: SI-SYSTEM).

Der zweite Schritt wurde in Liberia, Myanmar und USA noch nicht vollzogen. Daher ist es manchmal nötig, Einheiten zu konvertieren. Wie genau muss man dabei sein? Das hängt sehr von der Aufgabenstellung ab. 1999 verglühte beispielsweise die NASA-Sonde „Mars Climate Orbiter“, weil die Navigationssoftware von Lockheed Martin (das ist die Firma, die Südkorea mit ihren Abwehrraketen vor nordkoreanischen Angriffsraketen schützen soll) in der US-Krafteinheit (lbf) ausgelegt war. Die NASA verwendete SI-Einheiten und die richtige Einheitenkonversion wäre gewesen: $1 \text{ lbf} = 4.4482216152605 \text{ N}$. Kosten der fehlerhaften Einheitenkonversion: rund 200 Millionen US-Dollar.

Winkel und Raumwinkel

Traditionell misst man Winkel, indem man dem vollen Umfang eines Kreises einen Winkel von 360° zuordnet. Die Zahl 360 wurde deshalb ausgewählt, weil man sie durch 1,2,3,4, usw. ganzzahlig teilen kann, also bequem ganzzahlige Winkel für eine Reihe von Bruchteilen des vollen Kreiswinkels angeben kann. Ein Winkel von 90° wird also durch einen Kreisbogen von einem Viertel des vollen Kreises repräsentiert. Im Rahmen der euklidischen Geometrie ist die Länge des Kreisbogens dem Radius des Kreises proportional. Daher kann man auch das Verhältnis von Kreisbogenlänge s zu Radius r als Winkelmaß

heranziehen. Das ist die Winkelangabe

$$\alpha = s/r$$

im BOGENMASS. Ist der Raum euklidisch, so entspricht dem vollen Kreisumfang ein Winkel von 2π bzw. 2π rad im Bogenmaß.

Winkel haben ein Maß. Auch wenn dieses Maß eine dimensionslose Größe ist, also eine reine Zahl, muss Klarheit darüber herrschen, ob ein Winkel im Bogenmaß oder im Gradmaß angegeben ist. Daher verwendet man im ersten Fall die Marke (engl. „tag“) RAD und im zweiten Fall das Gradsymbol $^\circ$ bzw. die Marke DEG für die Angabe des Winkelmaßes. Während das Meter eine Einheit ist, sind die „tags“ nur Hilfsmaßeinheiten. Sie sind jedoch ebenfalls im SI-System definiert. Traditionell wird der Winkel von $1^\circ = 1 \text{ deg}$ in 60 Winkelminuten unterteilt ($1^\circ = 60'$). Eine Winkelminute wird mit $1'$ bezeichnet und wird in 60 Winkelsekunden unterteilt ($1' = 60''$). Diese Unterteilung ist manchmal rechnerisch unbequem. Daher kann man Winkel auch dezimal angeben, z.B. $30' = 0.5^\circ$.

Der RAUMWINKEL Ω ist analog zum Winkel definiert. Der durch eine Fläche \mathcal{A} auf der Sphäre einer Kugel mit dem Radius r aufgespannte Raumwinkel ist

$$\Omega = \mathcal{A}/r^2.$$

Er ist ebenfalls dimensionslos und wird mit dem „tag“ STERAD (sr) angegeben. Ein Raumwinkel von 4π sr umfasst die volle Kugeloberfläche.

Die Einheiten der visuelle Photometrie

Zu den Erscheinungen des Lichts, die das Auge wahrnehmen kann, gehören seine Stärke und seine Farbe. Bis Mitte des 19. Jahrhundert waren die visuellen Eindrücke die einzige Möglichkeit, diese Erscheinungen zu charakterisieren, denn für radiometrische Messungen fehlten insbesondere grundlegende Begriffe wie *Energie* und *Leistung*. Dieser Abschnitt soll das messtechnische Grundprinzip einer visuellen VERGLEICHSMETHODE bzw. Abgleichmethode exemplarisch erläutern, mit der man die Lichtstärke durch Vergleich mit einer Standardkerze quantifizieren kann kann.

Der physiologische Eindruck von Licht auf Menschen ist individuell verschieden und daher nicht objektivierbar. Objektivierbar ist hingegen, ob zwei Lichteindrücke gleich sind oder nicht. Die menschliche Sehempfindung ist sowohl hinsichtlich der Diskriminierung von Farbeindrücken als auch von Helligkeitseindrücken erstaunlich leistungsfähig. Das ist der Ansatzpunkt der Messmethoden der visuellen Photometrie. Die nachfolgend vorgestellten Messgrößen sind mit einem Index „v“ für „visuell“ gekennzeichnet. Das soll darauf hinweisen, dass sie durch visuelle Beobachtung nach dem Vergleichsprinzip gemessen wurden.

Wenn zwei gleiche Kerzen entzündet werden, dann beleuchten sie unter gleichen Umständen (gleiche Abstände usw.) ihre jeweils zugeordneten Flächen so, dass sowohl der Farbeindruck als auch die Lichtstärke als gleich empfunden werden. Vorausgesetzt dass Beobachtungsumstände und Farbeindruck gleich sind, kann man die Stärke einer unbekanntes Lichtquelle danach bemessen, ob man zwei, drei oder mehr Kerzen benötigt,

um Gleichheit der Lichtstärke zu empfinden. Man kann daher eine genau spezifizierte Standardkerze und genau spezifizierte Beobachtungsumstände für einen Vergleich heranziehen, um ein Maß für die Lichtstärke zu definieren. Das ist die physikalische Grundlage der photometrischen SI-Einheit CANDELA für die LICHTSTÄRKE I_v :

$$\begin{aligned} [I_v] &= 1 \text{ Candela} = 1 \text{ cd} . \\ [I_v] &= [\text{cd}] \end{aligned}$$

Zur Notation: Die Größe, deren Einheit oder physikalische Dimension angegeben werden soll, steht hier zwischen eckigen Klammern. Wenn nicht anders angegeben, folgt dahinter die SI-Einheit (Beispiel: erste Zeile) oder in eckigen Klammern einfach nur die physikalische Dimension (Beispiel: zweite Zeile). Eine typische Haushaltskerze leuchtet mit einer Lichtstärke von $I_v \approx 1 \text{ cd}$.

Die Einheit des von einer punktförmigen Lichtquelle der Lichtstärke I_v in einen Raumwinkel Ω gehenden LICHTFLUSSES

$$\Phi_v = I_v \Omega$$

heißt LUMEN:

$$[\Phi_v] = 1 \text{ Lumen} = 1 \text{ lm} = 1 \text{ cd sr} = 1 \text{ Candela Sterad} .$$

Der auf eine Fläche \mathcal{A} orthogonal auftreffende Lichtfluss Φ_v ruft dort eine BELEUCHTUNGSSTÄRKE

$$E_v = \Phi_v / \mathcal{A} ,$$

hervor, d.h. die Beleuchtungsstärke ist gleich dem pro Flächeneinheit auftreffenden Lichtfluss. Eine punktförmige Modell-Lichtquelle hat per Definitionem keine Richtungspräferenz und strahlt in alle Richtungen gleich viel Licht ab. Die Beleuchtungsstärke, mit der die Oberfläche einer Kugel bestrahlt wird, nimmt aus diesem Grunde quadratisch mit dem Abstand bzw. dem Radius r ab. Empirisch ergibt sich in der Tat, dass man vier Kerzen benötigt, um mit ihnen den gleichen Helligkeitseindruck in einem Abstand von 2 m zu erzielen, wie mit einer Kerze im Abstand von 1 m.

In Abb. 2 ist die Schattenmethode zur Messung der Lichtstärke dargestellt. Ein Stab wirft zwei Halbschatten auf einen Projektionsschirm. Der Lichteindruck dort stammt von jeweils nur einer der beiden Lichtquellen A oder B, während außerhalb des Halbschattens der von der Überlagerung beider Quellen hervorgerufene Lichteindruck A+B vorliegt. Die Lichtquellen haben nur dann gleiche Lichtstärke, wenn sowohl Beleuchtungsstärke als auch Farbeindruck der beiden Halbschattengebiete übereinstimmen. Für farbgleiche Lichtquellen kann man die Lichtstärke dadurch messen, dass man das Abstandsgesetz $E_v \propto r^{-2}$ zum Abgleich ausnutzen.

Ausgehend vom Modell des Lichtstrahls kann das von Lichtquellen ausgehende Licht modellmäßig als ein Lichtbündel gedacht werden, d.h. als ein Bündel von Strahlen. Die Strahlen divergieren, d.h. sie laufen in unterschiedliche Richtungen auseinander. Der Lichtfluss kann graphisch veranschaulicht werden, indem man in die Einhüllende eines Lichtbündels eine dem Zahlenwert des Lichtflusses entsprechende Anzahl von Strahlen einzeichnet. Sie

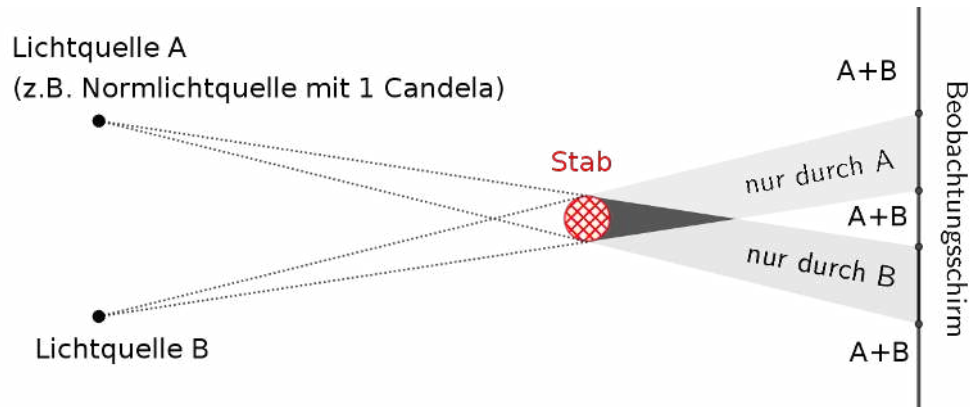


Abbildung 2: Visuelle Photometrie nach der Schattenmethode. Ein Stab entwirft zwei Halbschatten. Ob Helligkeit bzw. Farbe der Lichtquellen A bzw. B gleich sind, lässt sich visuell sehr gut entschieden.

repräsentieren die Flusslinien des Lichtes. Die Beleuchtungsstärke entspricht der Anzahl der orthogonal auf eine Fläche auftreffenden Flusslinien. Die Dichte der auf eine Fläche auftreffenden Strahlen liefert somit ein anschauliches Bild der Beleuchtungsstärke (s. z.B. Fig. 23). Wenn man aus einer solchen Skizze erkennt, dass eine optische Komponenten den Lichtfluss auf eine kleinere Fläche konzentriert, dann bedeutet das, dass dort die Beleuchtungsstärke steigt. Stellt man eine Blende in einen Strahlengang, so enden dort die Lichtstrahlen. Hinter der Blende ist der Lichtfluss und damit auch die Beleuchtungsstärke null. Wenn man zwei parallele Strahlenbündel zusammenführt bzw. sich unter einem kleinen Winkel schneiden lässt, so besteht der Fluss aus den Flusslinien beider Strahlenbündel, d.h. die Beleuchtungsstärken addieren sich. Da die Zahl der von einer Punktquelle ausgehenden Strahlen konstant bleibt, aber die beleuchtete Fläche quadratisch mit dem Quadrat des Abstands r zunimmt, folgt aus dem Modell anschaulich das Abstandsgesetz $E_v \propto r^{-2}$. Bei einer Abbildung können maximal alle von einem Objektpunkt ausgehenden Strahlen sich wieder in einem Bildpunkt vereinen. Daher kann die Lichtstärke des Bildes einer Punktquelle die Lichtstärke der Quelle nicht übersteigen, d.h. auch wenn es gelingt, alle von einer Quelle ausgehenden Lichtstrahlen durch optische Elemente in einem Fokuspunkt zusammenzuführen, ist die Lichtstärke maximal diejenige der Quelle.

1.4. Näherung

Aus der Sicht der Theorie liefern Rechnungen exakte Ergebnisse, exakte Zahlen. Aber möchte man immer das exakte Ergebnis ausrechnen? In der Physik rechnet man oft „unge nau“. Das ist innerhalb gewisser Regeln zulässig, weil empirische Ergebnisse grundsätzlich Schätzungen mit einer gewissen Unsicherheit sind (s. Kap. 1.5.4). Es gibt verschiedene Gründe, warum man zu Näherungen greift, beispielsweise wenn man ein Problem mathematisch nicht genau berechnen kann oder sich die Zeit für eine aufwendige Rechnung ersparen möchte.

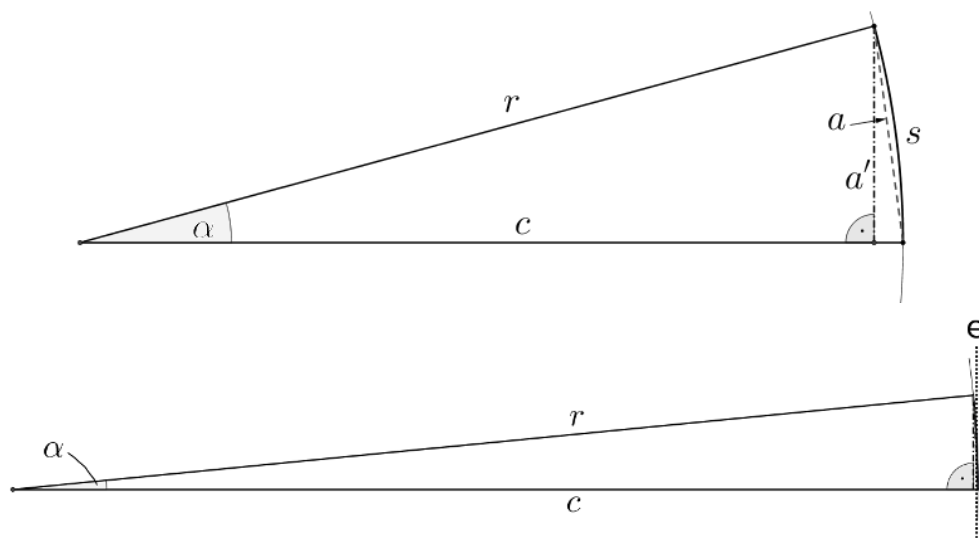


Abbildung 3: Für sehr kleine Winkel α werden Bogenlänge s , Sekante a und Gegenkathete a' näherungsweise gleich groß.

Die Zahl π

Wenn man unendlich viel Zeit hat, dann kann man das Ergebnis einer Rechnung, in der die Zahl π auftritt, im Prinzip exakt angeben, also mit unendlich vielen Dezimalstellen. Je nachdem, wie es für den Vergleich mit der durch eine stets mit einer Unsicherheit behafteten Messung nötig ist (s. Kap 1.5.4), wird man für die Zahl π jedoch eine Näherung einsetzen. Liegt die empirische Schätzgenauigkeit bei 1%, dann genügt es beispielsweise anstatt der Zahl π (mit unendlich vielen Stellen) die auf drei Stellen verkürzte Näherung $\pi \approx 3.14$ zu verwenden.

Näherung für kleine Winkel

Abb. 3 zeigt das Segment eines Kreises mit dem Radius r , der Bogenlänge s und dem Winkel $\alpha = s/r$. Das durch die Sekante a gebildete gleichschenklige Dreieck und das das rechtwinklige Dreieck mit der Gegenkathete a' , der durch r gebildeten Hypotenuse und der Ankathete c haben den gleichen spitzen Winkel $\alpha \neq 0$. Es gilt $s > a > a'$ sowie $r > c$. Wenn α jedoch immer kleiner wird, dann gleichen sich diese Seiten immer mehr an, so dass man für genügend kleine Winkel näherungsweise $s \approx a \approx a'$ und $r \approx c$ setzen kann. Ob man dann den Winkel nun korrekt durch $\alpha = s/r$ oder beispielsweise durch $\alpha \approx a/r$ oder $\alpha \approx a'/c$ ausdrückt, ist im Rahmen der Näherung für kleine Winkel gleich. Im Grenzfall $\alpha \rightarrow 0$ verschwindet die Krümmung von s und a' , a und s kommen näherungsweise auf der geraden Linie e zu liegen, die auf der Seite c orthogonal (rechtwinklig) steht.

Näherungsformeln für einige Funktionen

In Tab. 2 sind die Näherungsformeln einiger Funktionen durch Ausdrücke in Potenzen von x angegeben. Sie gelten für $x \ll 1$. Das Symbol „ \ll “ bedeutet „sehr klein gegen“, d.h.

Tabelle 2: Wichtige Näherungsformeln

$\frac{1}{1+x} \approx 1 - x$	$\ln(1+x) \approx x$	$\sin x \approx \tan x \approx x$	$\sinh x \approx x$
$\sqrt{1+x} \approx 1 + \frac{1}{2}x$	$e^x = \exp(x) \approx 1 + x$	$\cos x \approx 1 - \frac{1}{2}x^2 \approx 1$	$\cosh x \approx 1$

$x \ll 1$ bedeutet hier, dass die Größe x *sehr klein gegen 1* ausfallen soll. Das Argument der Winkelfunktionen ist selbstverständlich im Bogenmaß einzusetzen. Nähert man in einem Ausdruck mehrere darin auftretende Funktionen, so ist es wichtig, dass man stets bis zu Gliedern gleicher Potenz in x nähert und keines davon vergisst. Wenn man die Näherung bei der niedrigsten Potenz abbricht, die nicht verschwindet, spricht man von NÄHERUNG ERSTER ORDNUNG.

1.5. Messung

1.5.1. Präparation

Einer Messung geht üblicherweise eine PRÄPARATION voraus, d.h. die Vorbereitung einer geeigneten Mess-Situation. Das soll hier an einem Beispiel erläutert werden.

Um die Hypothese der Geradlinigkeit der Lichtausbreitung als Grundlage für die Geodäsie und Vermessungstechnik nutzen zu können, muss erst eine für die Messung geeignete Ausgangssituation hergestellt werden. Das Ziel der Präparation besteht im vorliegenden Fall darin, ein möglichst paralleles und in seiner Ausdehnung möglichst eng begrenztes Lichtbündel zu erzeugen.

Etwa 100 u. Z. beschrieb Heron von Alexandria das in Abb. 4a gezeigte Vermessungsgerät mit Winkelteilungen auf einer Scheibe. Zur Präparation gehört zum einen die möglichst horizontale Ausrichtung der Messscheibe für den Horizontalwinkel (Horizontalkreis). Ferner gehört dazu die Präparation eines Lichtbündels, das der Idealvorstellung eines Lichtstrahls möglichst nahe kommt. Diese Präparation geschieht durch eine Visiereinrichtung. Sie besteht aus zwei Blenden mit einem Fadenkreuz. Die erste Blende begrenzt die Querschnittsfläche des Strahlenbündel in der Blendenebene. Die zweite Blende begrenzt den Winkel, um den die Strahlen des Bündels von der Zentralrichtung abweichen können, indem stark divergente Lichtstrahlen ausgeschaltet werden. Dadurch wird die Parallelausrichtung des Bündels verbessert. Das Parallelausrichten eines Bündels nennt man KOLLIMATION. Das Lichtbündel ist umso besser kollimiert, je enger die Blenden sind und desto weiter sie voneinander entfernt stehen. Als Maß für die Güte einer Kollimation kann man den Öffnungswinkel bzw. die Strahldivergenz heranziehen. Die STRAHLDIVERGENZ α ist die Hälfte des Öffnungswinkels β . Da man bei der Kollimation ohnehin kleine Öffnungswinkel anstrebt, kann man diesen sehr gut durch die Näherung kleiner Winkel abschätzen. Für das in Abb. 4b gezeigte Kollimatorrohr kann die Strahldivergenz in dieser Näherung beispielsweise durch $\alpha \approx w/d$ angegeben werden. Wenn beide Blenden

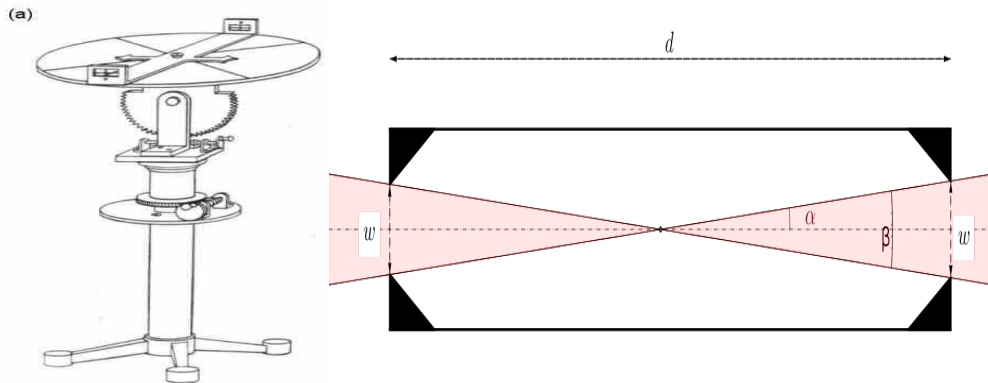


Abbildung 4: (a) Antikes Vermessungsgerät.

(b) Querschnitt durch ein Kollimatorrohr mit zwei Blenden der Breite w im Abstand d . Die Strahldivergenz α ist der größte Neigungswinkel, den ein Strahl des einfallenden Lichtbündels gegen die strichpunktiert gezeichnete Achse des Kollimators haben kann.

eines Visiers beispielsweise $w = 1 \text{ mm}$ breit sind und ihr Abstand $d = 1 \text{ m}$ beträgt, dann beträgt die Strahldivergenz ungefähr $\alpha \approx w/d = 0.001 \text{ rad}$.

Eine Kollimation mit zwei Blenden hat den Nachteil, dass man eine Verbesserung der Kollimation mit einer Abnahme des Lichtflusses bezahlt. In der Praxis stellt die Präparation daher einen Kompromiss zwischen Lichtfluss und der Annäherung an den idealen Lichtstrahl dar.

1.5.2. Vermessung der Welt in der Antike

Die beste kollimierte natürliche Lichtquelle großer Lichtstärke, die den Wissenschaftlern der griechischen Antike zur Verfügung stand, war die Sonne. Das Sonnenlicht fällt mit einem Divergenzwinkel von ungefähr $1 \times 10^{-2} \text{ rad}$ auf die Erde ein, was ungefähr einem halben Grad entspricht. Das ist fast schon parallel zu nennen. Deshalb ist der Schattenwurf im Sonnenlicht beispielsweise auf kleinere Entfernung recht scharf.

Diese ausgezeichnete Kollimation der „Sonnenstrahlen“ wurde ca. im 3. Jh. v. u. Z. von Eratosthenes zur ersten Bestimmung des Erdumfangs und des Abstandes von Erde und Mond bzw. Erde und Sonne herangezogen: Ihm war bekannt, dass ein tiefer Brunnen in Assuan das Sonnenlicht zur Mittagszeit der Sommersonnenwende reflektierte. Da Assuan nahezu auf dem Wendekreis liegt, fällt das Sonnenlicht dann senkrecht zur Erdoberfläche auf. Er kannte den Winkel des Schattenwurfs zur Mittagszeit in Alexandria und die Entfernung zwischen beiden Orten, woraus er den Erdumfang berechnen konnte. Die Methode ist in Abb. 5a skizziert. Selbstverständlich war seine Annahme paralleler Sonnenstrahlen streng genommen nicht richtig und seine Entfernungsbestimmung sehr ungenau. Aber dennoch war diese Messung eine Sternstunde der Menschheitsgeschichte. In der Physik kommt es oft nicht darauf an, dass man exakt misst, oder dass –wie im

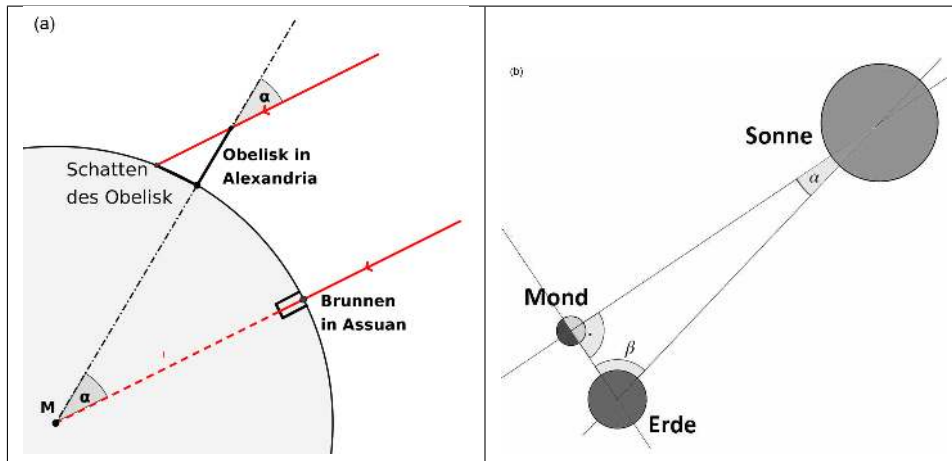


Abbildung 5: (a) Messung des Erdumfangs durch Eratosthenes. (b) Messung des Abstands Erde-Sonne durch Aristarch.

vorliegenden Fall– die Sonnenstrahlen exakt parallel wären, sondern nur darauf, dass die Messungen genau genug sind, um im Rahmen eines idealisierten Modells zu einer neuen Erkenntnis zu gelangen. In diesem Fall konnte Eratosthenes den Erdumfang größenordnungsmäßig korrekt bestimmen.

Ausgehend vom nunmehr bekannten Erdradius wurde die Entfernung von der Erde zum Mond aus den Beobachtungen bei Mond- und Sonnenfinsternis erschlossen. Vor mehr als 2000 Jahren ermöglichte dies Aristarch, schließlich die Entfernung von der Erde zur Sonne zu bestimmen (Abb. 5b): Wenn man vor Sonnenuntergang einen Halbmond beobachtet, so fällt das Licht der Sonne genau orthogonal zur Sichtlinie Erde-Mond ein. Aus dem rechtwinkligen Dreieck erhält man die Entfernung zur Sonne, indem man den Winkel β zwischen der Sichtlinie auf den Mond und der Sonne bestimmt.

1.5.3. Vermessung der Welt in der frühen Neuzeit

Auch in der frühen Neuzeit bleibt die TRIANGULATION, d.h. die Ausmessung von Dreiecken, die grundlegende Methode der Landvermessung. Abb. 6a illustriert die Methode der Entfernungsmessung durch die Bestimmung des Winkels, unter dem eine gegebene Strecke auf der Messlatte gesehen wird. Der Winkel wird um so kleiner, je weiter die Messlatte entfernt ist.

Als Winkelmessgerät wird meist ein Theodolit verwendet, wie er z.B. in Abb. 1b gezeigt ist. Er hat zwei Drehkreise, einen Horizontalkreis und einen Vertikalkreis, deren Drehachsen orthogonal zueinander stehen und sich in einem Bezugspunkt schneiden. Zwei Richtungen heißen NORMAL bzw. ORTHOGONAL bzw. rechtwinklig zueinander, wenn sie einen rechten Winkel miteinander einschließen. SENKRECHT bzw. vertikal wird eine Richtung genannt, wenn sie parallel zur lokalen Richtung des freien Falls ist. Sie wird z.B. durch das Senkblei bzw. Senklot ermittelt. Eine WAAGERECHE Richtung steht ortho-

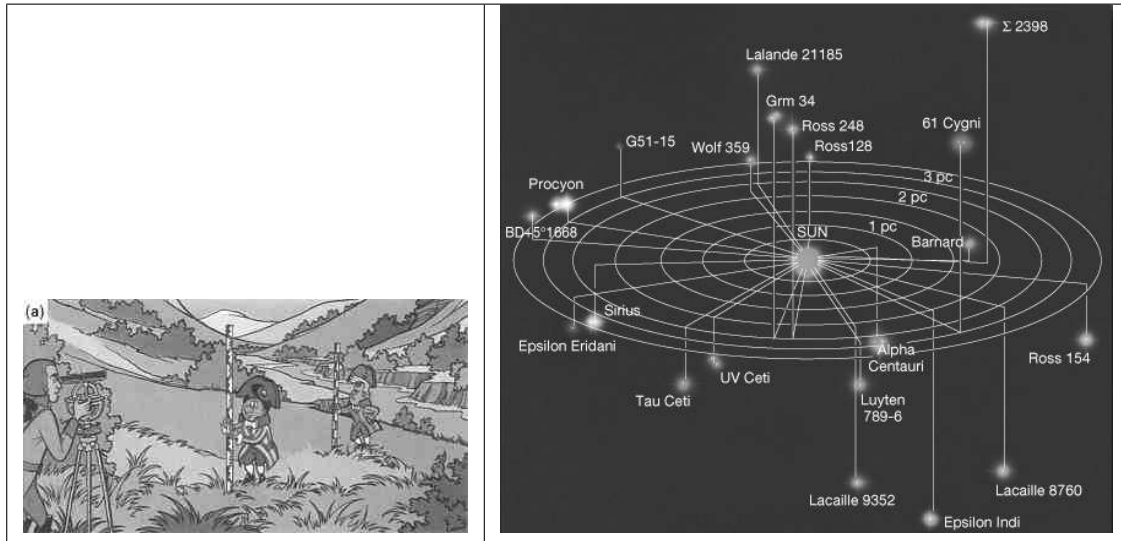


Abbildung 6: (a) Digidags bei der Landvermessung. Der Comic ist Kult bei den Geophysikern. (b) Sterne im Umkreis von einigen Parsec um unser Sonnensystem.

gonal zur senkrechten Richtung und wird z.B. mit einer Wasserwaage bestimmt. Zur Präparation eines Theodoliten gehört, dass er zuerst mit dem Lot so ausgerichtet wird, dass die Drehachse des Horizontalkreises senkrecht steht. Die entgegengesetzt zu der zum Erdmittelpunkt zeigende Richtung der Drehachse weist zum Zenit und ist Bezugsrichtung für den Polarwinkel ϑ , der an der Winkelteilung des Vertikalkreises abgelesen wird. Winkel ϕ in der waagerechten Ebene bzw. der Horizontalebene werden am Horizontalkreis abgelesen und als AZIMUTWINKEL bezeichnet.¹

Die erste genauere kosmische Entfernungsbestimmung wurde im 17. Jh. durch Cassini durchgeführt. Er bestimmte den Winkel, unter dem der Mars von Paris und Cayenne (Guayana) aus beobachtet wird, also mit einer sehr großen Basislinie, die durch terrestrische Triangulation genau genug bekannt war. Daraus lässt sich mit Hilfe der Keplerschen Planetengesetze die mittlere Entfernung zwischen Erde und Sonne berechnen. Dieser mittlere Radius der Erdbahn stellt die für die Astronomie wichtige Längeneinheit ASTRONOMISCHE EINHEIT dar:

$$1 \text{ Astronomische Einheit} = 1\text{AE} \simeq 1.5 \times 10^{11}\text{m.}$$

Die 1977 gestartete Sonde Voyager 1, das derzeit am weitesten von der Erde aus ins Weltall vorgedrungene Objekt, zu dem noch Funkkontakt besteht, hat inzwischen mehr als 140 AE zurückgelegt.

Der Radius der Erdbahn ist dann die Basislinie für die Entfernungsbestimmung der nächsten Sterne. Diese Methode der kosmischen Entfernungsbestimmung durch Triangulation

¹Ohne „th“, also nicht „Azimuth“, denn das Wort kommt aus dem Arabischen „as-samt“, اَلْسَمْت, und nicht aus dem Griechischen.

ist durch die Winkelauflösung der Teleskope begrenzt. Bei einer Auflösung von einer Winkelsekunde kann man damit Entfernungen von rund 200 000 AE bestimmen. Diese Entfernung heißt PARSEC (Parallaxensekunde):

$$1 \text{ Parsec} = 1 \text{ pc} \simeq 3 \times 10^{16} \text{ m.}$$

Der zu uns nächstliegende Stern (Proxima Centauri) ist beispielsweise etwas mehr als 1 pc von uns entfernt. Abb. 6b zeigt die uns am nächsten benachbarten Sonnensysteme. Ihre Entfernungen konnten allesamt mittels Triangulation ermittelt werden. Spätestens auf der Skala dieser Entfernungen wird offensichtlich, dass das, was man als Messung bezeichnet, nichts als eine gute Schätzung darstellt: Alle Messungen sind Schätzungen!

1.5.4. Messunsicherheit

Physikern stehen drei Möglichkeiten zur Verfügung, um zu einer quantitativen Beschreibung der Natur zu kommen:

1. Raten
2. Schätzen
3. Fachmännisches und vorschriftsmäßiges Schätzen

Anders als das reine Raten gehen Schätzmethode von empirischen Beobachtungen aus. Das fachmännische und vorschriftsmäßige Schätzen nennt man MESSEN. Es liegt in der Natur des Schätzens und somit des Messens, dass dem Ergebnis immer eine gewisse Unsicherheit anhaftet. Daher besteht eine MESSUNG stets aus einer Schätzung der interessierenden Größe (Messwert) sowie einer Schätzung der Schätzunsicherheit (Messunsicherheit). Je weiter man es aber in der Kunst des genauen Schätzens, also des Messens, gebracht hat, desto mehr kann man die Unsicherheit der Daten verringern.

Systematische Messfehler. Für Messungen verwenden Physiker in der Regel Messwert-Schätzinstrumente, die man meist kurz als MESSINSTRUMENTE bezeichnet. Es kann in der Praxis vorkommen, dass man eines auswählt, welches nicht zweckmäßig, nicht gut genug gebaut oder schlecht geeicht ist. Wenn man mehrere Alternativen für eine Messung hat (diese Wahl hat man aber nicht immer) und sich eine fehlerhafte daraus ausgesucht hat, dann hat man einen Fehler begangen, d.h. die Schätzung wird nicht fachmännisch ausfallen. Die aus solchen Fehlern resultierenden Abweichungen nennt man SYSTEMATISCHE MESSFEHLER. Wenn man erst einmal drauf gekommen ist, dass sie auftreten, und man auch noch die Ursache herausfindet, dann kann man sie manchmal vermeiden oder zumindest reduzieren, indem man beispielsweise den Bau des Instruments oder die Messmethode bzw. -methodik verbessert. Beispielsweise kann das Teleskop eines Theodoliten eine systembedingte Missweisung nach rechts oder nach links von der zu visierenden Richtung aufweisen. Diese systematische Messabweichung des Instruments kann man dadurch kompensieren, dass man das Fernrohr um 180° um seine Horizontalachse dreht und anschließend um 180° um die Vertikalachse. Eine linke Missweisung geht dann in eine gleich große rechte Missweisung über. Wenn man daher aus beiden Messwerten den Mittelwert bildet, mittelt sich die Missweisung heraus.

Zufallsbedingte Messunsicherheiten. Auch wenn man alle systematischen Messfehler beseitigt hat und die Messungen mit der gleichen Güte der Präparation durchführt, werden sich zwei Messungen i.A. unterscheiden. Wenn ein Geometer beispielsweise Messungen im Gelände durchführt, dann hängt das Resultat davon ab, wieviel Erfahrung der Geometer hat und wie viel Mühe, Sorgfalt und Zeit er für eine Messung aufwendet. Das Resultat könnte davon abhängen, ob er die Messung kurz nach einem Streit mit seinem Vorgesetzten durchgeführt hat oder nach einer durchzechten Nacht, ob er einen Kaffee getrunken hat oder ob er am Ende des Arbeitstages im Gelände müde geworden ist. Solche Einflussfaktoren sind nicht systematisierbar und folglich sind die Messwerte und Messabweichungen es auch nicht. Denn lägen systematisierbare Einflussfaktoren vor, beispielsweise weil man einen Trend dabei feststellen oder es irgendwie schaffen würde, den Einfluss systematisch zu beschreiben, dann könnte man ihn auch prinzipiell korrigieren, d.h. es würde auf die Korrektur eines systematischen Fehlers hinauslaufen.

Man muss sich leider mit der Tatsache abfinden, dass Messresultate i.A. von einer Vielzahl komplexer Einflussfaktoren abhängen können, die man einfach nicht alle unter Kontrolle hat. Im Hinblick auf diese kann man nur die Achseln zucken und aufgeben, weiter nach systematischen Ursachen zu suchen. Die meist in einem gewissen begrenzten Bereich schwankenden Messresultate, die trotz aller Bemühungen auftreten, die Messungen unter kontrollierten gleichen Bedingungen ablaufen zu lassen, nennt man zufällig. Damit soll ausgedrückt sein, dass man keine weiteren Anstrengungen unternehmen will, sie weiter zu systematisieren.

Schwankungen der Messergebnisse, die man nicht mehr weiter systematisieren kann oder will, heißen ZUFÄLLIG.

Was ist Zufall? Dass man Schwankungen als zufällig deklariert, liegt oft daran, dass man vor dem Einfluss einer großen Vielzahl komplexer Einflüsse einfach kapituliert. Manchmal wäre es vielleicht möglich, gewisse Einflussfaktoren durch vermehrte Anstrengung des Experimentators besser zu kontrollieren, aber man entscheidet sich bewusst dafür, darauf zu verzichten, weil sich der dazu nötige Aufwand hinsichtlich der zu klärenden Fragestellung nicht rechtfertigen ließe oder unsinnig wäre. Aber selbst wenn man bereit wäre, alle Anstrengungen auf sich zu nehmen, so lässt sich das Ziel der vollständigen Kontrolle einer Messung u.U. nicht erreichen, weil man für die gemeinsame Unsicherheit zweier Messgrößen auf eine prinzipielle untere Schranke stößt, die durch die Heisenbergsche Unschärferelation gesetzt ist (s. „Romanos Physikkurs 2“).

Wenn Physiker aus Resignation, Pragmatismus oder Einsicht aufgegeben haben, die Ursache dafür herauszufinden, warum sich zwei Messergebnisse unterscheiden, obwohl die Messungen nach Maßgabe der Möglichkeiten in gleicher Weise präpariert wurden, dann schreiben sie es dem Zufall zu. Wird zu den Schwankungen eines physikalischen Messwerts bei wiederholten Messungen die Feststellung getroffen, dass sie per Zufall eingetreten sind, dann möchte man damit kundtun, dass man es nicht mehr für sinnvoll hält, weiter nach einer kausalen Erklärung für unterschiedlich ausfallende Messresultate zu

suchen.

Standards für Messwert und Messunsicherheit. Liegt Datenmaterial aus sehr vielen Messungen unter gleichen oder systematisch variierten Bedingungen vor, dann kann die zufällig Verteilung der Messwerte durch eine statistische Analyse charakterisiert und quantifiziert werden. Die Analyse verbessert die Qualität der Messwertangabe und gibt Aufschluss darüber, in welchem Maße es gelungen ist, die unkontrollierbaren Einflussfaktoren zu reduzieren bzw. wie aussagekräftig das Messresultat ist. Manchmal gelingt es auch, durch die statistische Analyse systematischen Messfehlern auf die Spur zu kommen und sie zu eliminieren.

Nachfolgendes Beispiel soll die grundlegende Problematik und den Umgang mit Messunsicherheiten illustrieren: Ein Geometer habe die Aufgabe im Feld eine Strecke von genau 3 km mit einem Feldzirkel abzumessen und dort einen Grenzstein einzuschlagen. Genau? Nun, so genau er eben kann! Dazu wendet er den standardmäßig auf 2.000m eingestellten Zirkel 1500-mal an. Das ist dann eine Strecke mit einer NOMINALEN DISTANZ von 3 km. Wenn die Messung dieser Entfernung nun mehrfach, etwa von n unterschiedlichen Geometern durchgeführt wird, so wird es nicht allzu sehr überraschen, wenn die abgesteckten Endpunkte nicht übereinstimmen. Das wird auch dann zu erwarten sein, wenn die Geometer alle ihr Handwerk verstehen, ihre Feldzirkel präzise geeicht worden sind, etc., und sie also keine systematischen Messfehler begangen haben. Die Abweichungen sind nicht systematisierbar und daher als zufällig anzusehen. Jeder einzelne j -te Feldzirkelschritt $x_{i,j}$ der $j = 1, \dots, 1500$ Schritte des i -ten Geometers ($i = 1, \dots, n$) unterliegt dem Zufall und entspricht nur ungefähr der nominalen Schrittweite von 2.000m. In der Praxis wird die Schrittweite davon ein wenig nach oben oder nach unten hin abweichen, denn da muss die Strecke auch schon mal im Matsch eines strömendem Regens oder im winterlichem Schnee durch Äcker, Wiesen oder Wälder hindurch abgesteckt werden. Unter solchen Umständen wäre es offensichtlicher Unsinn, wenn man versuchen wollte, die unendlich vielen möglichen Einflussfaktoren akribisch zu erfassen oder zu berücksichtigen. Damit ist auch die vom i -ten Geometer gemessene Gesamtstrecke

$$X_i = \sum_{j=1}^{1500} x_{i,j}$$

zufallsbedingt, d.h. es werden sich i.A. n unterschiedliche Strecken X_i mit $i = 1, \dots, n$ (also X_1, X_2, \dots, X_n) von ungefähr 3 km ergeben. Sind die Abweichungen zufällig, so gibt es auch kein Kriterium a priori, welche dieser Strecken man als die Strecke von 3 km deklarieren sollte. Wenn nur eine Messung ($n = 1$) gemacht wurde, dann bleibt einem nichts anderes übrig als diese eine zu akzeptieren, aber wenn mehrere vorliegen, so gibt es mehrere Möglichkeiten, eine verlässlichere Schätzung (d.i. Messung) auszuwählen, als wenn man sich einen der Werte der n Geometer willkürlich herausgreifen würde. Eine davon ist der MEDIAN. Man bestimmt ihn, indem man die Messungen nach der Größe reiht und denjenigen nimmt, der in dieser Reihung der mittlere ist. Bei $n = 11$ ist beispielsweise der fünfte der elf Grenzsteine der Median.

Eine andere Möglichkeit ist der ARITHMETISCHEN MITTELWERT

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = X_1 + \frac{1}{n} \sum_{i=1}^n \Delta X_i, \quad (1.1)$$

wobei $\Delta X_i = X_i - X_1$ die Abweichung des i -ten Grenzsteins vom ersten ist, was man z.B. mit einem einfachen Metermaß mit einer Unsicherheit von 1 cm bestimmen kann.

Es ist sinnvoll, sich auf eine Konvention über Schätzungen zu einigen, so dass alle Physiker von derselben Größe reden, wenn sie einen Schätzwert angeben. Andernfalls wären die Schätzungen für Messwert und Abweichung auch nicht objektiv und fair miteinander vergleichbar.

Einen Messwert gibt man physikalisch korrekt immer in der Form

$$(\text{Messwert} \pm \text{Messunsicherheit})\text{Einheit} \quad (1.2)$$

an, also z.B.

$$(3000.00 \pm 0.15)\text{m}$$

Die Angabe besteht aus zwei Schätzwerten, die gemäß der geltenden Konventionen zu bilden sind, und am Ende gibt man immer die physikalische Einheit an. Die Begriffe „Messwert“ und „Messunsicherheit“ beziehen sich also auf Größen, die nach bestimmten nationalen oder internationalen Regeln bzw. Standards geschätzt wurden. Die derzeit gültige internationale Konvention ist im „Guide to the Expression of Uncertainty in Measurement“ festgelegt.

Der erste Schätzwert ist der STANDARDMESSWERT oder kurz Messwert. Wenn Daten aus mehreren Messungen zur Verfügung stehen, dann folgt man in der Physik meist der Konvention, den ARITHMETISCHEN MITTELWERT \bar{X} gemäß Gl. 1.1 als Messwert anzugeben.

Der zweite Schätzwert ist die nach den derzeit geltenden Konventionen berechnete MESSUNSICHERHEIT bzw. UNSICHERHEIT. Nun hat der Mittelwert logischerweise die Eigenschaft, dass die Abweichungen der einzelnen Messwerte einmal positiv und einmal negativ sind, so dass sie sich in Summe gerade herausheben. Daher folgt aus Gl. 1.1, dass

$$\sum_{i=1}^n (X_i - \bar{X}) = 0. \quad (1.3)$$

Aus diesem Grund geht man von der Summe der Quadrate der Abweichungen aus. Da wegen Gl. 1.3 nur $n - 1$ der n Summanden voneinander unabhängig sind, definiert man als Maßzahl der STANDARDISIERTEN KONVENTION für die Messunsicherheit die STANDARDABWEICHUNG

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}. \quad (1.4)$$

Wenn die Zahl n sehr groß ist, macht es keinen nennenswerten Unterschied, ob man durch n oder $n - 1$ dividiert. Daher verwendet man wegen

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + n\bar{X}^2 = \sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2$$

oft näherungsweise

$$\sigma \approx \sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2} \quad (1.5)$$

als Schätzwert der Standardabweichung, denn das lässt sich schneller berechnen. Das Verhältnis σ/\bar{X} wird **RELATIVE ABWEICHUNG** genannt.

Eine weitere Konvention betrifft die Angabe der Anzahl der signifikanten Ziffern bei der Ablesung von einem Messinstrument oder bei der Angabe eines Messergebnisses. Sie werden oben vielleicht gestutzt haben, als ich für die zwei Meter des Feldzirkels immer 2.000m angegeben habe. Warum nicht eine Null mehr oder eine weniger? Der Grund liegt darin, dass der Feldzirkel nur auf Millimeter genau geeicht bzw. gefertigt ist. Die Konvention ist, dass die vorletzte Ziffer, die man angibt, sicher sein sollte. Die letzte Ziffer kann im Rahmen des Unsicherheitsbereichs schwanken. Wenn die Fertigungstoleranz z.B. die Feldzirkel nur bis auf $\pm 3\text{mm}$ Genauigkeit herzustellen gestattet, dann würde man angeben: $(2.000 \pm 0.003)\text{m}$ bzw. eine Toleranz von 1.5‰. Wenn man mehr nicht weiß und keine weitere verlässliche Information über die Unsicherheit hat, weil keine statistische Auswertung vorgenommen wurde, würde man also beispielsweise die Messunsicherheit mit 1.5‰ abschätzen und die Gesamtstrecke also mit 3.000km oder besser mit der Behelfsschätzung $(3.000 \pm 0.0045)\text{km}$ angeben.

Um ein Gefühl dafür zu bekommen, wie sich die Formel für die Standardabweichung auswirkt, betrachten wir fünf Geometer A, B, C, D und E, welche die Strecke von 3 km abstecken. Die Distanz X_A , gemessen von Geometer A, sei dabei die geringste.

1. **Beispiel:** Von X_A weichen die anderen gemessenen Distanzen folgendermaßen nach oben hin ab:² $\Delta X_B = 0$, $\Delta X_C = 20\text{ cm}$, $\Delta X_D = \Delta X_E = 40\text{ cm}$. Das Symbol Δ steht hier für die Bildung von Differenzen. Die mittlere Abweichung von X_A ist $(2 \cdot 0 + 20 + 2 \cdot 40)/5\text{ cm} = 20\text{ cm}$ und daher ist der Grenzstein für die nominale Entfernung 3 km auf der Position $\bar{X} = X_A + 20\text{ cm}$. Die Summe aller Abweichungen von diesem Mittelwert ist null, aber für die Quadrate errechnet man $\sum (X_i - \bar{X})^2 = 1600\text{ cm}^2$ und somit eine Standardabweichung $\sigma = 20\text{ cm}$.
2. **Beispiel:** Von X_A weichen die anderen Distanzen folgendermaßen nach oben hin ab: $\Delta X_B = 10\text{ cm}$, $\Delta X_C = 20\text{ cm}$, $\Delta X_D = 30\text{ cm}$, $\Delta X_E = 40\text{ cm}$. Der Mittelwert \bar{X} , dem wir die nominale Entfernung 3 km zuordnen, liegt dann ebenfalls auf der Markierung $\bar{X} = X_A + 20\text{ cm}$. Die Summe aller Abweichungen von diesem Mittelwert

²Selbst wenn Sie diese Abweichungen auf einen Zehntel Millimeter genau messen könnten, würde es für das Ergebnis keinen Sinn machen, das so genau anzugeben. Eine Angabe auf Zentimeter, das zeigt die nachfolgende Analyse, reicht völlig aus.

ist gleichfalls null. Man errechnet aber $\sum(X_i - \bar{X})^2 = 1000 \text{ cm}^2$ und somit eine Standardabweichung $\sigma \approx 16 \text{ cm}$.

Im ersten Beispiel würde man die Messung mit $3.0000 \text{ km} \pm 20 \text{ cm}$ und im zweiten mit $3.0000 \text{ km} \pm 16 \text{ cm}$ angeben. Die Angabe $\pm 20 \text{ cm}$ bzw. $\pm 16 \text{ cm}$ stellt die Unsicherheit dar, mit welcher die Zuverlässigkeit der Messung hier quantifiziert wurde. Bei beiden Beispielen ist die zweite Ziffer der Messunsicherheit die statistisch unsichere, und man gibt die Unsicherheit bis auf eben diese Ziffer gerundet an. Mehr Ziffern sind nicht nötig, aber weniger dürfen es auch nicht sein. Jedenfalls sind das die Konventionen, an die man sich als Physiker halten sollte, wenn man ein Messergebnis fachmännisch angeben will. Man sollte aber weder das Messresultat für die tatsächliche Entfernung nehmen noch die statistische Abweichung für die tatsächliche Abweichung. Aufgrund der Fertigungstoleranz des Feldzirkels kann die Abweichung von der tatsächlichen Entfernung weit größer ausfallen – und das gilt auch für die tatsächliche Abweichung. Bei der Angabe der Unsicherheit sollte man also i.A. erläutern, wie man zu ihrer Beurteilung gekommen ist. Man darf sich nicht einfach auf die stumpfe Anwendung einer Formel zurückziehen. Bei der Analyse eines Messresultats ist das kritische Urteilsvermögen gefragt.

Da hier recht ausführlich auf zufallsverteilte Messresultate eingegangen wurde, mag der Eindruck entstehen, man müsste stets eine größere Messreihe durchführen, um ein Messresultat gültig angeben zu können. Das ist nicht nur zeitraubend, sondern oft auch unnötig. Eine plausible Schätzung kann man fallweise auch anders gewinnen: Ein Schullineal hat beispielsweise eine Millimeterteilung, und man kann bestenfalls auf einige Zehntel Millimeter abschätzen. Eichfehler des Lineals und andere Einflüsse werden nicht viel größer ausfallen. Daher ist es auch ohne Durchführung einer Messreihe eine gute Schätzung, wenn man die Messunsicherheit z.B. mit $\pm 0.5 \text{ mm}$ spezifiziert.

Zu langen Messreihen mit einer anschließenden statistischen Analyse wird man greifen, wenn beispielsweise folgende Gründe vorliegen: Wenn man sonst keine Handhabe hat, die Unsicherheit ausreichend genau abzuschätzen, wenn man feineren Einflüssen auf die Spur kommen will oder eine besonders hohe Verlässlichkeit der Angabe zur Schätzgenauigkeit anstrebt. Langwierige statistische Untersuchungen sind beispielsweise unumgänglich, wenn man die Güte unterschiedlicher Messgeräte bzw. Messverfahren charakterisieren will, um das geeignete Verfahren für ein Messvorhaben auszuwählen. Vorausgesetzt, dass die Messungen nach jedem der zu vergleichenden Verfahren fehlerfrei und fachmännisch durchgeführt werden, ist dasjenige Verfahren als das beste zu bewerten, welches die kleinste Unsicherheit aufweist.

1.6. Präzisierung

Alle Überlegungen und Schlussfolgerungen gingen bis hier hin von der Hypothese der Geradlinigkeit der Lichtausbreitung aus. Diese Hypothese hat sich in vielen der bis hier hin besprochenen Situationen hervorragend bewährt. Trotzdem kann es geschehen, dass man mit einer Beobachtung konfrontiert wird, die einer bewährten Hypothese widerspricht. Wenn man beispielsweise Zucker in einen Trog Zucker schüttet und ihn mit Wasser auffüllt, so verläuft der Lichtstrahl darin nicht mehr geradlinig, sondern gekrümmt (Abb. 7).

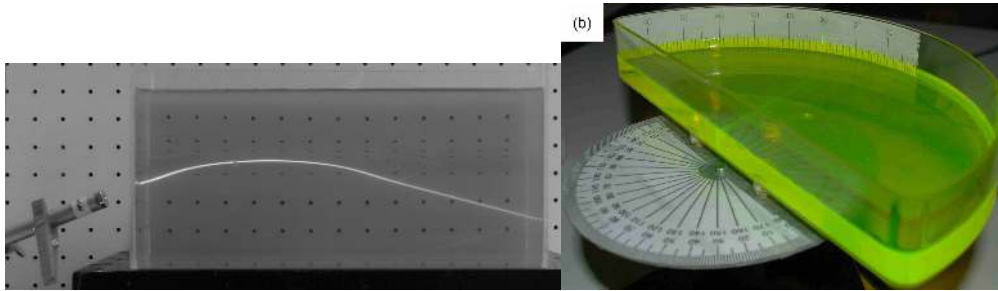


Abbildung 7: (a) Gekrümmter Lichtstrahl in einem inhomogenen Medium. Der von einem Laser (links) ausgehende Strahl durchquert einen Tank mit einer Zuckerslösung, deren Konzentration von oben nach unten hin zunimmt. (b) Lichtbrechung. Die obere Hälfte eines Strahls mit strichförmigem Profil verläuft in Luft, für die untere Hälfte findet Brechung beim Übergang in Wasser statt. Oben kann der Einfallswinkel und unten und der gebrochene Winkel ($\beta \approx 22.5^\circ$) abgelesen werden.

Wie gehen Physiker mit dem Widerspruch zur um? Das bis hierhin Bewährte will man nicht gleich über Bord werfen, denn es war ja eine äußerst erfolgreiche Naturbeschreibung. Die Ursache für den gekrümmten Lichtstrahl ist leicht zu erkennen: Die Zuckerkonzentration nicht mehr überall gleich ist. Sie nimmt von oben nach unten hin zu, d.h. das Medium ist inhomogen. Durch weitere Untersuchungen kann man feststellen, dass sich das Licht dem ideal geradlinigen Verlauf um so mehr annähert, je homogener das Medium ist, durch das es verläuft.

Aus dieser Erfahrung heraus wagt man eine neue Hypothese, eine Hypothese über ein idealisiertes Medium, das (absolut) homogene Medium:

Lichtwege sind Geraden, wenn sie in einem homogenen Medium verlaufen.

Diese nachgebesserte Hypothese ersetzt nun die ursprüngliche Hypothese „Lichtwege sind Geraden“ aus Abschnitt 1.1. Diese Art der evolutionären PRÄZISIERUNG einer physikalischen Hypothese ist, wie sich noch an vielen Beispielen zeigen wird, ein Grundmuster beim Fortschreiten physikalischer Erkenntnis. Interessant ist die Hypothese auch aus dem Grund, weil sie sich auf ein Medium bezieht, das es in dieser Perfektion nicht gibt. Keines der Medien, mit denen man im Alltag zu tun hat (z.B. Gläser oder die Luft der Atmosphäre), ist bei strenger Betrachtung absolut homogen. Die Hypothese ist eine Aussage über ein idealisiertes Modellmedium, das für weitere Schlüsse zugrunde gelegt wird.

Ein wichtiger Spezialfall der Lichtausbreitung in inhomogenen Medien ist der Fall, wo zwei unterschiedliche homogene Medien A und B eine ebene Grenzfläche miteinander haben. In den jeweils homogenen Bereichen, d.h. innerhalb der Medien A und B, ist die Lichtausbreitung geradlinig. Die Grenzfläche stellt jedoch eine massive Inhomogenität dar. Hier ändert sich der Verlauf des Lichts dramatisch (Abb. 7b). In dem Medium, von dem her der Lichtstrahl einfällt, beobachtet man reflektierte Strahlen. Wenn das zweite

Medium transparent ist, beobachtet man auch TRANSMITTIERTE Strahlen. Im Rahmen der Geometrischen Optik sind erst einmal die rein geometrischen Aspekte des Phänomens von Interesse.

1.6.1. Bezugsэлеmente und Konventionen für Winkel

Flächennormale und Einfallsebene. Alle an Reflexion und Brechung beteiligten Strahlen liegen in einer Ebene. Sie wird als EINFALLSEBENE bezeichnet. In ihr liegt auch die FLÄCHENNORMALE der Grenzfläche. Daher bietet es sich an, die Flächennormale als Bezugslinie zu wählen und die Richtung der Strahlen durch Winkel relativ zur Flächennormale anzugeben. Abb. 8 skizziert einen Strahl, der zunächst in einem Medium A verläuft und unter einem Winkel α auf die Grenzfläche G auftrifft. Dort kann er unter dem Winkel α_r ins Medium A zurück reflektiert werden oder er tritt unter einem Winkel β ins Medium B über.

Drehsinn. Als Konvention für den DREHSINN bzw. für Winkel wird im Folgenden die rechtshändige Vorzeichenkonvention verwendet: Setzt man die rechte Hand so auf die in Abb. 8b gezeichnete Einfallsebene, dass der Daumen nach oben zeigt, dann sind die Finger im mathematisch positiven Drehsinn gekrümmt. Winkel sind positiv, wenn sie bei Draufsicht gegen den Uhrzeigersinn verlaufen. In Abb. 8a sind die Winkel α und β negativ, in Abb. 8b positiv.

Ablenkwinkel. Die in Abb. 8b gestrichelte Linie ist die jeweilige Verlängerung des ursprünglichen Strahlverlaufs. Wird der Strahl beim Übertritt von einem Medium A in ein Medium B nach rechts und somit zur Flächennormale hin abgelenkt, so bezeichnet man das Medium A als das optisch dünnere Medium und B als das optisch dichtere. Der durch $\delta = \beta - \alpha$ gegebene ABLENKWINKEL ist dann negativ. Für den umgekehrten Strahlverlauf, also vom optisch dichteren Medium B ins optisch dünnere Medium A, ist β der Einfallswinkel und α der aus der Brechung resultierende Winkel. Daher ist der Ablenkwinkel $\delta = \alpha - \beta$ betragsmäßig gleich groß, aber positiv. Tritt ein Strahl durch ein System von mehreren Grenzflächen, so ist der Ablenkwinkel des Gesamtsystems gleich dem Ablenkwinkel des von der letzten Grenzfläche ausgehenden Strahls relativ zur Richtung des auf die erste Grenzfläche einfallenden Strahls. Die Gesamtablenkung ist gleich der Summe der Ablenkwinkel an jeder der Grenzflächen.

1.6.2. Reflexionsgesetz

Reflektierte Strahlen treten nur einfallsseitig auf. Im Fall, der in 8 gezeigt ist, fällt der Strahl im Medium A ein, und dort tritt dann auch der reflektierte Strahl unter einem Winkel α_r auf. Für diesen findet man das REFLEXIONSGESETZ

$$\alpha_r = -\alpha. \quad (1.6)$$

Für einen aus B kommenden Strahl wird der Strahl ins Medium B zurückreflektiert, und zwar mit dem Winkel $\beta_r = -\beta$ (Abb. 8b).

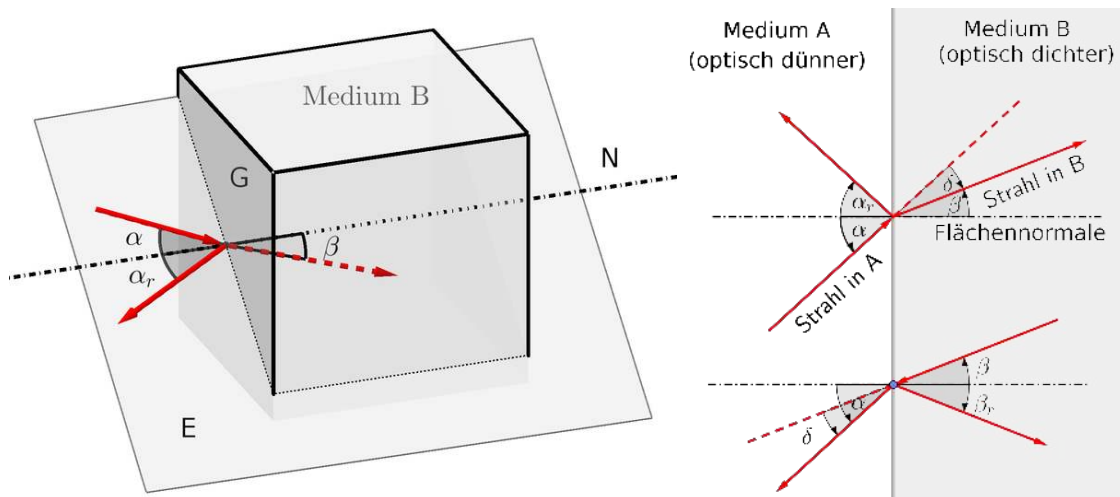


Abbildung 8: (a) Ein aus dem Medium B bestehender Würfel ist in eine aus dem Medium A bestehende Umgebung eingebettet. An der Grenzfläche G (dunkel hervorgehoben) treten Reflexion und Brechung auf. Alle Strahlen sowie die Flächennormale N (strichpunktierte Linie) liegen in der Einfallsebene E. Die Winkel α (Winkel im Medium A), α_r (Reflexionswinkel) und β (Winkel im Medium B) werden relativ zur Flächennormale gemessen. (b) Draufsicht auf die Einfallsebene. Verlauf des transmittierten Lichtstrahls beim Übergang von einem optisch dünneren Medium A zu einem optisch dichteren Medium B (oben) und umgekehrt (unten). Bezugslinien sind die Flächennormale (strichpunktiert) bzw. die Verlängerung des einfallenden Strahls (gestrichelt). Der Drehsinn der Winkel relativ zu ihrer jeweiligen Bezugslinie ist durch die Pfeilrichtung kenntlich gemacht.

1.6.3. Brechungsgesetz

Wenn sich die Winkel α und β unterscheiden, spricht man von einer BRECHUNG des Lichts bzw. von einem gebrochenen Strahl. Es ist dabei so, dass der Strahlverlauf von A nach B auf dem gleichen Weg erfolgt wie jener von B nach A. Man sagt dazu, dass Lichtwege „umkehrbar“ sind (Abb. 8b).

Die älteste uns bekannte experimentelle Untersuchung des Zusammenhangs der beiden Winkel wurde im 2. Jh. u.Z von Claudius Ptolemäus durchgeführt. Abb. 9a gibt seine Messresultate wieder. Ptolemäus gelang es nicht, das dahinter stehende Gesetz anzugeben. Es ist jedoch erkennbar, dass α und β im Grenzfall immer kleinerer Winkel proportional zueinander werden (Abb. 9b). Daher kann man für kleine Winkel die Geometrie der Brechung näherungsweise durch

$$\alpha \approx n_{AB}\beta. \quad (1.7)$$

beschreiben (PTOLEMÄISCHES BRECHUNGSGESETZ). Mit Hilfe dieses Brechungsgesetzes

konnte man schon sehr früh ein erstes Verständnis für grundlegende brechende optische Elemente und darauf aufbauende Instrumente entwickeln. Für die nachfolgenden Betrachtungen werden wir, der historischen Entwicklung folgend, zunächst einmal vom Ptolemäischen Brechungsgesetz (Gl. 1.7) ausgehen.

Die in Gl. 1.7 auftretende Proportionalitätskonstante n_{AB} heißt BRECHZAHL bzw. BRECHUNGSINDEX. Sie hängt von den beiden Materialien A und B ab, die an der Grenzfläche zusammentreffen. Wertet man die Daten von Ptolemäus für kleine Winkel aus, so ergibt sich beispielsweise für die Grenzfläche Glas-Luft (A=G(las), B=L(uft)) $n_{AB} = n_{LG} \approx 1.5$. Kommt der Lichtstrahl umgekehrt aus dem Medium B, so gilt für die Winkel der gleiche Zusammenhang, nämlich Gl. 1.7, d.h. es ist $\beta \approx n_{BA}\alpha$ mit $n_{BA} = 1/n_{AB}$. Ist $n_{AB} > 1$, so bezeichnet man das Medium A als das optisch dünnere Medium und B als das optisch dichtere. Empirisch zeigt sich, dass die Brechzahlen jeweils dreier Medien A, B und C miteinander in der Beziehung

$$n_{AB} = \frac{n_{AC}}{n_{BC}} \quad (1.8)$$

stehen. Daher kann man eine Tabelle der Brechzahlen relativ zu einem gemeinsamen Referenzmedium V aufstellen und erhält dann die relative Brechzahl zwischen zwei anderen Medien aus Gl. 1.8. Als gemeinsames Referenzmedium V hat man das Vakuum gewählt, d.h. die Brechzahlen sind üblicherweise relativ zum absoluten Vakuum (als extrapolierter Grenzfall) tabelliert und werden dann als n_A, n_B, \dots bezeichnet, wobei man den Index V für das gemeinsame Referenzmedium unterdrückt. Diese Brechzahlen können als Materialeigenschaften angesehen werden.

In den nachfolgenden Kapiteln werden stets Grenzflächen zwischen Luft und Glas betrachtet und der relative Brechungsindex n_{LG} für den Übergang vom optisch dünnen Medium Luft in das optisch dichte Medium Glas einfach mit n bezeichnet. Für einen zunächst in Luft verlaufenden Strahl, der in Glas eintritt, ist der Ablenkwinkel

$$\delta = \beta - \alpha \approx (1 - n)\beta. \quad (1.9)$$

negativ. Für einen zunächst im Glas verlaufenden Strahl, der dann in Luft übertritt, ist $\delta = \alpha - \beta \approx (n - 1)\beta$ und somit positiv.

1.6.4. Brechung am Prisma

Das Prisma ist ein brechendes optisches Element mit planen Grenzflächen, das es ermöglicht, den Verlauf von Lichtstrahlen in vorhersagbarer Weise zu beeinflussen. Abb. 10a zeigt die Brechung an einem durch zwei ebene Flächen begrenztem Medium B, das optisch dichter als seine Umgebung A ist, d.h. das Medium hat eine Brechzahl $n > 1$ relativ zur Umgebung. Wenn sich die beiden Grenzflächen unter einem Winkel $\varepsilon \neq 0$ schneiden, so stellt dies ein PRISMA dar. Ist $\varepsilon = 0$, dann sind die beiden Grenzflächen parallel zueinander, und man spricht von einer PLANPARALLELEN PLATTE, .

Sind β_1 bzw. β_2 die Winkel des im Medium verlaufenden Strahls relativ zu den beiden Flächennormalen, so sind umgekehrt gesehen $-\beta_1$ bzw. $-\beta_2$ gerade die Winkel, um welche

α	β		
	Luft-Wasser	Luft-Glas	Wasser-Glas
10°	8.0°	7.0°	9.5°
20°	15.5°	13.5°	18.5°
30°	22.5°	19.5°	27.0°
40°	28.0°	25.0°	35.0°
50°	35.0°	30.0°	42.5°
60°	40.5°	34.5°	49.5°
70°	45.5°	38.5°	56.0°
80°	50.0°	42.0°	62.0°

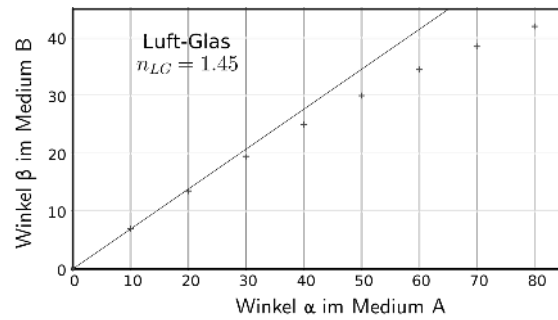


Abbildung 9: Vor fast 2000 Jahren durch Ptolemäus ermittelte Messwerte für die Lichtbrechung [2]. Im Diagramm rechts ist der Winkel α in Luft gegen den Winkel β in Glas aufgetragen. Die Ausdehnung der Kreuze entspricht einer Messunsicherheit von $\pm 0.5^\circ$.

die beiden Grenzflächen gegen den Strahl geneigt sind. Die Grenzflächen sind daher um $\varepsilon = \beta_1 - \beta_2$ relativ zueinander geneigt. Beim Übertritt vom optisch dünnen ins optisch dichte Medium wird der Strahl gemäß Gl. 1.9 um $\delta_1 \approx (1 - n)\beta_1$ abgelenkt und beim Übertritt vom optisch dichten ins optisch dünne Medium um $\delta_2 \approx (n - 1)\beta_2$. Insgesamt erfährt der Strahl eine Ablenkung um $\delta = \delta_1 + \delta_2 \approx (1 - n)(\beta_1 - \beta_2)$. Daraus folgt für die Gesamtablenkung

$$\delta \approx (1 - n)\varepsilon. \quad (1.10)$$

Im Rahmen der Ptolemäischen Näherung für das Brechungsgesetz kommt man somit zu dem bemerkenswerten Schluss, dass die Ablenkung eines Prismas nicht von β_1 abhängt und somit auch nicht vom Einfallswinkel α . Dazu muss u.a. der Prismenwinkel ε ausreichend klein sein.

Bei einer planparallelen Platte ($\varepsilon = 0$) kommt es insgesamt zu keiner Ablenkung, weil sich die Ablenkungen δ_1 und δ_2 kompensieren (Abb. 10b). Aber es tritt ein Parallelversatz des Strahls um eine Strecke Δs auf, der mit zunehmendem Einfallswinkel α und zunehmender Plattendicke anwächst.

Die von der Brechzahl abhängige Ablenkung durch ein Prisma lässt sich demonstrieren, indem man eine prismenförmige Glasküvette in einen wassergefüllten Glastrog einsetzt (Abb. 11). Wenn die Küvette mit Luft ($n < 1$) gefüllt ist, erfolgt die Ablenkung eines Laserstrahls in gleicher Drehrichtung wie ε , wenn sie mit Glyzerin gefüllt ist ($n > 1$), in entgegengesetzter Drehrichtung. Wird sie mit Wasser gefüllt, tritt keine Ablenkung auf. Aus der Ablenkung kann die relative Brechzahl n bestimmt werden.

1.6.5. Das Snellius'sche Brechungsgesetz

Das Christentum, das im Jahre 380 n. u. Z. zur Staatsreligion im römischen Reich wurde, hat wesentlich zum Verlust des Wissens der antiken Naturwissenschaften beigetragen. Das byzantinische Reich „saß“ auf den antiken Quellen, aber das Christentum machte der

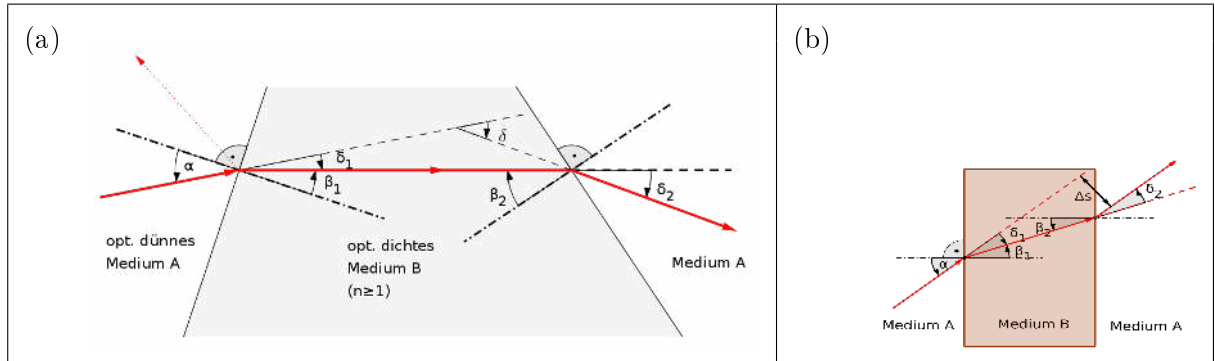


Abbildung 10: (a) Strahlableitung durch ein Prisma. (b) Brechung an einer planparallelen Platte. In beiden Fällen ist angenommen, dass das Medium B das optisch dichtere Medium ist, d.h. $n > 1$.

antiken Wissenschaftskultur in wenigen Jahrzehnten den Garaus. Die letzten Vertreter der hellenistischen Philosophie und Wissenschaft wurden von christlichen Geistlichen oder Herrschern verfolgt, gefoltert, hingerichtet oder ermordet. Die Ermordung der Physikerin und Mathematikerin Hypatia im Jahr 415 steht stellvertretend dafür, wie auch die durch die griechische Philosophie denkbar gewordenen ersten Ansätze einer Frauenemanzipation durch religiöse Eiferer systematisch zunichte gemacht wurden. Sankt Justinian – er wird von der orthodoxen Kirche als christlicher Heiliger verehrt – ließ nicht nur „heidnische“ Schriften der antiken Mathematiker, Physiker und Philosophen verbrennen, sondern ließ auch die letzte noch lehrende Platonische Akademie seines Kaiserreiches schließen. Für die Wissenschaftskultur des Abendlandes brach damit ein mehr als tausendjähriges dunkles Zeitalter an.

Vier Jahrhunderte nach dieser kulturellen Katastrophe im Abendland ging im Morgenland durch das Aufkommen einer im Vergleich zum damaligen Christentum wesentlich toleranteren Religion wieder die Sonne auf: Zwischen 800 und 1200 kamen Wissenschaft und Kultur durch die Wiederentdeckung der antiken hellenistischen Quellen im „*Goldenen Zeitalter des Islam*“ wieder zur Blüte (Abb. 12a). Durch den Gebrauch der arabischen Ziffern, der Einführung der Zahl 0 und des Dezimalsystems durch Al-Chwarizmi (أبو إسحاق إبراهيم بن محمد بن أدهم الخوارزمي) sowie durch die nach ihm benannten „Algorithmen“ bzw. „Algorithmen“ werden kompliziertere algebraische Rechnungen möglich (Das Wort Algebra geht auf den Titel seines 825 erschienenen arabischen Buches „al-gabr“ zurück).

Die antike geometrische Optik konnte die Brechung nur für kleine Winkel gut beschreiben. Die Diskrepanz zwischen den Messungen und dem approximativen Ptolemäischen Brechungsgesetz sind jedoch aus Abb. 9 deutlich erkennbar. Für Rechnungen mit größeren Winkeln müssen die trigonometrischen Funktionen bekannt sein, wie z.B. $\cos \alpha$, $\sin \alpha$, oder $\tan \alpha$.³ Im 10. Jahrhundert führte der persische Wissenschaftler Abu l-Wafa die Tangensfunktion ein, bewies den Sinussatz und erstellte Tabellen der trigonometri-

³Zur Konvention: Eigentlich $\cos(\alpha)$, d.h. das Funktionsargument α sollte in Klammern hinter dem Funktionsbezeichner stehen. Wenn es nicht zu Missverständnissen kommen kann, schreibt man bei trigonometrischen Funktionen stattdessen oft nur $\cos \alpha$. Bei der Exponentialfunktion $e^\alpha = \exp(\alpha)$ lässt man die Klammern auf der rechten Seite hingegen nicht wegfallen.

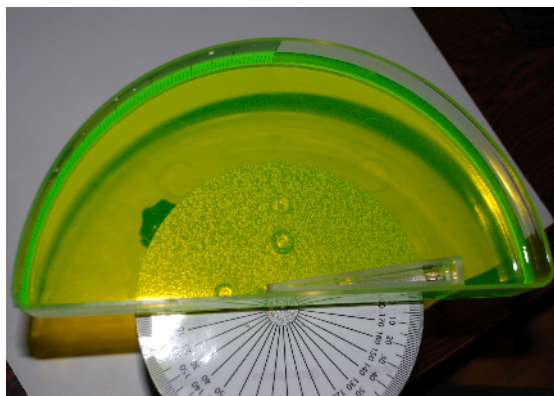


Abbildung 11: Demonstration der Prismenablenkung. Ein Laserstrahl wird durch eine halbkreisförmige Wasserwanne hindurch gerichtet. In den Strahlweg wird eine prismenförmige Küvette eingesetzt. Füllt man die Küvette mit Wasser, findet keine Ablenkung statt, wohl aber, wenn man sie mit Luft oder Glycerin füllt. Im ersten Fall sind der Drehsinn des brechenden Winkels und der Ablenkung gleich gerichtet, im zweiten Fall entgegengesetzt.

schen Funktionen. Im Jahr 984 fand Ibn Sahl (إيبن سَهْل، ابْنُ سَهْل) die korrekte Form des Brechungsgesetzes für zwei Medien mit den Brechzahlen n_A und n_B :

$$n_A \sin \alpha = n_B \sin \beta. \quad (1.11)$$

Für kleine Winkel gilt im Bogenmaß näherungsweise (Physiker rechnen Winkel fast immer im Bogenmaß)

$$\sin \alpha \approx \alpha, \quad (1.12)$$

d.h. das Ptolemäischen Brechungsgesetz, Gl. 1.7, ist als Grenzfall kleiner Winkel im allgemeineren Brechungsgesetz enthalten. Dies ist ein Musterbeispiel für den evolutionären Charakter der historischen Entwicklung der Physik: Ältere, ungenaue Gesetze oder Theorien werden in der Weise abgelöst, dass sie als Grenzfall in den neuen, genaueren Gesetzen oder Theorien bestehen bleiben. Dies liegt daran, dass sich die Naturphänomene, die zur älteren Theorie geführt haben, ja nicht geändert haben.

Aufkommender religiöser Fundamentalismus beendete zwar die Blütezeit der arabisch-persischen Naturwissenschaften, doch das Wissen wurde erfolgreich an den Okzident weitergereicht. Eine zentrale Rolle für die Übertragung der arabischen Wissenschaftskultur in den europäischen Raum spielte das Kalifat von Cordoba. In der Folge entstand durch Renaissance und Humanismus eine Neuorientierung an der Antike. Reformation und Aufklärung führten schließlich auch das Abendland aus einer langen Periode der Dunkelheit heraus. Das heute als SNELLIUS'SCHE BRECHUNGSGESETZ bekannte Gesetz, Gl. 1.11, wurde erst sechs Jahrhunderte später von Willebrord Snell für den abendländischen Kulturraum wiederentdeckt, denn das Wissen über die trigonometrischen Funktionen erreichte das christliche Abendland erst relativ spät. Die erste Sinustabelle des Abendlandes



Abbildung 12: (a) Blütezeit der arabisch-persischen Wissenschaften: Al-Chwarizmi (Algebra), Abu-I-Wafa (Tangensfunktion), Al-Biruni, Abu Sad (Bestimmung des Erdradius), Al-Haitham (Erfindung der Lupe), Al-Tusi (Trigonometrische Funktionen, Massenerhaltungssatz in Physik und Chemie). (b) Peuerbachs Sonnenuhr an der Südseite des Wiener Stephansdoms.

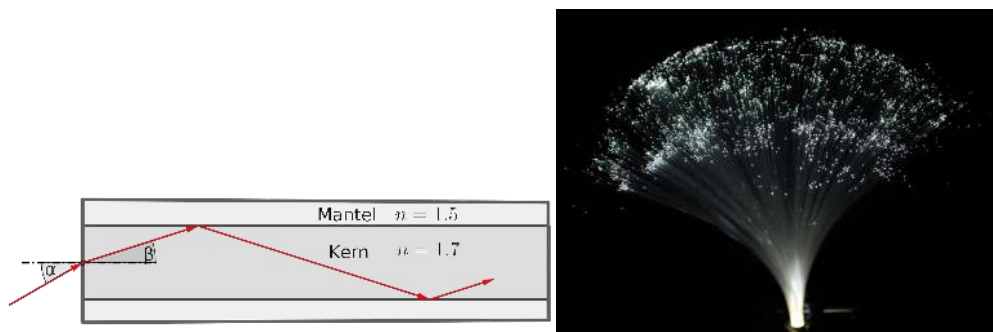


Abbildung 13: (a) Lichtleitung durch Totalreflexion in einem Kern mit höheren Brechzahl als die des Mantels. Der Eintrittswinkel α an der Stirnfläche muss so klein sein, dass $90^\circ - \beta$ den Winkel α_T der Totalreflexion nicht überschreitet. (b) Demonstration der Lichtleitung durch Glasfasern.

wurde beispielsweise erst Mitte des 15. Jahrhunderts durch Georg von Peuerbach erstellt (Abb. 12b).

Totalreflexion. Wenn im Snellius'schen Brechungsgesetz, Gl. 1.11, das Medium A das optisch dichtere Medium ist, gibt es für Einfallswinkel $\alpha > \alpha_T$ keine Lösungen mehr. Der Strahl kann nicht mehr gebrochen werden und wird vollständig (total) reflektiert. Der Winkel α_T ist durch

$$\sin \alpha_T = n_B/n_A \quad (1.13)$$

bestimmt und heißt Grenzwinkel der Totalreflexion. Diese Totalreflexion ist beispielsweise die physikalische Ursache des Naturphänomens der Fata Morgana. Die Messung des Reflexionswinkels α_T der Totalreflexion ist auch die Grundlage für Methoden zur raschen und präzisen Bestimmung der Brechzahl. Die Brechzahl lässt sich beispielsweise mit dem

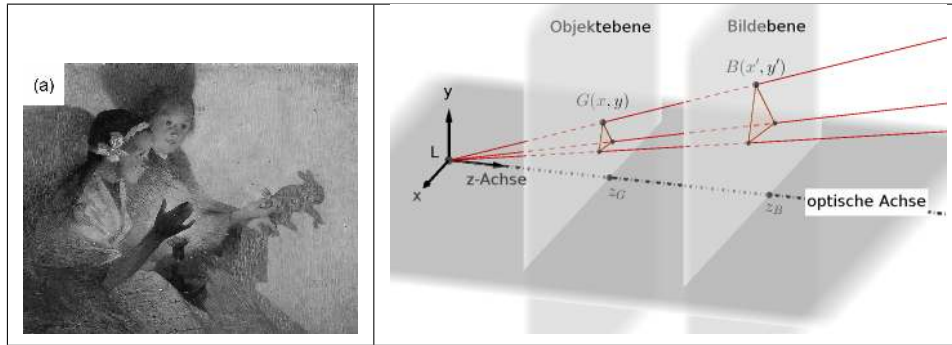


Abbildung 14: (a) Gemälde von Puigaucau: Chinesisches Schattentheater. (b) Modell des Schattenwurfs. Konstruktion des durch eine punktförmige Lichtquelle L von einem Objekt G entworfenen Schattenbildes B . Die gestrichelte Linie ist die optische Achse des Abbildungssystems.

Abbe-Refraktometer⁴ auf 4–5 Stellen genau messen. Solche Geräte werden beispielsweise in der Chemie, Pharmazie oder bei der Lebensmittelkontrolle eingesetzt.

Optische Datenübertragung. Die Totalreflexion wird zur verlustarmen Lichtleitung verwendet. Lichtleitfasern bestehen aus einem hochbrechenden Kern und einem Mantel mit geringerer Brechzahl (Abb. 13). Sie ermöglichen eine Informationsübertragung mit einer höheren Datenrate als es mit kabelgebundenen elektrischen Signalen möglich ist.

2. Bilder

2.1. Das physikalische Bild

„Wir machen uns innere Scheinbilder oder Symbole der äußeren Gegenstände, und zwar machen wir sie von solcher Art, dass die denknotwendigen Folgen der Bilder stets wieder die Bilder seien von den naturnotwendigen Folgen der abgebildeten Gegenstände.“ (Heinrich Hertz: Die Prinzipien der Mechanik, 1894)

Die Klärung des Begriffs Bild soll in diesem Kapitel auf zwei Ebenen geführt werden. Zum einen soll diskutiert werden, was ein BILD bzw. eine ABBILDUNG im landläufigen Sinn ausmacht, und zum zweiten, was mit einem PHYSIKALISCHEN BILD bzw. physikalischen Modell gemeint ist.

Obwohl das in Abb. 14a gezeigte Bild nur in einigen wenigen Charakteristika der Farben und Formen mit realen Objekten übereinstimmt, rekonstruiert unser Verstand es so, dass die Darstellung für uns einen Sinn ergibt. Der Bezug, den wir zwischen Bild und gedachtem Gegenstand herstellen, ist unsere Interpretation. Unsere Erfahrung kann z.B.

⁴Ernst Abbe, Physiker und Sozialreformer, gründete die Carl-Zeiss-Stiftung und gehörte zu den ersten Unternehmern, die den Achtstundentag für ihre Arbeiter einführten.

nahelegen, auf dem Bild zwei Kinder zu erkennen, die das Schattenbild eines Hasen auf eine Projektionsfläche werfen. Dass dies aber nur eine der möglichen Interpretationen ist, wird klarer, wenn man den Aspekt des Schattenwurfs herausgreift. Wenn man nichts weiter als den Schatten sähe, der auf der Projektionsfläche entsteht, kann man sich ihn durch einen Hasen hervorgerufen denken oder durch eine bestimmte Konfiguration zweier Hände. Beides kann man in den Schatten hineininterpretieren. Bei Vexierbildern sticht die Ambivalenz möglicher Interpretationen besonders deutlich hervor.

Bilder müssen genügend Merkmale haben, um zu erreichen, dass ein bestimmtes Objekt darin wiedererkannt wird kann, dass das Bild mit einer bereits bekannten Erfahrung identifiziert werden kann oder dass es im Verstand gewisse Vorstellungen bzw. Assoziationen auslöst, die eine sinnvolle Deutung ermöglichen. In Abb. 14a macht sich das Wiedererkennen beispielsweise an der Identifikation der ähnlichen Form fest. Im Extremfall genügt es, wenn Bilder nur in einem einzigen für die Interpretation wesentlichen Merkmal mit dem Objekt übereinstimmen. Weitere Übereinstimmungen müssen sie nicht unbedingt haben. Diese Erfahrung kann man leicht beim Betrachten von Bildern moderner Maler machen. Wenn man etwas als Bild bezeichnet, dann ist es insbesondere keine identische Kopie bzw. ein Klon des Objekts. Bilder lassen i.A. nicht zu, eindeutig auf den Gegenstand zurück zu schließen und können daher i.A. auch mehr als eine Deutung haben. Sie erfüllen ihren Zweck, wenn sie Sinn stiften.

Für Physiker ist das Denken in Bildern schon auf Grund der Einsicht eine Selbstverständlichkeit, dass ihnen die vollständige Realität eines „Ding an sich“, eines Objekts der Natur, wegen der Begrenztheit unserer Sinneswahrnehmungen in der Regel verborgen bleibt. Daran ändert auch der Einsatz von Instrumenten zur Erweiterung unserer Sinneswahrnehmungen nichts. Physiker können nur Aussagen über die Erscheinungen treffen, d.h. nur über das sprechen, wie ihnen gegenüber die Dinge in Erscheinung treten (s. auch Kap. 12.7). Von diesen Erscheinungen der Dinge machen sie sich Bilder, untersuchen deren Relationen und versuchen die sich daraus ergebenden Konsequenzen zu verstehen. Diese Verstandestätigkeit spielt sich meist in einer meist mathematischen Modellwelt ab. Das physikalische Bild der Realität mag manchmal nur eine sehr skizzenhafte Karikatur der Wirklichkeit sein. Die Begriffe der Modellwelt können völlig abstrakt sein, und es kann sogar so sein, dass nur die Beziehungen zwischen den Begriffen Bedeutung haben. Die Modellwelt muss insbesondere nicht alle Details der Realität abbilden. Aber das physikalische Modell muss so klar definiert sein, dass man darüber logische Urteile fällen kann und dass man das Resultat der Überlegungen empirisch prüfen kann. Ob ein physikalisches Bild zutreffend ist oder nicht, entscheidet sich darin, dass die experimentelle Prüfung den theoretischen Resultaten im Rahmen der Beobachtungsgenauigkeit nicht widerspricht.

Die typischen Arbeitsschritte der physikalischen Methode bestehen aus dem Entwurf eines (oft nur von einigen wenigen Fakten ausgehenden) physikalischen Modells, seiner logischen Analyse und der empirischen Überprüfung der Konsequenzen. Dies soll nun am Beispiel der Bildentstehung beim Schattenwurfs und bei der Lochkamera veranschaulicht werden.

2.2. Schattenwurf und Lochkamera

Das geometrische Modell des SCHATTENWURFS (Abb. 14a) ist die strahlenoptische Projektion von einer zweidimensionalen Fläche auf eine andere. Im Strahlenmodell der geometrischen Optik dürfen sich Strahlen folgenlos in einem Punkt kreuzen. Daher spricht auch nichts dagegen, vom Modell einer punktförmigen Lichtquelle auszugehen. Man mag einwenden, dass es punktförmige Lichtquellen in der Realität nicht gibt, aber so lange etwas nicht durch die Postulate eines Modells verboten ist, darf man im Rahmen des erlaubten Modells „Gedankenexperimente“ durchführen, und das heißt hier, dass es nicht von vornherein unzulässig ist, die Konsequenzen einer punktförmigen Lichtquelle zu durchdenken. Der Tatsache, dass Lichtquellen erfahrungsgemäß ausgedehnt sind, kann man gerecht werden, indem man sie als dichte Ansammlung punktförmiger Lichtquellen auffasst. Es macht also keine prinzipiellen Schwierigkeiten, von theoretischen Resultaten für punktförmige Quellen auf solche endlicher Ausdehnung überzugehen.

Im Modell der punktförmigen Lichtquelle gehen alle Lichtstrahlen von der im Punkt L befindlichen Quelle aus und verbinden genau einen Gegenstandspunkt G der Objektebene mit genau einem auf einer Bildebene liegenden Bildpunkt B (14b). Eine Abbildung, bei der jeder durch den Bildpunkt B gehende Strahl von genau einem Gegenstandspunkt G herkommt, heißt STIGMATISCH (d.h. punktförmig). Die stigmatische Abbildung ist der Modellfall der absolut scharfen Abbildung. Die einander eineindeutig zugeordneten Punkte G und B werden als konjugierte Punkte bezeichnet. Das Wort „KONJUGIERT“ wird in der Physik verwendet, wenn man auszudrücken will, dass eine Beziehung zwischen zwei Objekten, Größen o.ä. besteht, die in einem gewissen Sinne von ihrer Funktion her auch umgekehrt gilt.⁵ Mit der Umkehrbarkeit ist hier folgendes gemeint: Wenn ein Punkt A in einem betrachteten Fall gerade als Gegenstandspunkt fungiert und in einen Bildpunkt B abgebildet wird, so kann der Punkt B genauso als Gegenstandspunkt fungieren und wird dann in den zu ihm konjugierten Punkt A abgebildet, der nun als Bildpunkt fungiert.

Für die mathematische Analyse ist es zweckmäßig, den Ursprung eines kartesischen Koordinatensystems in den Punkt L zu legen, und die z -Achse orthogonal zur Objektebene zu wählen. Die z -Achse durchstößt die Objektebene bei der Koordinate z_G und die Bildebene bei z_B . Bei der Abbildung wird ein Objektpunkt mit den Koordinaten (x, y, z_G) auf einen Bildpunkt mit den Koordinaten (x', y', z_B) abgebildet. Die den x - y -Koordinaten des Objektpunkts entsprechenden Bildkoordinaten sind hier durch einen Strich gekennzeichnet.

Sind Objekt- und Bildebene parallel zueinander, dann wird jede in der Objektebene liegende Gerade auf eine dazu parallele Gerade abgebildet. Da die von L ausgehenden Strahlen somit von parallelen Geraden geschnitten werden, kann der Strahlensatz angewandt werden. Jede orthogonal zur z -Achse liegende Strecke wird um denselben Skalenfaktor

$$V = \frac{z_B}{z_G} \quad (2.1)$$

⁵Beispiele: „komplexe Zahl“ und „konjugiert komplexe Zahl“, „Zustandsvariable und konjugierte Zustandsvariable“.

gestreckt, d.h. das Abbildungssystem ist bezüglich der z -Achse rotationssymmetrisch. Die rotationssymmetrische Achse eines optischen Systems, wird als OPTISCHE ACHSE bezeichnet. Insbesondere werden die x -Koordinaten um den gleichen Skalenfaktor V gestreckt, wie die y -Koordinaten. Die bijektive Zuordnung zwischen Gegenstands- und Bildpunkt kann als lineare Koordinatentransformation aufgefasst werden:

$$\begin{aligned}x' &= Vx \\y' &= Vy\end{aligned}\tag{2.2}$$

Eine beliebige Strecke $\ell = \sqrt{\Delta x^2 + \Delta y^2}$, die in einer Ebene normal zur optischen Achse liegt, wird daher auf eine Strecke

$$\ell' = \sqrt{\Delta x'^2 + \Delta y'^2} = V\ell$$

abgebildet, d.h. mit dem gleichen Faktor V skaliert. Folglich wird jede Figur in der Objektebene auf eine geometrisch ähnliche Figur in der Bildebene abgebildet. Nur unter dieser Voraussetzung liegt eine Ähnlichkeitstransformation vor, bei der alle Winkel erhalten bleiben. Wenn die Bild- und Objektebene gegeneinander geneigt sind, ist die Streckung beim Schattenwurf offensichtlich nicht mehr für alle Richtungen in der Bildebene gleich. Ein in der Objektebene liegender Kreis wird daher in der Bildebene auf eine Ellipse abgebildet.

Auch für Abbildungen mit einer Lochkamera (Abb. 15) ist leicht nachvollziehbar, dass strenge Ähnlichkeitsabbildungen nur mit einem rotationssymmetrischen optischen System möglich sind: Im Modell, das in Abb. 16 gezeigt ist, verläuft eine zur Objektebene orthogonale Gerade durch den Lochpunkt L. Die Bildkonstruktion ergibt nur dann ein dem Objekt ähnliches Bild, wenn auch die Bildebene orthogonal zu dieser Gerade ist. Dann ist das Abbildungssystem rotationssymmetrisch und diese Gerade die optische Achse. Vergleicht man Abb. 14b mit Abb. 16, so ist der Unterschied nur, dass z_G für die Lochkamera negativ ist. Während der Skalierungsfaktor $V = z_B/z_G$ beim Schattenwurf stets positiv ist (und das Bild somit aufrecht steht), ist es bei der Lochkamera „auf dem Kopf“, d.h. es erscheint als ein um 180° um die z -Achse gedrehtes Bild. Beim Schattenwurf ist stets $V > 1$. Daher entsteht immer ein vergrößertes Bild. Bei der Lochkamera kann das Bild hingegen sowohl größer als das Objekt ausfallen (für $z_B > |z_G|$), als auch kleiner (für $z_B < |z_G|$).

Das bis hier hin besprochene Basismodell kann für endliche Ausdehnung der Lichtquelle bzw. des Lochs entsprechend erweitert werden. Alle aus solchen geometrisch-optischen Modellen abgeleiteten Konsequenzen können empirisch auf die Zweckmäßigkeit des physikalischen Bildes geprüft werden. Sie haben sich in vielen Experimenten im Rahmen der Beobachtungsgenauigkeit als zutreffend erwiesen. Es soll hier nicht verschwiegen werden, dass bei genaueren Untersuchungen Abweichungen zu Tage treten, die im Rahmen der Theorie der Geometrischen Optik nicht verstanden werden können. Sie haben mit dem Phänomen der Beugung zu tun. Für nicht allzu anspruchsvolle experimentelle Beobachtungen ist das Modell der strahlenoptischen Projektion jedoch eine akzeptable erste Näherung.

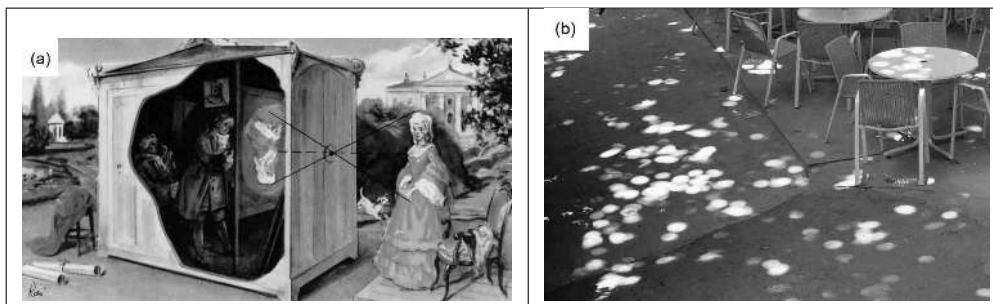


Abbildung 15: Lochkamera. (a) Camera Obscura. (b) Natürliche Lochkamera: Die kreisrunden Flecken (sog. Sonntaler) sind Bilder der Sonne, welche von (unregelmäßig geformten) Lücken zwischen den Blättern des Baums entworfen werden.

Abb. 15b zeigt sogenannte Sonntaler [5]. Die kleinen kreisrunden Scheiben kann man im Sommer unter wolkenlosem Himmel häufig bei Waldspaziergängen auf dem Boden sehen. Sie werden durch kleine Lücken zwischen den Blättern verursacht, die als „Löcher“ einer Lochkamera fungieren. Für die Abbildung ist die Form der „Löcher“ unerheblich. Sie können kreisförmig, viereckig oder sonst wie sein. Stets ergeben sich für die Projektionen der Sonne kreisförmige Scheiben. Bei einer Höhe der Löcher von ca. 5 m über dem Boden beobachtet man ungefähr 5 cm große Scheibendurchmesser. Daraus kann man schließen, dass der Abstand der Erde von der Sonne nur ca. 100 Sonnendurchmesser beträgt. Da die Sonne einen Abstand von 1 AE hat, lässt sich der Durchmesser der Sonne mit Hilfe des Strahlensatzes zu 1×10^{-2} AE abschätzen.

2.3. Planspiegel

SPIEGEL sind optische Elemente, deren Funktion auf der Reflexion an glatten Oberflächen beruht. Für Planspiegel ist die reflektierende Oberfläche eine Ebene. Für alle von einem Gegenstandspunkt G ausgehenden Strahlen lässt sich der Strahlverlauf mit dem Reflexionsgesetz (Gl. 1.6) konstruieren. Nach der Reflexion verlaufen sie so als ob sie von einem Punkt B herkämen, der sich aus der Rückverlängerung der reflektierten Strahlen ergibt (gestrichelte Linien in Abb. 17). B ist der scheinbare bzw. virtuelle Ausgangspunkt dieser Strahlen. Man bezeichnet ihn deshalb als „virtuell“, weil der Punkt B hinter der spiegelnden Ebene liegt und die Strahlen nicht tatsächlich von B ausgehen. Um B geometrisch zu konstruieren, genügt der Schnittpunkt zweier beliebiger Strahlen. Wenn es möglich ist, wählt man besonders einfache „Konstruktionsstrahlen“. Bei der Konstruktion kommt es nicht darauf an, ob entlang der gezeichneten Konstruktionslinien tatsächlich Licht verläuft oder nicht. Es ist eine Konstruktion im Rahmen einer geometrischen Analyse des Modells. Hier gibt es nur einen „einfachen“ Konstruktionsstrahl, nämlich den Strahl s_1 in Abb. 17a, der orthogonal auf den Planspiegel auftrifft (oder auftreffen würde). Er wird in sich zurückreflektiert, also mit $\alpha_r = 0$. In der Rückverlängerung verläuft er entlang s'_{1r} . Nun nimmt man noch einen beliebigen anderen Konstruktionsstrahl s_2 ,

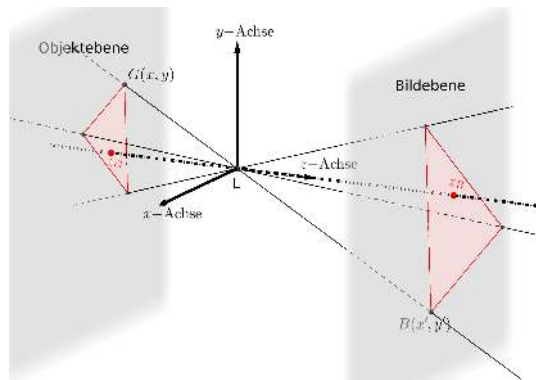


Abbildung 16: Modell der Abbildung einer Lochkamera. Der Gegenstandspunkt $G(x, y)$ wird von der Objektelebene auf den Bildpunkt $B(x', y')$ in der Bildebene abgebildet. Die gestrichelt gezeichnete optische Achse verläuft entlang der z -Achse, geht durch den Lochpunkt L und steht auf den beiden parallelen Ebenen orthogonal.

und damit ist der Schnittpunkt B bestimmt.

Im Unterschied zum Schattenwurf ist man frei, jede beliebige zur Spiegelebene orthogonale Achse als optische Achse heranzuziehen. Der Schnittpunkt der optischen Achse mit der Spiegelebene ist der Zentralpunkt Z . Die geometrische Konstruktion zeigt, dass der Skalierungsfaktor für die x - y -Ebene $V = 1$ ist, d.h. eine Spiegelung bildet die Umrisse eines ebenen Objekts identisch (und damit aufrecht) von der Objektelebene auf die Bildebene ab. Aber es gibt zwei wesentliche Unterschiede zwischen Schattenwurf und Spiegelung:

1. Bei der Spiegelung ist nicht nur ein einziger durch G gehender Strahl beteiligt, sondern alle: Jeder durch G gehende Strahlgerade entspricht eine durch B gehende Strahlgerade.
2. Die Abbildung ist für jede zur Spiegelebene parallele Ebene ähnlich, aber darüber hinaus auch noch in der dritten Dimension (nicht ähnlich, sondern) spiegelähnlich, weil der Skalierungsfaktor in der z -Richtung $V_z = -1$ ist.

Das dreidimensionale Abbild des Spiegels ist damit kein Bild (das wäre es, wenn $V_z = 1$ wäre), sondern ein SPIEGELBILD. Ein Spiegelbild kann man sich als Aufeinanderfolge zweier Abbildungsoperationen denken, nämlich einer Inversion und einer Drehung von 180° um die Flächennormale.

Eine Inversion bzw. Punktspiegelung hat einen invarianten Punkt, der in sich selbst abgebildet wird. Das ist der Zentralpunkt Z . Die Inversion ordnet jedem Objektpunkt G einen Punkt C derart zu, dass Z die Strecke \overline{GC} halbiert (s. Abb. 17b) Wegen der Inversion hat die Spiegelung eine Vertauschung der Händigkeit zur Folge (Abb. 17b). Im Gegensatz zu einem Bild, ist im Spiegelbild der Drehsinn vertauscht. Durch eine Spiegelung geht beispielsweise ein rechtshändiges Koordinatensystem in ein linkshändiges über

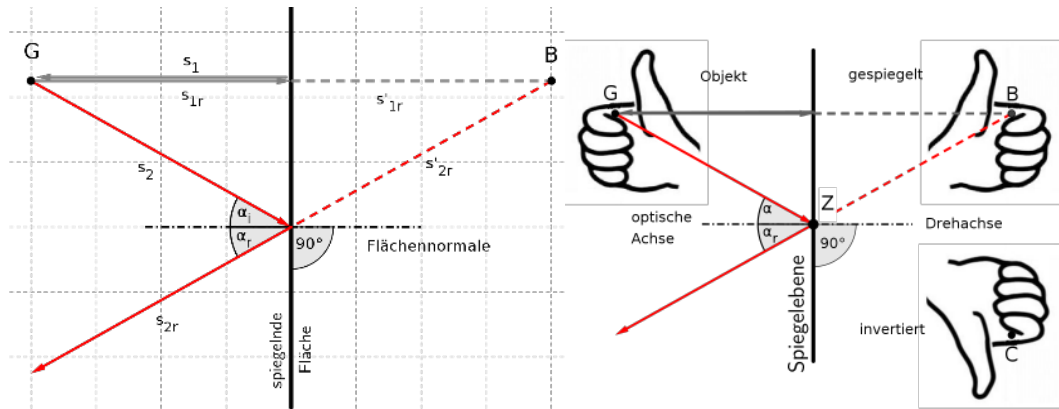


Abbildung 17: (a) Reflexion an einer spiegelnden Ebene (dicke Linie). Man konstruiert zuerst die Flächennormale (strichpunktiert) sowie den Verlauf zweier beliebiger Strahlen, die vom Gegenstandspunkt ausgehen. Der durch Rückverlängerung der Konstruktionsstrahlen (gestrichelt) sich ergebende Schnittpunkt ist der virtuelle Bildpunkt. (b) Die Konstruktion aller Bildpunkte hat eine zum Gegenstand entgegengesetzte Händigkeit und stellt ein Spiegelbild dar.

und eine rechtshändige Schrift in eine linkshändige (Abb. 27). Wenn man ein Spiegelbild wieder spiegelt, also zwei Spiegelungen durchführt, entsteht ein virtuelles dreidimensionales Bild im Maßstab 1:1, weil zwei Inversionen und zwei Drehungen um 180° wieder auf den Ausgangszustand zurückführen.

2.4. Theorie der perfekten strahlenoptischen Abbildung

Hier soll die Frage der Eigenschaften einer perfekten Abbildung durch Strahlen untersucht werden. Das ist eine ganz abstrakte Untersuchung, also eine mathematische. Welche Realitäten dem Modell des Strahls entsprechen, d.h. ob es sich um Licht-, Neutronen- oder Elektronenstrahlen handelt, ist dabei einerlei. Die einzig relevante Eigenschaft, die berücksichtigt werden muss, besteht darin, dass Strahlen durch Geraden modelliert werden, wie es der Situation entspricht, wenn der Objekt- und Bildraum der Abbildung aus homogenen Medien besteht.

Von einer idealen Abbildung fordert man, dass sie stigmatistisch ist, d.h. jeder Punkt des Objekts muss auf genau einen Punkt des Bildes abgebildet werden, und zwar so, dass alle Geraden, die durch einen Objektpunkt verlaufen, auf Geraden abgebildet werden, die durch den Bildpunkt verlaufen. Die eindeutig einander zugeordneten Objekt- und Bildpunkte stellen ein konjugiertes Punktepaar dar. Ihre Koordinaten sollen bezüglich des gleichen kartesischen Koordinatensystems angegeben sein. Um die Notation einfach zu halten, seien die Objektpunkte mit (x, y, z) bezeichnet und die Bildpunkte mit (x', y', z') , d.h. letztere sind mit einem zusätzlichen Strich gekennzeichnet. Das soll auch für andere Punkte gelten, d.h. alle Punkte des Objektraums sind ungestrichelt und alle Punkte, die

sich auf den Bildraum beziehen, gestrichen.

Es wird verlangt, dass das Bild treu ist, sich aber um einen Skalenfaktor V vom Objekt unterscheiden kann, d.h. die Abbildung soll eine geometrische Ähnlichkeitstransformation sein. Bildet man beispielsweise ein Dreieck ab, so soll das Ergebnis ein ähnliches Dreieck sein, d.h. die drei Winkel des Dreiecks sollen unverändert bleiben. Das ist äquivalent zur Forderung, dass der Skalenfaktor für jede Raumrichtung gleich und die Abbildung somit isotrop ist. Das abbildende optische System muss dann mindestens rotationssymmetrisch sein und es genügt, irgendeine Ebene zu betrachten, welche von der optischen Achse und einer dazu orthogonalen Richtung aufgespannt wird. Das soll hier die $x - z$ -Ebene sein, wobei die z -Achse des hier gewählten Koordinatensystems entlang der optischen Achse liege.

Kollineationen. Die allgemeinste Abbildung, die eine durch die lineare Beziehung

$$Ax + Bz + C = 0, \quad (2.3)$$

mit den Konstanten A, B, C gegebene Objektgerade auf eine Bildgerade

$$A'x' + B'y' + C' = 0, \quad (2.4)$$

mit i.A. anderen Konstanten A', B', C' abgebildet, hat die Form

$$\begin{aligned} x' &= \frac{a_{11}x + a_{13}z + b_1}{L_0(x, z)} \\ z' &= \frac{a_{31}x + a_{33}z + b_3}{L_0(x, z)} \end{aligned}$$

mit der linearen Funktion

$$L_0(x, z) = a_{01}x + a_{03}z + b_0 \quad (2.5)$$

und Konstanten a_{ij} und b_i , wobei $i, j = 0, 1, 3$. Spezialfälle dieser allgemeinen Transformation, die in der Mathematik als Kollineation bezeichnet wird, sind die linearen Transformationen mit $L_0 = 1$ und darunter die homogenen Transformationen, bei denen zusätzlich $b_1 = b_2 = b_3 = 0$ ist. Drehung und Inversion sind Spezialfälle der linearen homogenen Transformationen.

Gewöhnlich unterscheidet man Geraden, die sich schneiden von solchen, die zueinander parallel sind. In der projektiven Geometrie werden parallele Geraden jedoch als sich schneidende Geraden aufgefasst, deren Schnittpunkt im Unendlichen liegt. Für alle Punkte mit

$$L_0(x, z) \rightarrow 0$$

rückt nämlich der Bildpunkt ins Unendliche, d.h. die durch einen Punkt in dieser Ebene verlaufenden Geraden werden sich in einem Bildpunkt schneiden, der ins Unendliche rückt und des weiteren werden sie im Grenzübergang auf parallele Geraden abgebildet. Aus

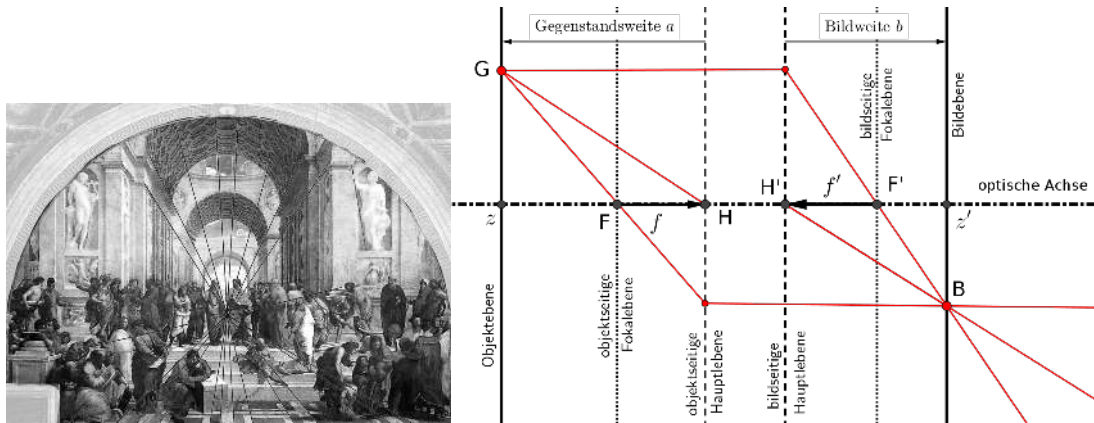


Abbildung 18: (a) Renaissance-Fresco „Die Schule von Athen“ von Raffael. Im Zentrum der zentralperspektivischen Darstellung stehen die antiken Philosophen Platon und Aristoteles. Auf dem Bild sind u.a. dargestellt: Archimedes (Physiker), Euklid (Mathematiker), Hypatia (Mathematikerin und Physikerin) und Ptolemäus (Physiker). (b) Projektive Abbildung: Alle von G ausgehenden Strahlen laufen wieder in B zusammen.

Gründen die später klar werden, bezeichnet man die Ebene $L_0(x, z) = 0$ als Fokusebene. Wie sich noch herausstellen wird, gehört zu jedem Punkt der Fokusebene im Bildraum eine Schar paralleler Geraden mit einer von der Position des Objektpunkts abhängigen Steigung. Die projektive Geometrie ist in der Malerei für die räumliche Darstellung von Bedeutung. In Abb. 18 erkennt man, dass bei der Zentralprojektion Geraden weiterhin Geraden bleiben, parallele Objektstrukturen im Unendlichen aber auf einen zentralen Punkt zusammenlaufen.

Um zu beweisen, dass die allgemeine Kollineation (und folglich auch jeder ihrer Spezialfälle) in der Tat Gl. 2.3 in Gl. 2.4 transformiert, multipliziert man Gl. 2.4 mit L_0 und erhält

$$A'L_1(x, z) + B'L_3(x, z) + C'L_0(x, z) = 0.$$

Durch Umformung lässt sich dieser Ausdruck auf die durch Gl. 2.3 geforderte Form bringen. Folglich ist das Bild einer Gerade wiederum eine Gerade.

Rotationssymmetrische Kollineationen. Kollineare Abbildungen sind in der Regel noch keine Ähnlichkeitsabbildungen. Das sind sie nur dann, wenn die Skalierung in jeder Richtung gleich groß ausfällt. Daraus folgt als Mindestforderung, dass die z -Achse rotationssymmetrisch ist. Die Ersetzung $x \rightarrow -x$ muss daher die Koordinate z' unverändert lassen und zu $x' \rightarrow -x'$ führen. Das geht nur, wenn $a_{31} = a_{01} = 0$ und $a_{13} = b_1 = 0$. Das reduziert die Transformation auf

$$\begin{aligned} x' &= \frac{a_{11}x}{a_{03}z + b_0} \\ z' &= \frac{a_{33}z + b_3}{a_{03}z + b_0} \end{aligned}$$

Indem man die vier Koeffizienten

$$\begin{aligned} f &= a_{11}/a_{03} & f' &= (b_3/a_{03} + FF')/f \\ F &= -b_0/a_{03} & F' &= a_{33}/a_{03} \end{aligned}$$

einführt, kann man die Koordinatentransformation in der symmetrischen Form

$$x' = f \frac{x}{z - F} \quad (2.6)$$

$$z' = F' + \frac{ff'}{z - F} \quad (2.7)$$

bzw.

$$x = f' \frac{x'}{z' - F'} \quad (2.8)$$

$$z = F + \frac{ff'}{z' - F'} \quad (2.9)$$

aufschreiben. Aus Gl. 2.9 folgt die NEWTONSCHE GLEICHUNG

$$(z - F)(z' - F') = ff'. \quad (2.10)$$

Mit ihr kann die Position z' eines Bildpunktes für eine gegebene Position z eines Objektpunktes berechnet werden und umgekehrt. Da diese Position nicht von y abhängt, werden alle in einer zur z -Achse orthogonalen Ebene (Objektebene) liegenden Objektpunkte auf eine zur z -Achse normalen Bildebene abgebildet. Der LATERALE SKALIERUNGSFAKTOR $V = \frac{\Delta x'}{\Delta x}$, wobei Δx die Differenz zweier Punkte auf der x -Achse ist, folgt direkt aus Gl. 2.6 bzw. Gl. 2.8:

$$V = \frac{\Delta x'}{\Delta x} = \frac{x'}{x} = \frac{f}{z - F} = \frac{z' - F'}{f'}, \quad (2.11)$$

Da die Skalierung gemäß Gl. 2.11 nicht von der x -Koordinate abhängt und daher auch nicht für jede andere Strecke, die orthogonal zur z -Achse liegt, wird jede Strecke in der x - y -Ebene um den gleichen Faktor V skaliert. Infolgedessen werden in einer zur z -Achse orthogonalen Objektebene liegende Dreiecke auf geometrisch ähnliche Dreiecke abgebildet. Für alle Punkte auf der Objektebene liegt daher eine Ähnlichkeitstransformation vor. Gilt das aber auch für die dritte Dimension entlang z ? Um diese Frage zu beantworten, berechnet man den LONGITUDINALEN SKALIERUNGSFAKTOR $V_l = \frac{\Delta z'}{\Delta z}$. Aus Gl. 2.11 erhält man:

$$V_l = \frac{\Delta z'}{\Delta z} = \frac{f'}{f} V^2, \quad (2.12)$$

Ist $|f'| = |f|$, so ist die gleiche Skalierung in allen drei Raumdimensionen nur dann möglich, wenn $V = \pm 1$, d.h. nur für Spiegelung, Inversion und Identität. Das stimmt mit

dem Befund für plane Spiegel überein (Kap. 2.3): Durch Reflektion entsteht ein dreidimensionales Spiegelbild im Maßstab 1 : 1. In optischen Anwendungen ist man jedoch an vergrößernden (Mikroskop) oder verkleinernden (Kamera) Ähnlichkeitsabbildungen interessiert. Diese sind wegen Gl. 2.12 nur für eine Abbildung von einer Ebene auf eine andere möglich, d.h. Abbildungen mit $V \neq \pm 1$ müssen sich zwangsläufig auf zweidimensionale Abbildungen beschränken. Mit einem Mikroskop kann man leicht selbst die Erfahrung machen, dass die Bildschärfe in z -Richtung (Tiefenschärfe) mit der Vergrößerung rasch abnimmt, weil V_l proportional zum Quadrat des lateralen Skalierungsfaktors V ist (Gl. 2.12). Bei sehr hoher Vergrößerung kann man daher praktisch nur auf eine Ebene „scharfstellen“.

Wenn man Punkte mit konstantem x betrachtet, dann bilden diese eine Gerade parallel zur optischen Achse. Gemäß Gl. 2.11 ist deren Bild die Gerade $x' = m'(z' - F')$ mit der Steigung $m' = x/f'$. Alle zur Achse parallele Geraden werden demnach auf Geraden abgebildet, die durch den Punkt F' gehen. Umgekehrt gehen alle Geraden, die mit einer Steigung $m = f/x'$ durch den Punkt F verlaufen, in Geraden über die parallel zur optischen Achse sind und die Koordinate y' haben. Den Punkt F bezeichnet man als objektseitigen BRENNPUNKT bzw. FOKALPUNKT und den Punkt F' als bildseitigen Brennpunkt bzw. Fokuspunkt. Parallelstrahlen und Fokalstrahlen schneiden sich in den beiden durch die konjugierten HAUPTPUNKTE $H = F + f$ und $H' = F' + f'$ verlaufenden HAUPTEBENEN, was man sich leicht durch eine Skizze klarmachen kann (Abb. 18b). In der Hauptebene liegende Objektpunkte werden im Maßstab 1:1 auf die konjugierte Hauptebene abgebildet (Gl. 2.11). Ein objektseitig durch H verlaufender Strahl verläuft bildseitig mit gleicher Steigung durch H' , d.h. solche Strahlen werden bei der Abbildung nicht abgelenkt. Falls H und H' zusammenfallen, bezeichnet man sie als MITTELPUNKTSTRAHLEN. Parallelstrahlen, Fokalstrahlen und ablenkungsfreie Strahlen genügen, um für eine mathematische Projektion zu jedem Objektpunkt den zugehörigen Bildpunkt zu konstruieren:

Abbildungssätze der projektiven Geometrie

1. Fokalstrahlen werden zu Parallelstrahlen.
2. Parallelstrahlen werden zu Fokalstrahlen.
3. Hauptstrahlen (Mittelpunktstrahlen) erfahren keine Ablenkung.

Je nach Position der Objektebene relativ zur Fokalebene kann man folgende Fälle unterscheiden:

Vergrößerung ($V > 1$)	für	$ z - F < f$
Identität ($V = 1$)	im Punkt	H
Inversion ($V = -1$)	im Punkt	H'
Verkleinerung ($V < 1$)	für	$ z - F > f$

Wenn man nun noch die positiv definierte GEGENSTANDSWEITE $a = H - z$ und die BILDWEITE für die (positive oder negative) Koordinate $b = z' - H'$ einführt, erhält man

aus dem Newtonschen Gesetz und Gl. 2.11 die **ABBILDUNGSGLEICHUNGEN**

$$\frac{1}{af'} - \frac{1}{bf} = \frac{1}{ff'} \quad (2.13)$$

$$\frac{x'}{x} = \frac{f}{f'} \frac{b}{a} \quad (2.14)$$

Diese Abbildungsgleichungen können als Referenz für den Vergleich mit der i.A. unvollkommenen Abbildung durch Spiegel, Linsen oder komplexeren optischen Systemen herangezogen werden. Sie lassen sich auch auf Abbildungen mit Elektronenmikroskopen, Linsen für Synchrotronstrahlung, Neutronenmikroskopen oder Gravitationslinsen anwenden, kurzum, auf jede Art von optischer Abbildung, bei der gerade Strahlen von einem Objektpunkt ausgehen und als gerade Strahlen in einem Bildpunkt enden. Es ist ein Wesenszug der Physik, dass sie die ganze Vielfalt der Welt auf nur wenige Grundprinzipien und Beziehungen zurückführen kann.

3. Sphärische optische Elemente

OPTISCHE ELEMENTE (z.B. Linsen) sind die Grundbausteine zu OPTISCHER KOMPONENTEN (z.B. Mikroskopobjektive). Optische Systeme bzw. Instrumente (z.B. Mikroskop) setzen sich wiederum aus i.A. mehreren optischen Komponenten zusammen. Grundlage für das geometrisch-optische Verständnis optischer Elemente sind die Gesetzmäßigkeiten für Reflexion und Brechung. Ist die Grenzfläche eines optischen Elements nicht mehr plan, müssen Reflexion und Brechung lokal konstruiert werden. Dazu benötigt man den Begriff der **TANGENTIALEBENE** (Abb. 19). Für jeden beliebigen Punkt P einer gekrümmten Grenzfläche ist dies die Ebene, welche für eine gewisse Umgebung eines Punkts P die Grenzfläche nur in ebendiesem Punkt berührt. Die **LOKALE FLÄCHENNORMALE** steht orthogonal bzw. normal zur Tangentialebene und geht durch den Punkt P. Die Gesetze der Reflexion und Brechung wendet man für diesen Punkt P so wie bisher an, aber bezüglich seiner lokalen Flächennormalen.

In diesem Kapitel werden optische Elemente mit sphärisch gekrümmten Grenzflächen besprochen (Kap. 3) und im nächsten daraus zusammengesetzte Systeme (Kap. 4). Wenn es um abbildungsoptische Anwendungen geht, interessiert erstens die Frage, unter welchen Umständen zumindest näherungsweise eine perfekte optische Abbildung realisiert werden kann, und zweitens, wie die charakteristischen Parameter H, H', f, f' dieser näherungsweise perfekten Abbildung mit den Geometrie- und Materialeigenschaften der Elemente oder Systeme zusammenhängen.

Beide nachfolgenden Kapitel beschränken sich auf eine geometrisch-optische Näherung erster Ordnung, was im Wesentlichen auf eine Näherung für kleine Winkel hinausläuft. Das ist für ein grundsätzliches Verständnis der abbildungsoptischen Funktion erst einmal ausreichend. Wenn man ein wirklich gutes Teleskop oder Mikroskop aus optischen Elementen aufbauen will, muss man selbstverständlich genauere Betrachtungen und Berechnungen anstellen, was hier nur als Ausblick angedeutet werden kann (s. Kap. 4.5).

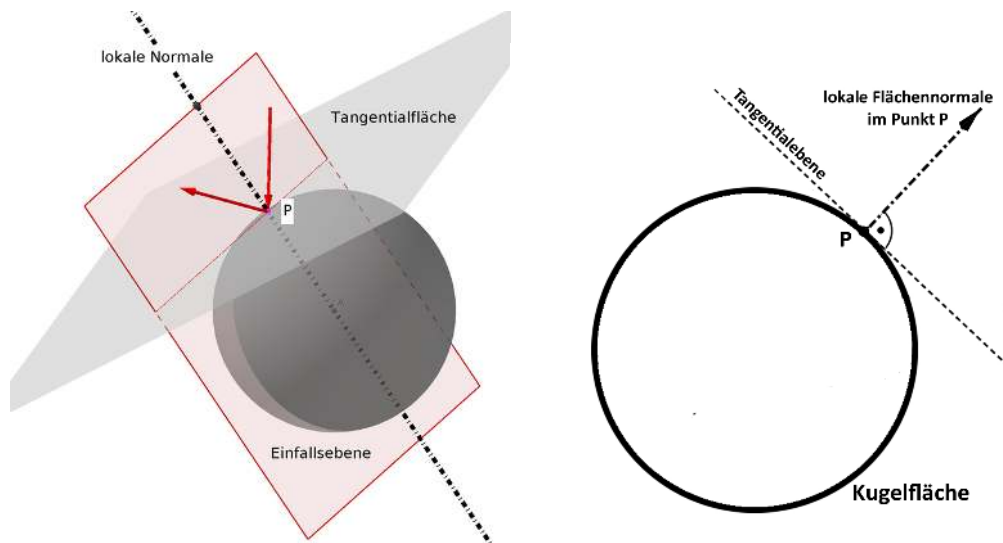


Abbildung 19: (a) Lokale Flächennormale und Tangentialebene im Punkt P einer gekrümmten Fläche. (b) Zweidimensionale Darstellung in der Einfallsebene.

3.1. Sphärische Spiegel

Sphärische Spiegel werden sowohl für Aufgaben der Abbildung als auch der Beleuchtung eingesetzt. Hier soll zunächst ihre Funktion als abbildungsoptisches Element betrachtet werden.

Die reflektierende Fläche eines sphärischen Spiegels liegt auf einer Kugel mit dem Mittelpunkt M und dem Radius r (Abb. 20). Fällt auf sie ein paralleles Strahlenbündel ein, so ist die durch M verlaufende Gerade die optische Achse bezüglich der Rotationssymmetrie vorliegt. Sie schneidet die reflektierende Fläche im Scheitelpunkt S . Zur Beschreibung wählt man ein Koordinatensystem mit S als Nullpunkt und mit der z -Achse entlang der optischen Achse, wobei ihre Orientierung in Richtung des einfallenden Strahls sei. Für einen KONKAVSPIEGEL (20a) ist die z -Koordinate von M negativ, für einen KONVEXSPIEGEL (20b) positiv. Im Vergleich hat ein Planspiegel den Vorteil, dass er perfekte dreidimensionale Spiegelbilder erzeugen kann, aber den Nachteil, dass das nur im Maßstab 1:1 möglich ist. Sphärische Spiegel haben den Vorteil, dass ein vergrößertes oder ein verkleinertes Spiegelbild verwirklicht werden kann, aber dafür muss der Vorzug des Planspiegels, dreidimensionale Spiegelbilder zu erzeugen, geopfert werden: Sphärische Spiegel können nur ebene Objekte in ebene Spiegelbilder abbilden. Ein weiterer Nachteil ist, dass die Abbildung nur im paraxialen Limes, d.h. im Limes für achsennahe Strahlen, perfekt wird. Die charakteristischen Parameter H, H', f, f' der perfekten Abbildung im paraxialen Limes sollen nun bestimmt werden.

Es ist sofort einsichtig, dass die Spiegelebene zugleich Gegenstands- und Bildebene für eine Abbildung im Maßstab 1:1 ist. Die Koordinate der Hauptebene ist für nur leicht gekrümmte Spiegel bzw. für achsennahe Strahlen daher identisch mit dem Scheitelpunkt

und die beiden Hauptpunkte fallen zusammen:

$$H = H' = S.$$

Als nächstes betrachtet man Parallelstrahlen. Sie unterscheiden sich durch ihre x -Koordinate. Von achsennahe bzw. paraxiale Strahlen spricht man, wenn $|x|$ sehr klein ist. Für einen im Punkt P auftreffenden Parallelstrahl erhält man den Schnittpunkt N mit der optischen Achse durch Anwendung des Reflexionsgesetzes. Für Konkavspiegel wird er tatsächlich vom Lichtstrahl durchquert und damit ist N ein reeller Punkt. Für einen Konvexspiegel handelt es sich um einen durch Rückverlängerung des Strahls konstruierten virtuellen Punkt, der hinter der spiegelnden Fläche liegt. Fällt man von N aus das Lot auf die durch M und P verlaufende Gerade, so ergibt sich Punkt L . Die Punkte L , M und N bilden ein rechtwinkliges Dreieck. Deren Ankathete hat die Länge $r/2$. Sie ist kleiner als die Hypotenuse, doch nähert sich die Länge der Hypotenuse mit kleiner werdendem Winkel α an die Länge der Ankathete an. Dabei läuft der Schnittpunkt N auf den Punkt F' zu. Alle Parallelstrahlen verlaufen daher näherungsweise durch diesen Punkt F' , wenn $|x|$ sehr klein wird. Im Limes $|x| \rightarrow 0$ liefern sphärische Spiegel also ein perfektes Spiegelbild mit bildseitigem Brennpunkt F' . Kehrt man den Strahlverlauf um, so ist leicht einzusehen, dass alle von F' ausgehenden Strahlen anschließend parallel sind. Daher fallen F und F' zusammen und es ist $f' = f$. Aus Gl. 2.13 folgen damit die Abbildungsgleichungen

$$\frac{x'}{x} = \frac{b}{a} \quad (3.1)$$

$$\frac{1}{a} - \frac{1}{b} = \frac{1}{f} \quad (3.2)$$

Hier ist a die stets positive z -Koordinate der Objektebene und b die positive oder negative Koordinate der Bildebene. Für einen Konkavspiegel ergibt sich die durch $f = H\text{-}F$ definierte Brennweite zu

$$f = f' = r/2 \quad (3.3)$$

und für einen Konvexspiegel zu

$$f = f' = -r/2. \quad (3.4)$$

Da f für Konvexspiegel negativ ist, ist b stets positiv. Das Spiegelbild eines Konvexspiegels ist daher stets virtuell, aufrecht und verkleinert.

Perfekt wird eine Abbildung mit sphärischen Spiegeln nur für den Grenzfall $r \rightarrow \infty$ des Planspiegels und für den paraxialen Grenzfall $|x| \rightarrow 0$. Für weiter von der Achse entfernte Parallelstrahlen oder zur Achse stärker geneigte Strahlen machen sich zunehmend Abbildungsfehler bemerkbar. Wenn diese Strahlen das Spiegelbild allzu störend beeinflussen, kann man sie beispielsweise durch Blenden beseitigen, was allerdings auf Kosten der Beleuchtungsstärke des Bildes geht.

Rotationssymmetrische Spiegel mit gekrümmten Oberflächen werden beispielsweise für Teleskope verwendet, insbesondere für Satellitenteleskope, die ferne Röntgenquellen er-

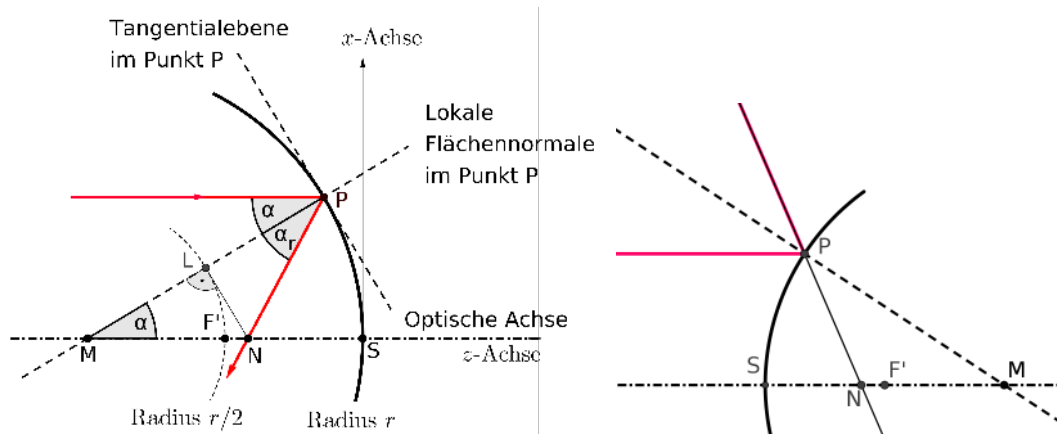


Abbildung 20: Skizze des Strahlengangs in der Einfallsebene (a) für einen sphärischen Konkavspiegel. (b) für einen sphärischen Konvexspiegel.

kunden.⁶ Es muss sich dabei nicht unbedingt um sphärische Spiegel handeln. Parabolspiegel bilden beispielsweise alle parallel zur optischen Achse verlaufenden Strahlen perfekt in den Brennpunkt ab, weisen aber u.U. größere Abbildungsfehler für Strahlen auf, die zur optischen Achse geneigt sind.

Abschließend soll noch kurz etwas über den Einsatz gekrümmter Spiegel für beleuchtungstechnische Anwendungen gesagt werden: Da Sonnenlicht fast parallel ist, kann man die Beleuchtungsstärke durch einen sphärischen Spiegel bzw. einen Parabolspiegel „aufkonzentrieren“. Die maximale Beleuchtungsstärke ergibt sich im Punkt F'. Ein dort befindlicher Körper, der das Licht absorbiert, kann sehr hohe Temperaturen erreichen. Das ist der Grund, warum man F bzw. F' als Brennpunkt bezeichnet. Umgekehrt kann man eine gleichmäßige Bestrahlungsstärke über eine große Fläche erreichen, wenn man eine isotrop abstrahlende Lichtquelle in den Brennpunkt platziert. Da sich die Divergenz des Strahls dabei verringert, nimmt die Beleuchtungsstärke nicht so schnell mit der Entfernung ab, was z.B. für Leuchttürme wichtig ist.

3.2. Linsen

Alles, was zur abbildungsoptischen Manipulation von Strahlung (Licht, Elektronen, Ionen usw.) dient und nicht auf Reflexion beruht, bezeichnet man als LINSEN. Soll die geometrische Ähnlichkeit bei der Abbildung erhalten bleiben, muss das abbildende System rotationssymmetrisch sein.

Abb. 21 zeigt eine SPHÄRISCHE LINSE. Sie grenzt durch zwei sphärische Flächen mit den Mittelpunkten M_1 bzw. M_2 und den Radien r_1 bzw. r_2 ein Medium B mit der relativen Brechzahl $n = n_B/n_A$ von einem umgebenden Medium A mit der relativen Brechzahl 1 ab. Die optische Achse (Symmetrieachse) verläuft durch die Mittelpunkte.

⁶Z.B. der vom Deutschen Zentrum für Luft- und Raumfahrt gebaute Röntgensatellit ROSAT, dessen Spiegel laut Guinness-Buch der Rekorde die glattesten Oberflächen der Welt waren.

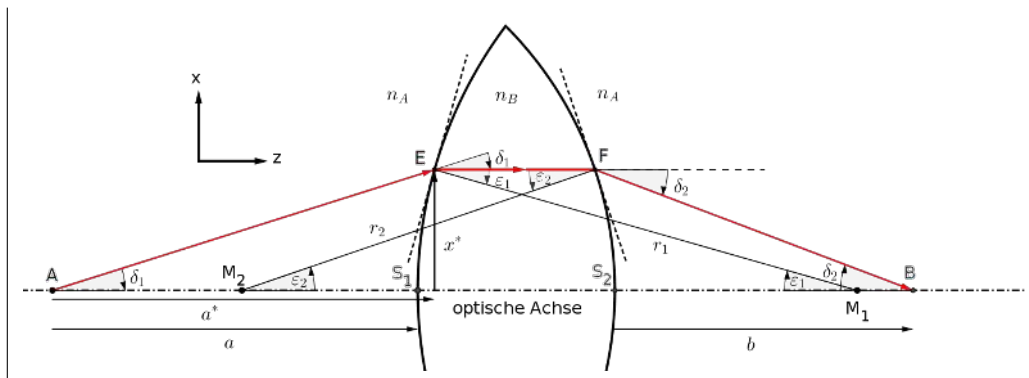


Abbildung 21: Skizze zur Herleitung der Linsenmacher-Gleichung.

Im Grenzfall unendlich großer Radien wird die Linse zur planparallelen Platte. Analog zum sphärischen Spiegel kann man daher schließen, dass der Scheitelpunkt S_1 dem Punkt H der abbildungsoptischen Hauptebene entspricht und der Scheitelpunkt S_2 dem Punkt H' . Zur Beschreibung wird ein kartesisches Koordinatensystem mit der z -Achse parallel zur optischen Achse eingeführt. Die mit großen Buchstaben bezeichneten Punkte, die auf der optischen Achse liegen, sollen zugleich für die z -Koordinaten stehen. Ferner seien die Größen $R_1 = S_1 - M_1$ und $R_2 = S_2 - M_2$ eingeführt. In Abb. 21 ist R_1 negativ und R_2 positiv. In diesem Fall handelt es sich um eine bikonvexe Linse. Bei einer bikonkaven Linse kehren sich die Vorzeichen von R_1 und R_2 um.

Im Inneren der Linse sei ein Strahl betrachtet, der mit der konstanten Koordinate x^* parallel zur optischen Achse verläuft. Das ist z.B. für den in Abb. 21 gezeigten Strahl von E nach F der Fall. An den Punkten E und F findet die Brechung genauso statt, wie bei einem von den beiden Tangentialflächen (gestrichelte Linien) gebildeten Prisma. Die Neigungswinkel der Flächennormalen relativ zur optischen Achse sind $\varepsilon_1 \approx x^*/R_1$ bzw. $\varepsilon_2 \approx x^*/R_2$. Die Gesamtablenkung ist $\delta = \delta_1 + \delta_2 = (1 - n)\varepsilon = (1 - n)(\varepsilon_2 - \varepsilon_1)$ bzw.

$$\delta = -D x^* \quad (3.5)$$

mit der BRECHKRAFT

$$D = (n - 1) \left[\frac{1}{R_1} - \frac{1}{R_2} \right]. \quad (3.6)$$

Bei Anwendung von Gl. 3.6 sind die Vorzeichen von R_1 und R_2 zu beachten. Für eine bikonvexe Linse ist beispielsweise $R_1 = r_1$ und $R_2 = -r_2$ und somit

$$D = (n - 1) \left[\frac{1}{r_1} + \frac{1}{r_2} \right]. \quad (3.7)$$

Für ein Medium mit $n > 1$ ist die Brechkraft einer bikonvexen Linse folglich positiv, diejenige einer bikonkaven Linse negativ. Gl. 3.6 wird als LINSENMACHER-GLEICHUNG bezeichnet. Mit der Brechkraft definiert sie die wichtigste Größe, die eine Linse charakterisiert. Die Einheit der Brechkraft

ist die DIOPTRIE

$$[D] = 1 \text{ Dioptrie} = 1 \text{ dpt} = 1 \text{ m}^{-1}.$$

Während der Betrag des Ablenkungswinkels für ein Prisma konstant ist, nimmt er für Linsen gemäß 3.5 proportional zum Abstand von der optischen Achse zu. Mit den Koordinaten $a = H - A$ und $b = B - H'$ relativ zu den Hauptpunkten H und H' gilt für kleine Winkel $\delta_1 \approx -x^*/a^* \approx -x^*/a$ und entsprechend $\delta_2 \approx -x^*/b$. Näherungsweise werden daher alle von einem Objektpunkt A ausgehenden Strahlen auf denselben Objektpunkt B abgebildet. Mit Gl. 3.5 folgt daraus die Abbildungsgleichung

$$\frac{1}{a} + \frac{1}{b} \approx D. \quad (3.8)$$

Sie entspricht der Abbildungsgleichung eines idealen Abbildungssystems, Gl. 2.13, wenn

$$f = \frac{1}{D} = -f', \quad (3.9)$$

d.h. wenn der objektseitige Brennpunkt F relativ zu U in negativer z -Richtung liegt und der bildseitige Brennpunkt F' in positiver Richtung.

Durch Vergleich mit Gl. 2.13 und Gl. 2.14 folgt für Linsen also die Abbildungsgleichung

$$\begin{aligned} V &\approx -\frac{b}{a} \\ \frac{1}{a} + \frac{1}{b} &\approx \frac{1}{f}. \end{aligned} \quad (3.10)$$

Linsen mit positiver Brennweite f stellen somit Sammellinsen dar, d.h. Parallelstrahlen laufen in einem bildseitigen Brennpunkt zusammen, während Linsen mit negativer Brennweite Zerstreuungslinsen sind. Ist $n > 1$, so fungieren bikonvexe Linsen als Sammellinsen und bikonkave als Zerstreuungslinsen. Für $n < 1$ ist es umgekehrt.

Zusammenfassend kann man feststellen, dass man mit Linsen näherungsweise die Bedingung für perfekte vergrößernde oder verkleinernde Abbildung erfüllen kann, wenn die auftretenden Winkel klein genug und die Strahlen im Inneren der Linse PARAXIAL verlaufen, d.h. nur kleine Winkel gegen die optische Achse haben.

4. Optische Instrumente

Die Lesebrille wurde im späten 13. Jahrhundert in den wirtschaftlich aufstrebenden Stadtrepubliken der Toskana erfunden. In Florenz (1321) und Pisa (1341) werden jene Universitäten gegründet, die das Wissen der arabischen Gelehrten für den abendländischen Raum rezipieren. Die Toskana wird zur Wiege der Renaissance, der Wiedergeburt der Antike, und knüpft mit der Wiederentdeckung des Platonismus wieder dort an, wo die Entwicklung der antiken Wissenschaft „im Westen“ abgebrochen ist. In Pisa und Florenz erhält der junge Student GALILEO GALILEI (1564-1641) jene Bildung, die ihm schon bald

ermöglichen wird, die Physik zu revolutionieren. Nicht zuletzt der hohe Stand der Brillenmacherkunst im Florenz der Renaissance ermöglichen Galileo Galilei den erfolgreichen Bau seiner optischen Instrumente.

Galilei gelangen mit dem Fernrohr sensationelle Beobachtungen. U.a. entdeckte er die vier großen Jupitermonde (Abb. 22a). Sie waren der schlagende Beweis, dass es etwas gab, das sich definitiv nicht um die Erde drehte und dass die „Heilige Schrift“ daher eine nachweisbare Falschaussage enthielt. Daran entzündete sich der Konflikt zwischen dem Wissenschaftler Galilei und der katholischen Kirche. Ein feierliches Begräbnis wurde nicht nur ihm (Abb. 22b), sondern auch seiner Tochter verweigert, obwohl letztere der Auflage nachkam, jede Woche sieben Psalmen zur Buße für ihren Vater zu beten. Das erscheint harsch, aber Theologen war klar, was Galileis Erkenntnis bedeutete: Wenn sich auch nur ein einziger Fehler objektiv nachweisen ließ, dann gab es kein Halten mehr. Dann war logischerweise zulässig, auch jeden anderen Satz der Heiligen Schrift anzuzweifeln. Es musste ein Weg gefunden werden zu entscheiden, welche Sätze richtig und welche falsch waren. Dieses durch die Physik aufgeworfene Problem wurde von der katholischen Kirche erst spät gelöst, nämlich durch das im ersten Vatikanischen Konzil aufgestellte Dogma von der Unfehlbarkeit des Papstes. Nachdem sich der Vatikan im Jahr 2008 von der Verurteilung Galileis durch die Inquisition distanzierte, war der Weg frei, in Rom die erste öffentliche Statue für den Vater der modernen Naturwissenschaften aufzustellen. Sie wurde durch den chinesisch-amerikanischen Nobelpreisträgers Tsung-Dao Lee entworfen und von der chinesischen Staatsregierung finanziert.

4.1. Theorie linearer optischer Systeme

Dieses Kapitel soll Sie in eine weitere typische physikalische Betrachtungsweise einführen. Bei dieser kommt es nicht darauf an, die Funktionsweise eines einzelnen Elements aus seinen Material- und Struktureigenschaften heraus zu verstehen, sondern das Funktionieren eines aus beliebigen Elementen aufgebauten Systems. Eine solche Betrachtungsweise wird beispielsweise eingesetzt, um die Funktion ganzer elektrischer oder elektronischer Netzwerke zu verstehen. Man nimmt dazu einen Systemstandpunkt ein. Im Fall optischer Systeme bedeutet das, dass man sich nur für den Zusammenhang zwischen der eingangsseitig eintretende Strahlgeometrie (Input) und resultierenden ausgangsseitigen Strahlgeometrie (Output) interessiert. Wie das System innen aufgebaut ist, ist dabei völlig egal. Man untersucht das System als „black box“, d.h. als eine obskure Schachtel unbekanntem Inhalts. Die vom System entworfenen Bilder heißen dabei REELL, wenn sie im Fall der Transmission hinter der ausgangsseitigen Grenzfläche entstehen und im Fall der Reflexion vor der Grenzfläche. In allen anderen Fällen sind die Bilder VIRTUELL.

Das Übertragungsverhalten des Modellsystems soll dabei durch einige wenige empirische Parameter und Regeln beschrieben werden, aus denen man das Gesamtverhalten für jede einfallende Lichtgeometrie in guter Näherung erschließen kann. Es geht in diesem Kapitel also um die Definition eines rein abstrakten mathematischen Modell-Spielzeugs und darum, den Umgang damit zu lernen.

Der Einfachheit halber wird das System als rotationssymmetrisch und linear vorausge-



Abbildung 22: (a) Physikstudenten der Universität Wien vor dem Gebäude der finnischen Botschaft am Vatikan. Von diesem Gebäude aus demonstrierte Galilei einigen Kardinälen, dass es Himmelsobjekte gab, die sich nicht um die Erde drehten. (b) Das Grabmal Galileo Galileis in der Kirche Santa Croce in Florenz durfte erst zweihundert Jahre nach seinem Tod errichtet werden.

setzt. Ersteres bedeutet, dass es eine optische Achse gibt. Linear ist das System dann, wenn alle Strahlablenkungen proportional zum Einfallswinkel sind. Ferner wird für die Konstruktion vorausgesetzt, dass alle Bezugsflächen Ebenen sind, die zur optischen Achse orthogonal stehen.

4.1.1. Systemcharakterisierung

Zur Charakterisierung linearer abbildungsoptischer Modellsysteme geht man vom Modell des idealen Abbildungssystems aus. Man zieht zum ersten das Übertragungsverhalten für parallele Strahlen in Betracht und zum zweiten Strahlen, welche bei der Übertragung von der Eingangs- zur Ausgangsseite keine Ablenkung erfahren. Das optische System gilt als vollständig charakterisiert, wenn man seine optische Achse, die Fokalfpunkte F und F' sowie seine Hauptpunkte $H = F + f$ und $H' = F' + f'$ relativ zu einem Bezugspunkt des optischen Systems (z.B. dem Scheitelpunkt der „hardwaremäßigen“ Eintrittsfläche) empirisch ermittelt hat. Das kann in folgenden Schritten geschehen:

1. Man lässt eingangsseitig paralleles Licht ein und bestimmt F' . Dann tut man dasselbe in umgekehrter Richtung und bestimmt F (23a)

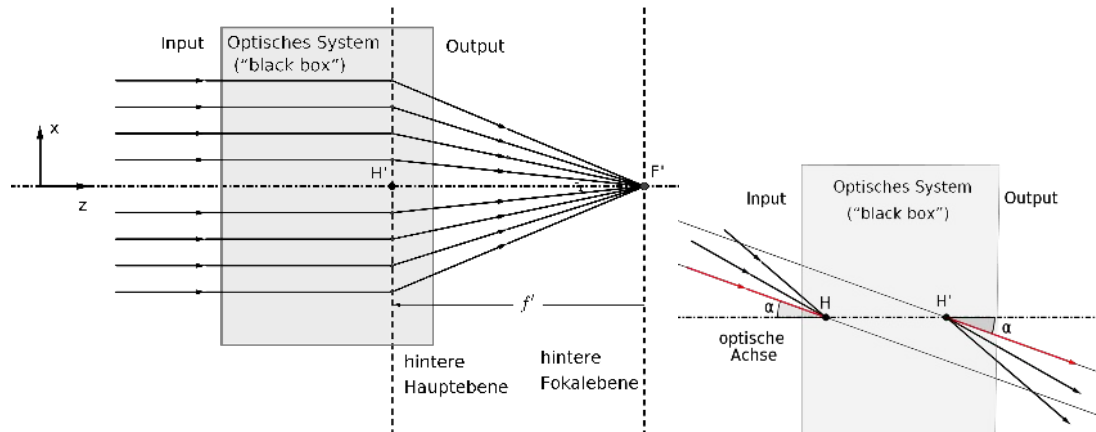


Abbildung 23: Charakteristika linearer optischer Systeme. (a) Parallelstrahlen werden zu Brennpunktstrahlen. Bei Annäherung an den Fokuspunkt F' erhöht sich die Beleuchtungsstärke. (b) Ablenkungsfreie Strahlen schneiden sich in den Hauptpunkten H und H' .

2. Man ermittelt die Hauptebenen, indem man z.B. die Lage der Punkte bestimmt, für die ein Strahl parallel zum Eingangsstrahl austritt (Abb. 23b).

Das Modellverhalten des optischen Systems ist damit dasjenige eines perfekten abbildenden Systems. Das Zusammenwirken von optischen Komponenten, die auf diese Weise charakterisiert wurden, kann durch entsprechendes Hintereinanderschalten von projektiven Abbildungen und gerader Lichtausbreitung zwischen den Komponenten beschrieben werden. Darauf soll hier nicht weiter eingegangen werden. Ist f' positiv, so konvergieren einfallende Parallelstrahlen auf der Ausgangsseite zum Brennpunkt F' hin (reeller Brennpunkt), ist sie negativ, so divergieren sie aus dem Brennpunkt heraus (virtueller Brennpunkt)..

In Kap. 3.1 und Kap. 3.2 wurde beispielsweise gezeigt, dass sich sphärische Spiegel und Linsen in der Näherung kleiner Winkel und paraxialer Strahlen wie lineare optische Systeme verhalten. Gl. 3.3 bzw. 3.4 sowie Gl. 3.9 verbinden dabei die Modellparameter f bzw. f' mit den konkreten Form- und Materialeigenschaften eines realen sphärischen Spiegels bzw. einer realen Linse.

4.1.2. Linsensysteme

Ein optisches System mit der Brennweite f bestehe aus zwei optischen Systemen mit den Brennweiten f_1 und f_2 . Der Abstand ihrer Hauptebenen sei $d = H_2 - H_1'$. Die ihnen

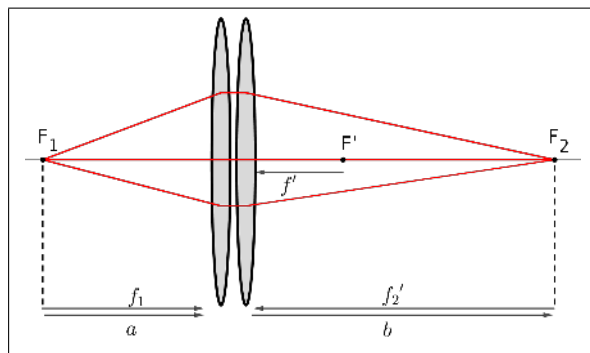


Abbildung 24: Konstruktion der Brennweite für ein aus zwei Linsen bestehendes Linsensystem.

äquivalenten Idealsysteme erfüllen die Abbildungsgleichungen

$$\begin{aligned}
 \frac{1}{a_1} + \frac{1}{b_1} &= \frac{1}{f_1} \\
 \frac{1}{a_2} + \frac{1}{b_2} &= \frac{1}{f_2} \\
 \frac{1}{a} + \frac{1}{b} &= \frac{1}{a_1} + \frac{1}{b_2} = \frac{1}{f}
 \end{aligned}
 \tag{4.1}$$

Da der Bildpunkt des ersten Systems gleich dem Objektpunkt des zweiten Systems ist, gilt für den Abstand der Hauptebenen $d = b_1 + a_2$ und somit folgt

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{a_1 b_2}.$$

Die Hauptebenen des Gesamtsystems sind hierbei $H = H_1$ und $H' = H'_2$, d.h. sie sind durch die eingangs- und ausgangsseitigen Hauptebenen gegeben.

In paraxialer Näherung, d.h. wenn die einfallenden und ausfallenden Strahlen näherungsweise parallel zur optischen Achse bleiben sollen ($a_1 \rightarrow \infty$, $b_2 \rightarrow \infty$), gilt $b_1 \approx f_1$ und $a_2 \approx f_2$ und somit

$$\frac{1}{f} \approx \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_1 f_2}.$$

Wenn man zwei Linsen mit den Brechkraften $D_1 = 1/f_1$ und $D_2 = 1/f_2$ direkt hintereinander anordnet (24), so dass der Abstand d der Hauptebenen gegen die Brennweiten vernachlässigt werden kann ($d \rightarrow 0$), gilt für die Brennweite $D = 1/f$ des zweilinsigen Systems näherungsweise

$$D \approx D_1 + D_2,
 \tag{4.2}$$

d.h. die Brechkraften addieren sich näherungsweise.

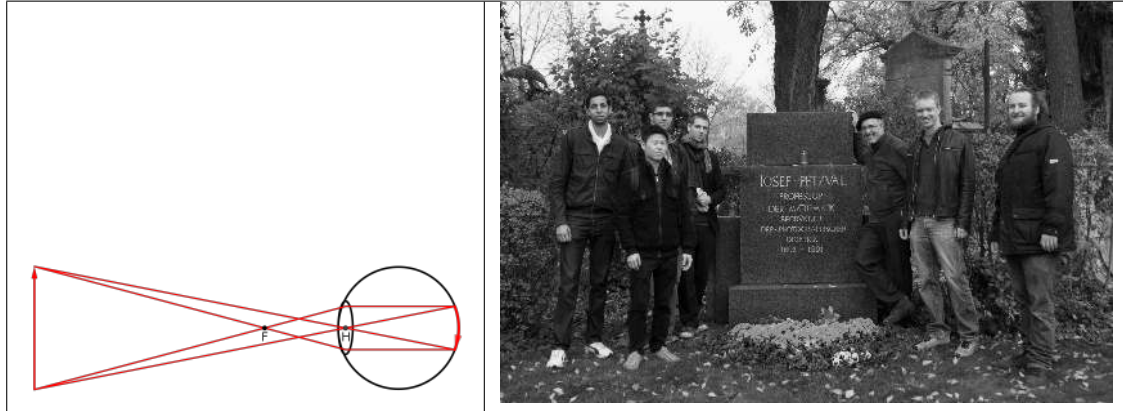


Abbildung 25: (a) Modellskizze zur Abbildung des Auges. (b) Physikerkursion zum Ehrengrab Petzvals auf dem Wiener Zentralfriedhof.

Man kann dieses Resultat auch direkt aus Gl. 4.1 ableiten, wenn man $a = a_1 = f_1$ setzt, weil dann die Strahlen hinter der ersten Linse parallel zur optischen Achse sind und von der zweiten Linse in den Brennpunkt der zweiten Linse abgebildet werden, wo $b = b_2 = -f'_2 = f_2$ ist (24). Die bildseitigen Hauptebenen der zweiten Linse und des Gesamtsystems bleiben dabei nur dann (näherungsweise) gleich, wenn $d = 0$, d.h. wenn die beiden Linsen direkt hintereinander angeordnet werden.

4.2. Auge

Das menschliche Auge ist ein komplexes brechendes optisches System. Der Einfachheit halber wird es als „Augenlinse“ bezeichnet und hier auf ein einfaches lineares Ersatzmodell im Sinne von Kap. 4.1 mit den Abbildungsgleichungen

$$V = \frac{x'}{x} \approx -\frac{b_A}{a} \quad (4.3)$$

$$\frac{1}{a} + \frac{1}{b_A} \approx D_A \quad (4.4)$$

reduziert. Hier ist D_A die nominale Brechkraft und b_A die nominale Bildweite des Auges. Das Bild entsteht auf der Netzhaut bzw. Retina (25). Die Fähigkeiten des Auges, die Details eines Objekts auflösen zu können, wird durch die Dichte der Sehzellen auf der lichtsensitiven Netzhaut eingeschränkt. Sie entsprechen den „Pixeln“ bzw. Bildpunkten des Bildes. Je näher man ein Objekt an das Auge heranrückt, desto größer fällt die Projektion des Objekts auf die Retina aus (Gl. 4.3). Doch unterhalb von ca. $a \approx 10$ cm kann das Auge die Bedingungsgleichung 4.4 für eine scharfe Abbildung nicht mehr erfüllen. Bei diesem kleinstmöglichen Sehabstand von 10 cm für scharfe Abbildung kann man Strukturen von etwa einem Zehntel Millimeter gerade noch auflösen. Daraus lässt sich mit Gl. 4.3 abschätzen, dass der Abstand x'_{min} der Sehzellen größenordnungsmäßig bei $x'_{min} \approx 10 \mu\text{m}$ liegt. Wenn man die lichtsensitiven Netzhautfläche mit ca. 1 cm^2 abschätzt,

liegt die Anzahl der Fotorezeptoren des Auges in der Größenordnung von einem Megapixel ($1 \text{ cm}^2/100 \mu\text{m}^2 = 10^6$). Aufgrund des beschränkten Pixelabstands y'_{min} kann man keine Strukturen mehr von einem Punkt unterscheiden, die kleiner sind als $x_{min} = (x'_{min}/b_A) a$. Wegen der begrenzten Brechkraft D_A des Auges bzw. seiner begrenzten Akkommodationsfähigkeit schränkt das sowohl die Fähigkeit ein, Strukturen weit entfernter Gegenstände als auch feine Strukturen im Nahbereich auflösen zu können.

Da die Bildweite b des Auges des Auges unveränderlich ist, muss Gl. 4.4 für einen gegebenen Objektstand a durch Anpassung von D_A erfüllt werden. Die reziproke Bildweite des Auges liegt bei ungefähr $1/b_A \approx 60 \text{ dpt}$ und entspricht einer Bildweite b_A , die etwas kleiner als 2 cm ist. Für sehr weit entfernte Objekte ist die reziproke Objektentfernung $1/a$ gegen die reziproke Bildweite vernachlässigbar und daher ist die vom Auge aufzubringende Mindestbrechkraft $D_{A<}$ für die Abbildung ferner Objekte $D_{A<} \approx 60 \text{ dpt}$. Rückt das Objekt näher an das Auge heran, muss sich die Brechkraft um $1/a$ erhöhen. In gewissem Ausmaß hat das Auge diese Fähigkeit der Änderung der Brechkraft. Wenn Objekte jedoch einen Sehabstand von ca. 10 cm unterschreiten, werden sie unscharf, d.h. das Auge kann für $1/a > 10 \text{ dpt}$ nicht mehr weiter akkommodieren.⁷ Durch Akkommodation kann das Auge also seine Brechkraft in den Grenzen

$$D_{A<} \lesssim D_A \gtrsim D_{A>}$$

mit der minimalen Brechkraft $D_{A<} = 60 \text{ dpt}$ und der maximalen Brechkraft $D_{A>} \approx 70 \text{ dpt}$ variieren.

4.3. Brille und Lupe

Wenn man eine Linse der Brechkraft D vor das Auge setzt, beträgt die effektive Brechkraft des Gesamtsystem gemäß Gl. 4.2

$$D_{eff} = D + D_A$$

Wenn eine Person fersichtig ist, aber im Nahbereich wegen fehlender Brechkraft nicht mehr akkommodieren kann, lässt sich das durch eine Brille mit positiver Brechkraft, d.h. mit einer Sammellinse, ausgleichen. Wenn Kurzsichtigkeit vorliegt, weil die Mindestbrechkraft des Auges zu hoch ist, kann das durch eine Linse mit negativer Brechkraft (Zerstreuungslinse) ausgeglichen werden. Für diese Korrekturen genügen Brillengläser von nur wenigen Dioptrien.

Verwendet man Sammellinsen sehr hoher Brechkraft, die man vor das Auge hält, so kann das Objekt weit näher an das Auge herangeführt werden als 10 cm . Indem man a

⁷Das ist selbstverständlich von Mensch zu Mensch unterschiedlich. Kurzsichtige Menschen können z.T. noch erheblich unterhalb von 10 cm scharf sehen und folglich noch feinere Strukturen auflösen. Vor Erfindung der Lupe wurden daher sehr feine Kunst- und Juwelierarbeiten von kurzsichtigen Handwerkern ausgeführt.

verringert, erhöht sich gemäß Gl. 4.3 die Bildgröße. Der arabische Wissenschaftler Al-Haitam erkannte und beschrieb erstmals diesen Vergrößerungseffekt von Sammellinsen und wurde damit zum Erfinder der LUPE.

Die Vergrößerung der Lupe definiert man so, dass man die Abbildung des unbewaffneten Auges für eine Normentfernung $a_0 = 25 \text{ cm}$ mit derjenigen vergleicht, welche sich bei gleicher Brechkraft des Auges mit einer Lupe der Brechkraft $D = 1/f$ ergibt. Die Abbildungsgleichungen sind:

$$\frac{1}{a_0} + \frac{1}{b_A} = D_A \quad (4.5)$$

$$\frac{1}{a_L} + \frac{1}{b_A} = D_A + D \quad (4.6)$$

Im ersten Fall ist das Abbildungsverhältnis $V_0 = -b_A/a_0$ und im zweiten Fall $V_L = -b_A/a_L$. Daraus ergibt sich die Lupenvergrößerung

$$m_L = \frac{V_L}{V_0} = \frac{a_0}{a_L} = 1 + a_0 D = 1 + \frac{25 \text{ cm}}{f}.$$

Die Normentfernung $a_0 = 25 \text{ cm}$ ist eine Konvention und wurde so gewählt, dass das Objekt in beiden zu vergleichenden Fällen ungefähr mit gleich geringer Anstrengung beobachtet werden kann.

4.4. Fernrohr und Mikroskop

In diesem Abschnitt wird exemplarisch die Funktion von optischen Instrumenten besprochen, die sich primär aus zwei optischen Einzelkomponenten zusammensetzen, die man gemäß ihrer Funktion im Abbildungssystem als OBJEKTIV und OKULAR bezeichnet.

Objektiv. Während es mit der Lupe eine Lösung für die Abbildung von Objekten mit höherer Auflösung gibt, die man nahe genug an das Auge heranbringen kann, gibt es keine analoge Lösung für ferne Objekte. Man scheitert nämlich nicht an Gl. 4.4 und damit an der Brechkraft, sondern an Gl. 4.3:

$$x' \approx -\frac{x}{a} b_A \geq x'_{min}.$$

Wenn a immer größer wird, so wird das Bild x' eines Objekts der Größe x irgendwann kleiner als der Abstand x'_{min} zweier Sehzellen. Wie kann man das Auflösungsvermögen steigern? Man könnte daran denken, b_A zu vergrößern, aber dazu bräuchte man größere Köpfe bzw. verlängerte Augäpfel. Das hat die Natur nicht vorgesehen. Man kann diesen grundlegenden Gedanken aber mit „künstliche Augen“ realisieren, bei denen die Bildweite $b > b_A$ größer ist als die des Auges. Bei dem in Abb. 26a gezeigten TELESKOP bildet das Objektiv (rechte Linse) auf ein Bild der Größe

$$x'_T = -\frac{x}{a} b$$

ab, das in einer Ebene liegt, die sich unmittelbar vor dem Okular (linke Linse) befindet. Der Abstand von Objektiv zu dieser Ebene ist die Bildweite b . Grob gesprochen kann man daher sagen, dass das Auflösungsvermögen des Teleskops um so größer ausfällt, desto länger es ist.

Die grundlegende Idee, die gerade für das Teleskops erläutert wurde, kann man auch auf die Betrachtung kleiner Objekte im Nahbereich anwenden, indem man die Abbildung ebenfalls mit einem Objektiv als „künstlichem Auge“ durchführt und dabei b sehr viel größer als b_A macht. Das geschieht in dem in Abb. 26b gezeigten MIKROSKOP. Sein Vorteil gegenüber der Lupe ist, dass die Brechkraft des Objektivs für die gleiche Vergrößerung kleiner bleiben kann als die zusammengesetzte Brechkraft des aus Lupe und Auge bestehenden Systems. Dadurch können die Linsenfehler der Abbildung, d.h. die Abweichungen von einer perfekten Abbildung, dramatisch verringert werden. Auf das Objektiv des in Abb. 26b gezeigten Objektivs folgt eine APERTURBLENDE. Sie hat den Zweck, Strahlen auszugrenzen, die zu Abbildungsfehlern führen würden. Die Funktion der Feldlinse ist die Vergrößerung des Sichtfeldes.

Okular. Sowohl beim Teleskop als auch beim Mikroskop stellt sich das Problem, wie man das vergrößerte Bild betrachten soll, denn das Auge besteht ja nicht nur aus der Retina, die man einfach nur in die Bildebene platzieren müsste, sondern hat selbst ein vorgeschaltetes Abbildungssystem. Man muss daher ähnlich wie mit der Brille Anpassungen für diese „Augenlinse“ vornehmen. Dazu gibt es prinzipiell zwei Möglichkeiten: Erstens kann man die „Augenlinse“ mit einer Zerstreuungslinse schlicht (teilweise) „wegkompensieren“ (Kompensationsokular). Im Vergleich zu einer Brille für Kurzsichtige, muss die dafür eingesetzten Zerstreuungslinsen eine höhere negative Brechkraft haben. Die zweite Möglichkeit besteht darin, dieses als „Zwischenbild“ bezeichnete Bild mit Hilfe einer Lupe in Lupenentfernung (d.h. in der Entfernung der Brennweite) zu betrachten (Lupenokular). Letzteres ist für Mikroskope die optimale Variante, weil die Gesamtvergrößerung des Mikroskops das Produkt der Objektivvergrößerung und der Lupenvergrößerung ist.

Fernrohre mit einem Okular nach dem Lupenprinzip fallen etwas länger aus. Ferner steht das Bild auf dem Kopf. Beides ist für astronomische Fernrohre nicht von Belang, wohl aber die höhere Vergrößerung durch das Lupenokular.

Würde man ein Kompensationsokular mit einer Zerstreuungslinse von -60 dpt realisieren, also das optische System des Auges vollständig wegkompensieren, so hätte man praktisch den Fall eines Auges mit einem sehr langen Augapfel realisiert. Daher ist klar, dass hier das Bild nicht für den Beobachter nicht auf dem Kopf zu stehen scheint. Dieser Vorteil und die Tatsache, dass die Fernrohre etwas kürzer ausfallen, macht Kompensationsokulare für terrestrische Fernrohre vorteilhafter.

Blenden und Feldlinsen. In 26 treten mehrere Blenden sowie eine Feldlinse auf. Die Aperturblende hat den Zweck, Strahlen auszugrenzen, die zu Abbildungsfehlern führen würden. Die Feldlinse steht meist an der Position des Zwischenbildes und dient der Vergrößerung des Sichtfeldes.

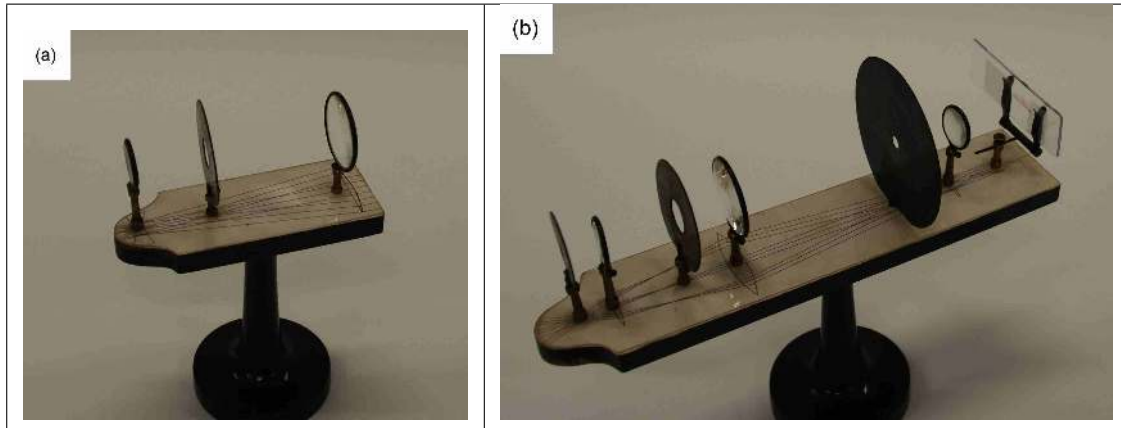


Abbildung 26: (a) Terrestrisches Fernrohr bestehend aus einer Sammellinse als Objektiv, einer Aperturblinse und einer Zerstreulinse als Okular. (b) Mikroskop mit Objekthalter, Objektiv, Aperturblinse, Feldlinse, Feldblende und Lupe. Das Zwischenbild entsteht in der Ebene der Feldblende.

4.5. Petzvals Revolution der Geometrischen Optik

Obwohl es Ende des 17. Jahrhunderts im Prinzip möglich war, optische Komponenten genauer zu berechnen und damit zu optimieren, bestand zunächst kein Bedarf, denn die Technologie der Linsenherstellung steckte noch in den Kinderschuhen. Mitte des 19. Jahrhunderts war die Perfektion der Linsenherstellung so weit gediehen, dass es nun auch Sinn machte, über die Korrektur von Linsenfehlern genauer nachzudenken. Schon aufgrund der einschränkenden Voraussetzungen bei der Herleitung der Linsenmachergleichung in Kap.3.2 sind zwei Linsenfehler offensichtlich:

1. Die SPHÄRISCHE ABERRATION. Sie tritt auf, weil der Schnittpunkt von Parallelstrahlen um so weiter vom Brennpunkt abweicht, je weiter diese von der optischen Achse entfernt sind.
2. Der ASTIGMATISMUS. Der Fehler tritt in dem Maße in Erscheinung, wie die Situation der Abbildung von der perfekten Rotationssymmetrie abweicht. Das kann daran liegen, dass die Achse eines Strahlenbüschels nicht parallel zur optischen Achse ist oder daran, dass das optische System nicht rotationssymmetrisch ist, weil der Radius der Krümmung in der x -Ebene ein anderer ist als für die y -Ebene. Der Extremfall einer astigmatischen Linse ist die Zylinderlinse. Astigmatische Linsenfehler eines optischen Systems (z.B. des Auges) können beispielsweise durch Kombination mit einer entsprechend orientierten Zylinderlinse korrigiert werden.

Die bahnbrechenden Arbeiten zur Korrektur von Linsenfehlern wurden unter der Leitung von Jozef Maximilian Petzval durchgeführt (25b). Sie führten gegen Ende des 19. Jahrhunderts zu einer Revolution in der Optik. Für seine präzisen Berechnungen konnte er sich auf eine militärische Computereinheit mit einer Rechenleistung von 13 Artillerieoffizieren stützen, die ihm vom österreichischen Erzherzog Ludwig zur Verfügung gestellt wurde.

Auf Grund ihrer Ausbildung waren Artillerieoffiziere für Routineberechnungen besonders qualifiziert und wurden damals als Rechner (im englischen Sprachraum auch als „Computer“ bezeichnet, also als Menschen, welche die Aufgabe des „computing“ erledigten, also des Berechnens.) für aufwändige Rechenaufgaben verwendet. Auf der Grundlage seiner Rechnungen konnten beispielsweise jene lichtstarken Objektive verwirklicht werden, ohne die der Siegeszug der Photographie undenkbar gewesen wäre. Das Verständnis der Ursachen der Linsenfehler und die darauf beruhende Verbesserung der Mikroskopobjektive durch Petzval und Ernst Abbe trugen wesentlich zur Revolution in der Medizin gegen Ende des 19. Jahrhunderts bei. Durch das Aufkommen leistungsstarker elektronischer Computer erhielt die „Computational Physics“ Mitte des 20. Jahrhundert einen dramatischen Entwicklungsschub. Mit Hilfe von Ray-Tracing-Programmen können seither auch die komplexesten Linsensysteme mit einer bis dahin unerreichten Genauigkeit berechnet werden.

Teil II.

Kinematik

Die KINEMATIK ist dasjenige Teilgebiet der Mechanik, das sich ausschließlich mit der Beschreibung von Bewegungen befasst. Ausgangspunkt ist die Klärung der Begriffe Raum, Zeit, Geschwindigkeit und Beschleunigung. Das Galileische Relativitätspostulat führt zu einer Vereinigung der Theorien von Raum und Zeit. In dieser Vereinigungstheorie, der Theorie der Raumzeit, spielt die Strukturkonstante der Raumzeit, $c = 299792458\text{m/s}$, eine zentrale Rolle. Um die Darstellung möglichst einfach zu halten, werden in diesem Teil erst einmal nur räumlich eindimensionale Bewegungen betrachtet. Die Erweiterung der Mechanik auf den dreidimensionalen Raum und die Begründung für die Wahl eines euklidischen Raumes erfolgt erst in Teil [V](#).

5. Raum

Die Geometrische Optik beruht einerseits auf einem Modell des Raumes (euklidischer Raum) und liefert andererseits dem Experimentator die Werkzeuge, um den Raum zu erkunden. In diesem Kapitel soll ein Abriss der sich daraus ergebenden Theorie des Raumes gegeben werden. Die grundlegende Hypothese über den physikalischen Raum lautet:

Der Raum ist homogen und isotrop.

Die erste Eigenschaft (Homogenität) ist eine eher „religiöse“ Glaubensaussage, während sich die zweite (Isotropie) wenigstens lokal prüfen lässt.

5.1. Homogenität

Physiker sind der Überzeugung, dass es allgemeine physikalische Gesetze gibt, die überall gelten. Sie drücken das durch die Zuschreibung aus, dass der Raum HOMOGEN sei. Mit dieser grundlegenden Arbeitshypothese drückt man aus, dass man fest daran glaubt, dass es physikalische Gesetze und Naturkonstanten gibt, die an jedem Ort der ungeheuren Weiten des Kosmos genau so gelten, wie im Heimatlabor. Das ist insofern „Religion“, denn es gibt keine Möglichkeit, die Zuschreibung der Homogenität zu prüfen. Man kann nicht vor Ort überprüfen, ob die als allgemeingültig erachteten Gesetze auf den Planeten entfernter Galaxien tatsächlich exakt so gelten wie bei uns. Aber man geht davon aus, dass das physikalische Geschehen in einer Welt, die relativ zu unserer um Milliarden Parsec ortsverschoben wäre, den gleichen Gesetzen unterliegt, wie in unserer Welt.

5.2. Isotropie

Neben der Homogenität schreibt man dem physikalischen Raum zu, ISOTROP zu sein. Damit meint man, dass die Physik auch in einer Welt, die relativ zu unserer gedreht wäre, nach den gleichen Gesetzen abläuft. Galileo Galilei vermittelt in seinem „Dialogo“ [3] die Erfahrung der Isotropie auf eine sehr anschauliche Weise:

„Schließt euch in einen möglichst großen Raum unter dem Deck eines großen Schiffes ein. Verschafft euch dort Mücken, Schmetterlinge und ähnliches fliegendes Getier, sorgt auch für ein Gefäß mit Wasser und kleinen Fischen darin. Beobachtet nun sorgfältig wie die fliegenden Tierchen nach allen Seiten des Zimmers fliegen. Auch die Fische werden ohne irgendwelche Umstände gleichermaßen nach allen Richtungen schwimmen.“

Die Hypothese der Isotropie ist lokal prüfbar, und wenn eine der vielen Beobachtungen, die durch die letzten Jahrhunderte gemacht wurden, eine Richtungsabhängigkeit des physikalischen Geschehens ergeben hätten, dann hätte das die Hypothese der Isotropie des Raumes falsifiziert. Aber bislang hat sie alle Prüfungen unbeschadet überstanden. Wenn man die Hypothese der Isotropie jedoch irgendwo als richtig akzeptiert, dann gilt sie wegen der Annahme der Homogenität des Raumes überall.

5.3. Bezugssysteme

In Kap. 1.6 wurden bereits Bezugselemente definiert (Winkel relativ zur Flächennormale, Ablenkung relativ zur Richtung des einfallenden Strahls) und Konventionen über Winkelangaben eingeführt. Das geht auch gar nicht anders, denn der Raum ist nach Voraussetzung isotrop. Richtungsangaben sind nicht absolut, sondern relativ, und daher wird sie ein Zugbegleiter nicht einfach zum Ausstieg nach links auffordern, sondern zum „Ausstieg in Fahrtrichtung links“. Im Raum gibt es keine Vorzugsorientierungen. Daher sind Konventionen über Richtungen, Winkel und Orientierungen erforderlich. Wegen der Homogenität des Raumes erfordert auch die Kommunikation, wo etwas ist, eine Konvention über einen als Referenz dienenden Raumpunkt. Jede Positionsangabe stellt eine Angabe RELATIV zu einem räumlichen Referenz- oder Bezugspunkt dar.

Ein aus mehreren Bezugselementen bestehendes System bildet ein BEZUGSSYSTEM. In einem eindimensionalen Raum käme man mit zwei Elementen aus: einen Bezugspunkt und einer Bezugsrichtung. Für einen Raum höherer Dimension werden weitere Bezugselemente benötigt. Wenn unterschiedliche Konventionen über Bezugselemente vorliegen, dann benötigt man einen Übersetzungsschlüssel, mit dem man eine Angabe relativ zu einem Bezugssystem auf ein anderes Bezugssystem übertragen bzw. transformieren kann.

5.4. Koordinatensysteme

Um quantitative Positionsangaben machen zu können, benötigt man außer der Festlegung eines Bezugssystems noch eine Vereinbarung über ein KOORDINATENSYSTEM. Dasselbe Koordinatensystem kann unterschiedlichen Bezugssystemen und demselben Bezugssystem können unterschiedliche Koordinatensysteme zugeordnet sein. Wenn man

$n = 1, 2, 3, \dots$ Koordinaten benötigt, um eine Position eindeutig festzulegen, liegt ein n -dimensionaler Raum vor.

5.4.1. Kreis- und Kugelkoordinaten

Im Dreidimensionalen wird die Position neben dem bereits in Kap. 1.2 kartesischen Koordinatensystem oft auch in KUGELKOORDINATEN angegeben. Das Bezugssystem für Kugelkoordinaten wird i.A. durch Festlegung der folgenden drei Elemente definiert: ein Bezugspunkt und zwei Bezugsrichtungen bzw. Bezugsebenen. Als Bezugspunkt kann man z.B. vom Ort des Beobachters ausgehen (topozentrisches Bezugssystem), vom Erdmittelpunkt (geozentrisches Bezugssystem) oder vom Sonnenmittelpunkt (heliozentrisches Bezugssystem).

Als Referenz des geozentrischen Bezugssystems wählt man meist die Ebene orthogonal zur Drehachse der Erde: die ÄQUATOREBENE. Messungen mit dem Theodolit erfolgen topozentrisch, d.h. der Bezugspunkt ist gewissermaßen der auf einen Punkt geschrumpfte Beobachter bzw. Theodolit. Bezugsebenen sind hier die Horizontalebene des Horizontalkreises und die Vertikalebene des Vertikalkreises, deren senkrecht bzw. vertikal stehenden Achsen sich im Bezugspunkt schneiden.

Koordinatentransformationen zwischen Kugelkoordinaten

Beispiel 1: Ein Meridian ist die Hälfte eines Großkreises, der die Äquatorebene orthogonal schneidet und durch den Pol (meist Nordpol) verläuft. Winkel zwischen Meridianen sind Azimutwinkel. Der bekannteste als Nullmeridian dienende Bezugsmeridian ist der Greenwich-Meridian mit dem Längengrad ϕ_G als Azimutwinkel. Bis zum Jahr 2010 war in der österreichischen Landesvermessung jedoch ein anderes Bezugssystem (bei gleichem Koordinatensystem) mit einem Längengrad ϕ_F in Gebrauch, das sich auf den Ferro-Meridian als Nullmeridian bezog. Der Ferro-Meridian geht auf einen antiken Geographen des 2. Jh. zurück, der die Kanarischen Inseln als Referenz vorschlug. Die Umrechnung der österreichischen auf die internationalen Koordinaten erfolgte gemäß der Koordinatentransformation $\phi_G = \phi_F + 17^\circ 40' 00''$.

Beispiel 2: In der Geodäsie ist es üblich, den Höhenwinkel β relativ zur Äquatorebene als eine der Positionskoordinaten zu wählen. In der Mathematik gibt man stattdessen den Polarwinkel ϑ an, der sich auf eine durch den Pol (bzw. Zenit) verlaufende, auf der Äquatorebene orthogonale Richtung bezieht. Die Koordinatentransformation zwischen Höhenwinkel und Polarwinkel lautet $\vartheta = 90^\circ - \beta$.

Um einen Punkt P im Raum zu charakterisieren, braucht man neben den beiden Winkeln noch eine Abstandskoordinate r vom Bezugspunkt. Im dreidimensionalen Raum ist der aus drei Koordinaten bestehende Satz (r, ϕ, ϑ) dann die Positionsangabe $P(r, \phi, \vartheta)$ in Kugelkoordinaten relativ zum vereinbarten Bezugssystem (Abb. 27). Kreiskoordinaten sind der Spezialfall der Kugelkoordinaten für zweidimensionale Räume. Hier genügt die Abstandsangabe r und Angabe des Azimutwinkels ϕ , d.h. hier ist die Positionsangabe durch $P(r, \phi)$ gegeben.

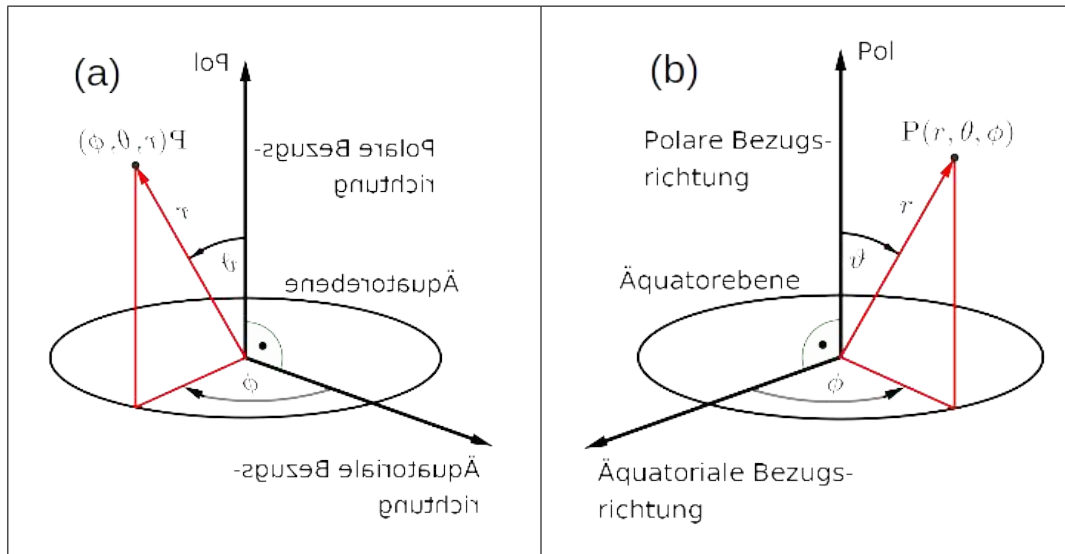


Abbildung 27: Skizze zur Definition der Kugelkoordinaten. (a) Linkshändiges System. (b) Rechtshändiges System.

5.4.2. Händigkeit

Für Kreiskoordinaten im Zweidimensionalen gibt es zwei Möglichkeiten für den positiven Umlaufsinn des Azimutwinkels ϕ , die spiegelsymmetrisch zueinander sind: Der Umlauf kann im Uhrzeigersinn oder gegen den Uhrzeigersinn verlaufen. Entsprechend können die Achsenrichtungen in einem kartesischen Koordinatensystem spiegelsymmetrisch zueinander stehen.

Im Dreidimensionalen benötigt ein Bezugssystem vier Elemente: Einen Bezugspunkt und drei Richtungen. In Abb. 27 ist ein Element die zur Äquatorebene orthogonale Polrichtung, ein weiteres eine äquatoriale Bezugsrichtung. Für die dritte Richtung sind noch zwei Richtungsoptionen frei: Im Dreidimensionalen gibt es zwei spiegelsymmetrische Richtungskonfigurationen, die sich durch ihre HÄNDIGKEIT unterscheiden. Wenn man Kugelkoordinaten zu wählen beabsichtigt, unterscheiden sich diese beiden Möglichkeiten durch den Umlaufsinn des Azimutwinkels ϕ und sind daher spiegelsymmetrisch zueinander. Die Zunahme des Winkels ϕ kann, wie links dargestellt, im Uhrzeigersinn (Rechtsdrehung) oder wie rechts gezeigt, gegen den Uhrzeigersinn (Linksdrehung) erfolgen.

Als „körperliches“ Bezugssystem für die „Händigkeit“ spiegelsymmetrischer Objekte nimmt man konventionsgemäß die rechte Hand und bezeichnet Objekte der gleichen Händigkeit als RECHTSHÄNDIG oder rechts und ihr Gegenstück als linkshändig oder links. Das in Abb. 27 rechts dargestellte Koordinatensystem bezeichnet man als RECHTSHÄNDIGES KOORDINATENSYSTEM. Wenn man den Daumen der rechten (!) Hand in Richtung des Pols weisen lässt, dann sind die Finger im positiven Umlaufsinn des Azimut gekrümmt. Die übliche Konvention ist, dass man für alle räumlichen Beschreibungen rechtshändige Koordinatensysteme zugrunde legt. Im Dreidimensionalen gehen rechts- und linkshän-

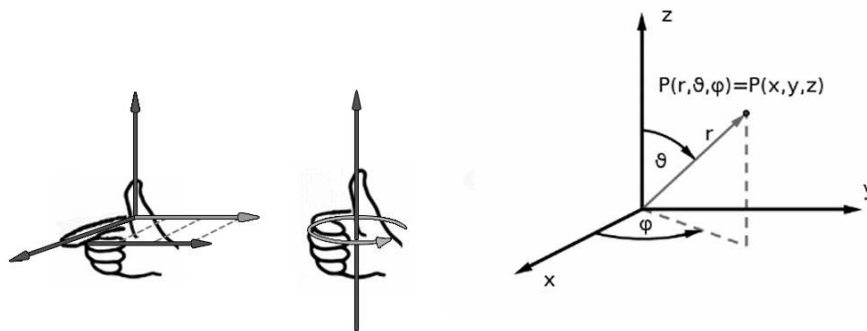


Abbildung 28: Rechtshändige Koordinatensysteme: Weist der Daumen der rechten Hand in die polare Richtung, geben die Finger die positive Drehrichtung des Azimutwinkels ϕ für ein rechtshändiges System der Kugelkoordinaten. Öffnet man den Zeigefinger, so zeigt dieser in die x-Richtung und der Daumen weiterhin in die (polare) z-Richtung, während die restlichen Finger in die y-Richtung eines rechtshändigen kartesischen Koordinatensystems weisen.

dige Koordinatensysteme durch eine Raumspiegelung bzw. eine einfache Spiegelung ineinander über. Im Zweidimensionalen ist das nicht der Fall, d.h. die Raumspiegelung ist gleichbedeutend mit einer Drehung um π .

5.5. Lineare Koordinatentransformationen

Koordinatentransformationen von einem Koordinatensystem S in ein Koordinatensystem S' gibt man durch Gleichungen an, die erklären, wie die Koordinaten von S' aus denen von S auszurechnen sind. Lineare Koordinatentransformationen haben dabei die Eigenschaft, dass eine lineare Beziehung zwischen den Koordinaten von S in eine lineare Beziehung zwischen den Koordinaten von S' transformiert wird.

Zur Illustration werden im Folgenden homogene lineare Koordinatentransformationen in zwei Dimensionen betrachtet. S sei ein zweidimensionales Koordinatensystem mit den Koordinaten (x, y) und S' ein zweidimensionales Koordinatensystem mit den Koordinaten (x', y') . Der Strich bei S' und x' bzw. y' stellt hier nichts weiter als ein Merkmal zur Unterscheidung der Systeme bzw. Koordinaten dar. Die durch

$$\begin{aligned} x' &= \alpha x + \beta y \\ y' &= \gamma y + \delta x \end{aligned} \tag{5.1}$$

erklärte Koordinatentransformation mit vier beliebigen Skalaren α, β, γ und δ ist die allgemeinste homogene lineare Koordinatentransformation, die man sich ausdenken kann. Durch eine lineare Koordinatentransformation werden lineare Beziehungen stets wieder

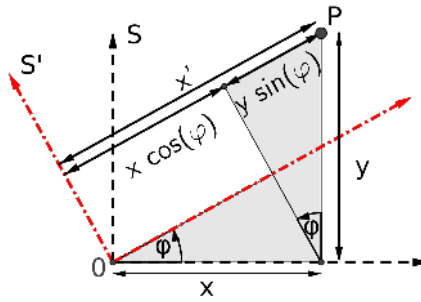


Abbildung 29: S und S' sind zwei kartesische Koordinatensysteme, die um den Winkel φ relativ zueinander gedreht sind. Die Koordinaten des Punktes P bezüglich S sind (x, y) und bezüglich S' sind sie (x', y') . Aus der Zeichnung geht hervor, wie die Koordinate x' aus den beiden Koordinaten des Systems S zu berechnen ist (Gl. 5.2).

in lineare Beziehungen transformiert (und somit Geraden in Geraden). Das kann man folgendermaßen zeigen: Die Gleichung $y' = m'x'$ für eine Gerade mit der Steigung m' stellt eine lineare Beziehung zwischen den Koordinaten in S' dar. Setzt man hier die Koordinatentransformation, Gl. 5.1, ein, so erhält man wieder eine Geradengleichung, nämlich $y = mx$ mit der Steigung $m = (m'\alpha - \delta)/(\gamma - m'\beta)$.

Beispiel: Gedrehte kartesische Koordinatensysteme in zwei Dimensionen

Auch wenn die gleichen rechtshändigen kartesischen Koordinatensysteme und der gleiche Bezugspunkt gewählt wurden, können sich die Koordinaten auf unterschiedliche Bezugssysteme beziehen, deren Achsen gegeneinander gedreht sind. Das ist beispielsweise bei den in Abb. 29 gezeigten Bezugssystemen S und S' der Fall, die relativ zueinander um den Winkel φ gedreht sind. Für den gleichen Ortspunkt P ergeben sich unterschiedliche Koordinaten, die mit (x, y) bezüglich S und mit (x', y') bezüglich S' bezeichnet seien.

Die Koordinaten (x', y') berechnet man aus den Koordinaten (x, y) gemäß dem homogenen linearen Gleichungssystem

$$x' = x \cos \varphi + y \sin \varphi \quad (5.2)$$

$$y' = y \cos \varphi - x \sin \varphi \quad (5.3)$$

Diese Koordinatentransformation ist ein Spezialfall der durch Gl. 5.1 gegebenen allgemeinen linearen Koordinatentransformation, bei dem die Koeffizienten $\alpha = \gamma = \cos \varphi$ und $\beta = -\delta = \sin \varphi$ von einem gemeinsamen Parameter abhängen, nämlich dem Drehwinkel φ . Ihre Herleitung gemäß Abb. 29 setzt voraus, dass ein euklidischer Raum vorliegt.

6. Zeit

Tempus aequabiliter fluit. Die Zeit verfließt gleichförmig. (Isaak Newton, „Philosophiae Naturalis Principia Mathematica“ (1687).

Newton postuliert, dass die Zeit gleichmäßig verfließt. Das bedeutet, dass physikalische Vorgänge unter sonst gleichen Bedingungen jederzeit in gleicher Weise ablaufen. Genauso wie jene der Homogenität des Raumes ist das eine metaphysische Aussage. Sie formuliert die Überzeugung der Physiker, dass es allgemeine physikalische Gesetzmäßigkeiten gibt, die zu Lebzeiten der Pharaonen oder Dinosaurier genauso galten wie sie auch in einer Million Jahren noch gelten werden und auch dann noch, wenn unsere Sonne längst erloschen sein wird. Wer an die Homogenität der Zeit glaubt, der glaubt an die ewige Gültigkeit allgemeiner physikalischer Gesetze: Vom Urknall bis in alle Ewigkeit sind Ladung und Energie eines abgeschlossenen Systems konstant und hat die Elementarladung den Wert 1.60×10^{-19} As. Aus dem Glauben an das ewige Leben der physikalischen Gesetze und Naturkonstanten, aus der Homogenität der Zeit, zieht man die Berechtigung, die in den letzten 400 Jahren aufgefundenen physikalischen Gesetze bis zum Urknall extrapolieren zu dürfen, also auf 14000000000 Jahre hinaus. Da war kein Physiker dabei, der die Newtonsche Zuschreibung hätte überprüfen können.

Die Zeit ist homogen.

Uhren. Die besondere Bedeutung des Postulats von der Homogenität der Zeit liegt darin, dass es ermöglicht, UHREN zu definieren, d.h. Verfahren bzw. Geräte zur Messung der DAUER eines Vorgangs anzugeben. Hat man beispielsweise mehrere Kerzen, die genau gleich lang sind, den gleichen Docht und Durchmesser haben und aus demselben Material bestehen, dann ist deren Brenndauer unter gleichen Bedingungen (gleiche Luftzusammensetzung, gleicher Luftdruck usw.) gleich lang, und zwar – wegen des Newtonschen Postulats von der Homogenität der Zeit – egal wann sie angezündet wurden, denn die Brenndauer hängt nicht vom Zeitpunkt ab, wann der Vorgang gestartet wurde. Hat die Kerze eine Teilung, wie z.B. die in Abb. 30a gezeigte Stundenkerze, dann kann man auf ihr die abgelaufene Zeit ablesen, denn sie ist proportional zur abgebrannten Strecke Δl . Wenn die Dauer eines interessierenden Vorgangs länger als die Brenndauer T einer einzelnen Kerze ist, ersetzt man jede völlig abgebrannte Kerze sofort durch eine neue. Der in Abb. 30b dargestellte Vorgang ist dann eine periodisch laufende Uhr mit der Periode T . Periodisch bedeutet, dass die Kerzenlänge $l(t)$ nach Ablauf einer Dauer von $n = 1, 2, \dots$ Perioden T wieder gleich ist, d.h. die Kerzenlänge ist eine PERIODISCHE FUNKTION der Zeit mit

$$l(t) = l(t + nT). \quad (6.1)$$

Im Sinne von Ernst Mach wird die Messung der Zeit damit zu einem Vergleich der Änderung der Zustände eines Dings gegen ein anderes, nämlich der Zustände der Kerze als Maß der Dauer mit anderen Zuständen eines Objekts, die sich während der Dauer der Beobachtung ändern (oder auch nicht).

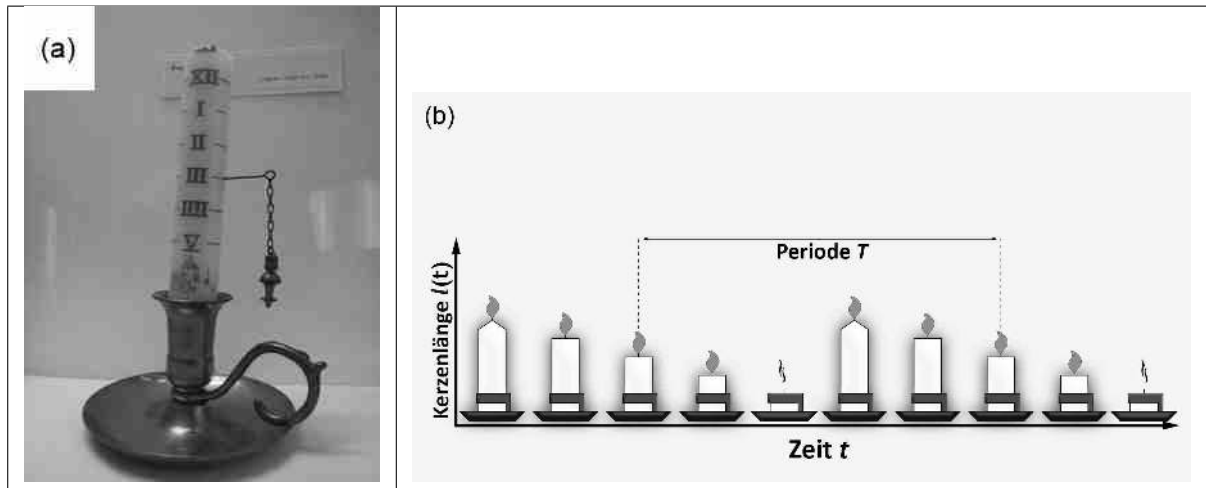


Abbildung 30: (a) Mittelalterliche Stundenkerze mit Zeiteinteilung. (b) Die Kerzenlänge als Zeitmaß.

In der Natur lassen sich physikalische Phänomene finden, die von sich aus periodisch sind. Ein Beispiel ist der Umlauf der Erde um die Sonne, der dem Umlauf der Sternbilder entspricht und das periodische Phänomen der Jahreszeiten hervorruft. Ein weiteres Beispiel ist die Rotation der Erde mit der Periode von 24 Stunden, welche der Rotation der Fixsternsphäre entspricht. Tatsächlich gehörte es bis 1967 zu den vornehmsten Aufgaben der Astronomie, durch genaue Beobachtung der Drehung der Erde gegen den Fixsternhimmel die Zeiteinheit SEKUNDE zu definieren.

Für technisch-praktische Uhren wählt man oft die periodischen Vorgänge bei Pendeln, Schwingquarzen oder atomaren Übergängen. Letzteres ermöglicht die Realisierung der derzeit genauesten Uhren, der ATOMUHREN. Daher ist die Basiseinheit der Zeit

$$[t] = 1 \text{ Sekunde} = 1 \text{ s}$$

heute durch Atomuhren realisiert, welche auf einem bestimmten Energieübergang eines bestimmten Cäsiumisotops beruhen.

6.1. Lokale Zeit

Will man historische EREIGNISSE angeben, benötigt man sowohl die Angabe wo als auch wann es stattfand. Man braucht dazu erstens Lineale und Uhren und zweitens eine Konvention über Bezugs- und Koordinatensysteme.

Setzt man relativ zueinander ruhende Lineale und Uhren voraus, so folgt aus dem Postulat der Homogenität von Raum und Zeit, dass sie überall und zu jeder Zeit baugleich realisiert werden können, denn sie sind damit ja nur an relativ zueinander verschobenen Ort- bzw. Zeitposition realisiert worden, an denen die gleichen physikalischen Gesetze gelten. Für die nachfolgenden Überlegungen wird von einem solchen gedanklich zulässigen Bezugssystem mit überall ruhenden baugleichen Linealen und Uhren ausgegangen.

Nicht nur jede Positionsangabe, sondern auch jede Zeitangabe stellt eine Angabe RELATIV zu einem Referenz- oder Bezugspunkt dar. Referenzereignisse für die Zeitkoordinate kann man erst einmal nur lokal festlegen. Bei der in Abb. 30 gezeigten Zuordnung zwischen der Kerzenlänge und der Zeit t wurde das Ereignis des Anzündens der ersten Kerze als Zeitnullpunkt gewählt. Für den Ort B bzw. Bethlehem gibt man beispielsweise die lokale Zeit relativ zum fiktiven Geburtszeitpunkt von Jesus Christus an und am Ort C bzw. Chonju in Nordkorea wird die lokale Zeit relativ zum fiktiven Geburtszeitpunkt von Kim Il-Sung gemessen. Um zu verdeutlichen, dass die Zahlenwerte der Zeitkoordinate von einer jeweilig lokalen Vereinbarung über den Zeitnullpunkt abhängen, werde ich jeder Zeitangabe ein in eckigen Klammern angegebenes Ortsetikett hinzufügen. Die Zeitkoordinate der in Bethlehem stattfindenden Ereignisse ist also die Jesus-Zeit $t[B]$ relativ zum Referenzereignis am Ort B. Auch wenn man überall Uhren mit identischem Gang bauen kann, wird ein Bewohner von C mit der Zeitangabe eines Ereignisses in B wegen der unterschiedlichen Referenzereignisse zunächst einmal nichts anfangen können.

6.2. Geschwindigkeit und Uhrensynchronisation

Wenn sich jemand von B nach C bewegt, dann verschwindet er zur lokalen Zeit $t_1[B]$ an der Ortskoordinate x_B von B und taucht zur lokalen Zeit $t_2[C]$ an der Ortskoordinate x_C in C wieder auf. Die zurückgelegte Strecke $|\Delta x|$ für einen gegebenen Weg kann im Prinzip mit entlang der Strecke ruhenden Linealen bestimmt werden. Aus der Koordinatendifferenz $\Delta x = x_C - x_B$ und den beiden Zeitangaben, kann man jedoch keinen sinnvollen Begriff der GESCHWINDIGKEIT im Sinne einer zurückgelegten Strecke durch Zeitspanne der Reise bilden. Die Differenz $t_2[C] - t_1[B]$ der lokalen Zeitangaben ist nicht die Reisedauer und das Verhältnis

$$\frac{\Delta x}{t_2[C] - t_1[B]}$$

ist wegen der beiden willkürlichen lokalen Referenzzeiten eine willkürliche Zahl.

Um Geschwindigkeiten messen zu können, muss man zuvor relativ zueinander ruhende, aber an unterschiedlichen Orten befindliche Uhren synchronisieren. Das kann man durch die Methode der EINSTEINSYNCHRONISATION erreichen. Sie beruht auf folgender einfacher Überlegung:

Man betrachtet eine Reise, die aus einem Hinweg von B nach C und aus einem Rückweg, der auf genau dem gleichen Weg von C nach B führt. Diesen Weg teilt man in beliebig kleine (besser: infinitesimal kleine) Teilstrecken ein. Wenn man für jede dieser Teilstrecken sicherstellen kann, dass sie auf dem Hin- wie Rückweg – abgesehen von der unterschiedlichen Richtung – in der gleichen Weise durchlaufen werden, dann dauert der Hinweg genauso lang wie der Rückweg. Für diesen Schluss kann man sich auf das Postulat von der Isotropie des Raumes berufen. Ob die Wege gleich durchlaufen wurden oder nicht, kann der Reisende durch eine mitgeführte Taschenuhr überprüfen, indem er für jede Teilstrecke feststellt, ob er in der Hin- wie Rückrichtung die gleiche Zeitdifferenz auf der relativ zu ihm in Ruhe befindlichen Taschenuhr abliest. Der Schluss ist auch dann

richtig, wenn der Gang der Uhr von der Geschwindigkeit abhängt. Wichtig ist nämlich nur, dass der Gang der Uhr von der Bewegungsrichtung unabhängig ist.

Es ergebe sich nun, dass der Reisende zur Zeit $t_1[B]$ in B startet, zur Zeit $t_2[B]$ wieder zum Ausgangsort B zurückkehrt und dass die Prüfung durch die Taschenuhr des Reisenden bestätigt, dass Hin- und Rückweg abgesehen von der unterschiedlichen Richtung gleich verlaufen sind. Im Einklang mit den Hypothesen über Raum und Zeit ist dann die Dauer der Hinreise gleich der Dauer der Rückreise und durch

$$\Delta t = \frac{1}{2} (t_2[B] - t_1[B]) \quad (6.2)$$

gegeben.

Das Ereignis der Ankunft des Reisenden in Chonju ist damit sowohl durch Ablesung der lokalen Zeit $t_0[C]$ bekannt als auch durch die Zeitkoordinate $t_0[B] + \Delta t$ in Jesus-Zeit. Egal ob man nun die eine oder die andere Skala zur universellen Zeitskala erklärt, die Uhren sind damit synchronisiert. Das für die Unterscheidung der lokalen Zeiten benötigte Symbol (eckige Klammer mit Orts-Etikett) kann damit entfallen. Relativ zu einem Bezugssystem S man kann sich von nun an auf die gleiche SYSTEMZEIT t beziehen, die auf synchronisierten Uhren abgelesen werden kann, die relativ zu S ruhen. Erst durch die Einsteinsche Uhrensynchronisation erhält der Begriff der GLEICHZEITIGKEIT zweier Ereignisse, die an unterschiedlichen Orten B und C stattfinden, einen konkreten Sinn.

Als Konsequenz wird die mittlere (Bahn-)Geschwindigkeit

$$u = \frac{\Delta x}{\Delta t}, \quad (6.3)$$

zu einem sinnvollen physikalischen Begriff, denn sowohl Δx als auch Δt beziehen sich auf Messungen mit Instrumenten, die relativ zum gleichen Bezugssystem in Ruhe sind.

Liegen überall ruhende synchronisierte Uhren vor, kann man übrigens überprüfen, ob die Hypothese von der Homogenität von Raum und Zeit konsistent sind, indem man die Dauer gleicher Vorgänge misst, die an verschiedenen Orten und zu verschiedenen Zeiten gestartet wurden. Aus solchen empirischen Tests hat sich bislang kein Widerspruch zur Grundhypothese der Homogenität ergeben.

6.3. Orts-Zeit-Diagramme

Mit überall baugleich vorliegenden Linealen und synchronisierten Uhren kann man definieren, was eine gleichförmig geradlinige Bewegung sein soll: Sie ist GLEICHFÖRMIG, wenn für jede Größe eines gewählten Streckenabschnitts gleiche Streckenabschnitte in gleichen Zeiten zurückgelegt werden. Sie heißt GERADLINIG, wenn sie auf einem solchen Weg erfolgt, dass die zurückgelegte Entfernung zwischen Anfangs- und Endpunkt die kürzest mögliche ist.

Eine geradlinige Bewegung kann durch eine Ortskoordinate x beschrieben und durch graphische „Fahrpläne“ in einem ORTS-ZEIT-DIAGRAMM dargestellt werden (Abb. 31). In dieses x-t-Diagramm kann man die bei einer Bewegung beobachteten Ereignisse eintragen.

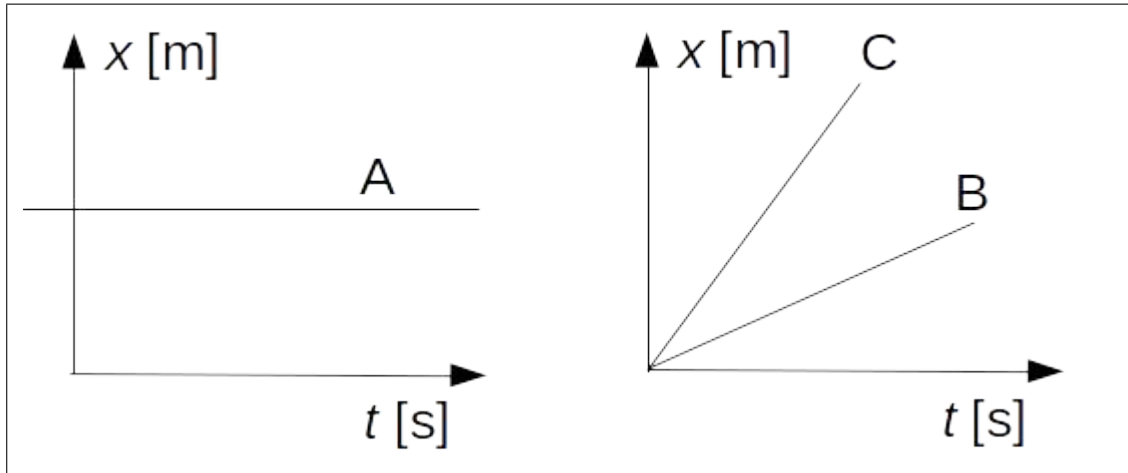


Abbildung 31: Darstellung dreier Weltlinien in Orts-Zeit-Diagrammen. A: Objekt ruht relativ zum Bezugssystem. B und C: Objekte bewegen sich bezüglich des Bezugssystems gleichförmig.

Die dabei entstehenden Kurven bzw. Linien werden als WELTLINIEN bezeichnet. Weltlinie A zeigt an, dass das Objekt trotz voranschreitender Zeit am gleichen Ort verbleibt, d.h. es ist die Weltlinie eines ruhenden Objekts, beispielsweise eines der ruhenden Systemlineale oder Systemuhren. Bei Weltlinie B und C werden gleiche Wegstrecken in gleichen Zeiten zurückgelegt. Somit sind es gleichförmige Bewegungen mit konstanter Geschwindigkeit. Die Geschwindigkeit ist für C größer als für B.

7. Äquivalenzprinzipien

7.1. Äquivalenzprinzip für invertierte Bezugssysteme

Bezugssysteme, die durch eine Rauminversion auseinander hervorgehen, also Bezugssysteme unterschiedlicher Händigkeit (s. Kap. 5.4.2), sind physikalisch äquivalent. Die Rauminversion ist wichtig für die Unterscheidung zwischen Skalaren und Vektoren.

Skalare und Vektoren. Im Sinne der Physik werden alle Größen als SKALARE bezeichnet, die keinen Bezug zum Raum haben und daher auch von Raumtransformationen nicht betroffen sind. Beispiele für Skalare sind Zeit, Masse, Temperatur oder Ladung. VEKTOREN im physikalischen Sinne sind Objekte mit einem sehr spezifischen Transformationsverhalten gegenüber einer Rauminversion: sie ändern bei Rauminversion das Vorzeichen. Der Ort x und die Geschwindigkeit u sind beispielsweise Vektoren.

7.2. Äquivalenzprinzip für gedrehte Bezugssysteme

Für Bezugssysteme, deren Bezugspunkte zusammenfallen, die aber relativ zueinander gedreht sind, geht man wegen der Isotropie des Raumes von folgendem Äquivalenzprinzip aus:

Relativ zueinander gedrehte Bezugssysteme sind physikalisch äquivalent.

Das bedeutet, dass alle physikalischen Gesetze, die man relativ zu diesen Bezugssystemen formuliert, die gleiche Form haben und alle Feststellungen, die man in einem Bezugssystem macht, in gleicher Weise auch im dazu äquivalenten Bezugssystem gemacht werden können. Eine der Konsequenzen dieses Äquivalenzprinzips soll hier untersucht werden. Um die Darstellung einfach zu halten, werden die nachfolgenden Überlegungen für eine Ebene durchgeführt. Sie lassen sich leicht auf den dreidimensionalen Raum erweitern.

Gegeben seien zwei Bezugssysteme S und S' mit gemeinsamem Bezugspunkt und den kartesischen Koordinaten (x, y) für S bzw. (x', y') für S' . System S' sei relativ zu S um einen Winkel φ gedreht. Umgekehrt ist S relativ zu S' um $-\varphi$ gedreht. Gesucht ist die Koordinatentransformation zwischen S und S' . Das Äquivalenzprinzip verlangt u.a. folgendes: Setzt man bezüglich S das Axiom, dass Lichtwege Geraden sind, dann ist S' nur dann ein physikalisch äquivalentes Bezugssystem, wenn auch relativ zu diesem Lichtwege Geraden sind. Die mathematische Konsequenz ist, dass lineare Relationen unter der Koordinatentransformation erhalten bleiben müssen. Lineare Koordinatentransformationen leisten das. Man kann also von der allgemeinst möglichen linearen Transformation, Gl. 5.1, ausgehen, wobei allerdings die Parameter α, β, γ und δ vom Winkel φ abhängen (und somit Funktionen von φ sind, aber nicht von den Koordinaten).

Eine Grundvoraussetzung der Äquivalenz ist, dass die Bezugselemente des einen Systems im anderen System die gleiche Beschreibung haben. Da der Nullpunkt von S und S' der gleiche Bezugspunkt sein soll, geht es nur noch um die relative Beschreibung der Achsenlagen. In S' erfüllt die x' -Achse die Gleichung $y' = 0$ für alle x' , während sie in S eine Gerade mit der Gleichung $y = vx$ ist, welche die Steigung

$$v = \frac{\Delta y}{\Delta x} = \tan \varphi$$

hat. Setzt man das in Gl. 5.1 ein, so erhält man $\delta = -v\gamma$. Umgekehrt erfüllt die x -Achse die Gleichung $y = 0$ für alle x in S und $y' = -vx'$ in S' . Daraus folgt $\alpha = \gamma$. Jede Koordinatentransformation zwischen zwei relativ zueinander gedrehten Bezugssystemen hat daher die Form

$$\begin{aligned} x' &= \gamma(v)x + \beta(v)y \\ y' &= \gamma(v)(y - vx) \end{aligned} \quad (7.1)$$

wobei die Koeffizienten β und γ vom Parameter $v = \tan \varphi$ abhängen.

Nun betrachtet man ein weiteres Bezugssystem S'' mit den Koordinaten (x'', y'') , dessen x'' -Achse relativ zu S die Steigung u hat und relativ zu S' die Steigung u' . Das Äquivalenzprinzip verlangt, dass die Koordinatentransformation die gleiche Form wie Gl. 7.1

hat und daher gilt:

$$\begin{aligned}x'' &= \gamma(u)x + \beta(u)y \\y'' &= \gamma(u)(y - ux)\end{aligned}\tag{7.2}$$

und

$$\begin{aligned}x'' &= \gamma(u')x' + \beta(u')y' \\y'' &= \gamma(u')(y' - u'x')\end{aligned}\tag{7.3}$$

Setzt man diese Gleichungen ineinander ein, so erhält man durch Koeffizientenvergleich

$$\gamma(u) = \gamma(u')\gamma(v) - v\beta(u')\gamma(v)\tag{7.4}$$

$$\beta(u) = \gamma(u')\beta(v) + \beta(u')\gamma(v)\tag{7.5}$$

$$\gamma(u) = \gamma(u')\gamma(v) - u'\beta(v)\gamma(u')\tag{7.6}$$

$$u\gamma(u) = (v + u')\gamma(u')\gamma(v)\tag{7.7}$$

Aus Gl. 7.4 und Gl. 7.6 kann man schließen, dass

$$K = \beta(v)/v\gamma(v) \quad (\text{gültig für beliebige } v)\tag{7.8}$$

eine für alle relativ zueinander gedrehten Bezugssysteme universell gültige Konstante ist und aus Gl. 7.6 und Gl. 7.7, dass zwischen den Steigungen u , v und u' die Relation

$$u = \frac{v + u'}{1 - Ku'v}\tag{7.9}$$

besteht.

Nun vergleicht man die Transformation für ein System S' , das um φ nach links gedreht wurde mit einem, das um $-\varphi$ nach rechts gedreht wurde. Im ersten Fall gilt für die x' -Achse $y = vx$ und im zweiten Fall $y = -vx$. Dann muss gelten

$$x' = \gamma(v)x + \beta(v)vy$$

und

$$x' = \gamma(-v)x - \beta(-v)vy.$$

Wegen der Isotropie des Raumes darf die Transformation nämlich nicht von der Richtung der Drehung abhängen. Daraus folgen die Symmetriebeziehungen

$$\gamma(v) = \gamma(-v)$$

$$\beta(v) = -\beta(-v)$$

Setzt man in Gl. 7.6 $u = 0$ und $u' = -v$ ein, so erhält man mit Hilfe von Gl. 7.8:

$$\gamma(v) = \frac{1}{\sqrt{1 + Kv^2}}\tag{7.10}$$

Damit kann man die Gleichungen der Koordinatentransformation explizit niederschreiben:

$$x' = \gamma(v)(x + Kvy)\tag{7.11}$$

$$y' = \gamma(v)(y - vx)\tag{7.12}$$

Die Konstante K heißt geometrische Strukturkonstante und ist dimensionslos. Sie verbindet die eindimensionalen Koordinatenräume x und y zu einem zweidimensionalen Koordinatenraum mit einer gemeinsamen Metrik.

Es sei nun eine Länge l betrachtet, deren erstes Ende im Ursprung liegt und deren zweites Ende sich in einer Entfernung $l^2 = x^2 + y^2$ befindet. In S' ergibt sich aufgrund der Koordinatentransformation die Länge

$$l'^2 = x'^2 + y'^2 = \gamma^2 [x^2 + 2Kvxy + K^2v^2y^2 + y^2 - 2vxy + v^2x^2]$$

Wenn man für gedrehte Bezugssysteme verlangt, dass in einem euklidischen Raum $K = K_e$ mit der positiven euklidischen Strukturkonstanten $K_e = 1$ gilt und somit $l = l'$ gilt, so folgt daraus

$$\gamma = \frac{1}{\sqrt{1 + \tan^2 \varphi}} = \cos \varphi$$

bzw.

$$\begin{aligned} x' &= \cos \varphi x + \sin \varphi y \\ y' &= -\sin \varphi x + \cos \varphi y \end{aligned} \quad (7.13)$$

Das ist wiederum die Koordinatentransformation, die in Abschnitt 5.5 abgeleitet wurde. Aus der Sicht des Physikers folgt Gl. 7.13 aus drei Annahmen: der Isotropie des Raumes, der Euklidizität des Raumes und dem Äquivalenzprinzip für gedrehte Bezugssysteme.

7.3. Galileisches Äquivalenzprinzip

Galileo Galilei hat die bereits in Kap. 5.2 zitierten Beobachtungen in einem ruhenden Bezugssystem um solche erweitert, die man in Bezugssystemen machen kann, die sich sowohl geradlinig als auch gleichförmig zueinander bewegen [3]:

„Nun lasst das Schiff sich mit jeder beliebigen Geschwindigkeit bewegen. Wenn die Bewegung gleichförmig und geradlinig ist, werdet ihr bei allen Erscheinungen nicht die geringste Veränderung eintreten sehen. Aus keiner derselben werdet ihr entnehmen können, ob das Schiff fährt oder still steht. Die Fische im Wasser werden sich nicht mehr anstrengen müssen, um nach dem vorangehenden Teile des Gefäßes zu schwimmen als nach dem hinteren. Sie werden sich vielmehr mit gleicher Leichtigkeit nach dem Futter begeben, auf welchen Punkt des Gefäßrandes man es auch legen mag. Auch die Mücken und Schmetterlinge werden ihren Flug ganz ohne Unterschied nach allen Richtungen fortsetzen. Niemals wird es vorkommen, dass sie gegen die dem Hinterteil zugekehrte Wand gedrängt werden, gewissermaßen müde von der Anstrengung, dem schnell fahrenden Schiffe nachfolgen zu müssen. Und doch sind sie während ihres langen Aufenthalts in der Luft von ihm getrennt.“

Seine Beobachtungen legen nahe, dass jeder physikalische Vorgang in allen Bezugssystemen, die sich in Ruhe oder in gleichförmig-geradliniger Bewegung zueinander befinden, in ununterscheidbar gleicher Weise abläuft. Erhebt man sie zu einem Prinzip, so etabliert dieses Galileische Äquivalenzprinzip, das man auch als RELATIVITÄTSPRINZIP bezeichnet, eine neue physikalische Symmetrie.

Die Bezugssysteme der Mechanik zerfallen demnach in Klassen von äquivalenten Bezugssystemen, die GALILEIKLASSEN. Alle zu einer bestimmten Galilei-Klasse gehörenden Bezugssysteme heißen GALILEISYSTEME dieser Klasse. Galileisysteme der gleichen Klasse bewegen sich relativ zueinander gleichförmig und geradlinig.

Galileisches Äquivalenzprinzip

Es existieren Klassen von Bezugssystemen, deren Bezugssysteme (Galileisysteme) sich sowohl gleichförmig als auch geradlinig zueinander bewegen. Die Galileisysteme der gleichen Klasse zeichnen sich dadurch aus, dass für in ihnen ruhende Beobachter die physikalischen Phänomene ununterscheidbar gleich ablaufen.

Galileisysteme der gleichen Klasse sind physikalisch äquivalent.

Es fällt am Anfang schwer, die Allgemeingültigkeit des Relativitätsprinzips so ohne weiteres zu akzeptieren, denn sie wurde von Galilei aus wenigen Fakten herausgelesen und visionär verallgemeinert. Dass das so erfolgreich war, macht das Genie dieses Mannes aus.

Beschleunigung. Da der Typus der gleichförmigen und geradlinigen Bewegung für das Relativitätsprinzip eine so herausragende Rolle spielt, definiert man für Bewegungen, die davon abweichen, den Gegenbegriff der BESCHLEUNIGTEN BEWEGUNG bzw. der BESCHLEUNIGUNG:

Ist eine Bewegung nicht geradlinig oder nicht gleichförmig, so heißt sie beschleunigt.

Ist B ein Bezugssystem und B' ein dazu beschleunigtes Bezugssystem, dann gehört es einer anderen Galilei-Klasse als B an. Das gilt auch umgekehrt: Beschleunigung ist ein relativer Begriff. Im Rahmen der Kinematik kann man nur sagen, dass ein Objekt *relativ* zu einem anderen Bezugssystem beschleunigt ist, aber nicht, dass es in einem absoluten Sinn beschleunigt ist. Erst im Rahmen der Dynamik kann man die Frage stellen, ob es ein gewissermaßen „natürliches Bezugssystem“ gibt, relativ zu dem Beschleunigungen uns als quasi-absolut erscheinen (s. Kap. C).

8. Spezielle Relativitätstheorie

Die SPEZIELLE RELATIVITÄTSTHEORIE wurde vor etwas mehr als einem Jahrhundert von ALBERT EINSTEIN entwickelt. Sie befasst sich mit einer speziellen Klasse von relativ zueinander bewegten Bezugssystemen, welche den Voraussetzungen des Relativitätsprinzips von Galileo Galilei genügen. Für diese Klasse von Bezugssystemen beantwortet seine

Theorie die Frage, wie sich bewegte Lineale und Uhren verhalten und wie die Koordinaten der relativ zueinander bewegten Bezugssysteme zu transformieren sind. Dieses Kapitel behandelt nur den rein kinematischen Aspekt seiner Theorie. Der dynamische Teil wird in Teil III dargestellt. Im Rahmen der von Einstein später entwickelten Theorie der Gravitation, die als Allgemeine Relativitätstheorie bekannt ist, ist die Spezielle Relativitätstheorie als jener Grenzfall enthalten, bei dem die Gravitationskonstante G verschwindet.

8.1. Die Axiome einer Theorie von Raum und Zeit

Die als spezielle Relativitätstheorie bezeichneten Theorie der Raumzeit geht von den folgenden drei Axiomen der Kinematik aus:

1. Homogenität der Zeit.
2. Homogenität und Isotropie des Raums.
3. Gültigkeit des Relativitätsprinzips.

Im Folgenden werden die logischen Konsequenzen des dadurch festgelegten Modells von Raum und Zeit hergeleitet. Da es primär um das Verständnis des Zusammenhangs von Raum und Zeit geht, genügt es, neben der Zeitkoordinate nur eine einzige Raumkoordinate zu betrachten, also eindimensionale Bewegungen zu untersuchen.

So lange man durch empirische Erfahrung auf keinen Widerspruch zu den drei axiomatischen Behauptungen stößt, gelten die daraus gezogenen Schlussfolgerungen und beanspruchen auch Gültigkeit für die physikalische (reale) Raum-Zeit. So lange wie Gravitation keine Rolle spielt, haben sich die Axiome und alle Schlussfolgerungen daraus empirisch bewährt.

8.2. Koordinatentransformationen zwischen Galileisystemen

Mit S , S' und S'' seien drei Galileisysteme bezeichnet, die zur gleichen Galilei-Klasse gehören. Ihre jeweiligen Bezugspunkte (Koordinatennullpunkte) bewegen sich entlang der x -, x' -, und x'' - Achse der jeweiligen Koordinatensysteme. Das Galileiprinzip gewährleistet, dass für jedes Galileisystem Raum und Zeit homogen sind und der Raum isotrop ist. Daher darf man sich für jedes der drei Galileisysteme vorstellen, dass dort relativ zum jeweiligen System ruhende Lineale und Uhren existieren, die überall (Homogenität des Raumes) und immer (Homogenität der Zeit) baugleich sind. Für ein gegebenes Galileisystem „ticken“ die dort ruhenden Uhren nicht nur im gleichen Takt, sondern sollen auch alle synchronisiert sein. In jedem der Galileisysteme ist durch die Einsteinsynchronisation definiert, was unter der Gleichzeitigkeit zweier Ereignisse an zwei unterschiedlichen Orten zu verstehen ist, nämlich dass sie zur gleichen auf den Uhren angezeigten Systemzeit an ihren jeweiligen Orten stattfinden. Damit kann auch in jedem Galileisystem eine klare Angabe darüber gemacht werden, mit welcher Geschwindigkeit sich ein im anderen Galileisystem ruhender Körper bewegt.

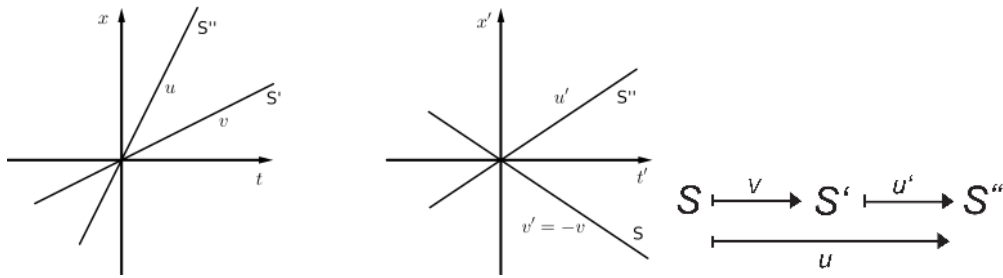


Abbildung 32: (a) Weltlinien des Ursprungs der Galileisysteme S' und S'' dargestellt in S . (b) Weltlinien des Ursprungs der Galileisysteme S und S'' dargestellt in S' . (c) Relativgeschwindigkeiten zwischen den drei Galileisystemen.

Um die nachfolgende Untersuchung der Koordinatentransformation zwischen Galileisystemen besonders einfach und durchsichtig zu machen, soll nun eine spezielle, aber zulässige Wahl der Koordinaten für drei sich relativ zueinander bewegendes Systeme getroffen werden: Man betrachtet drei in ihren jeweiligen Galileisystemen ruhende Uhren (Clocks) C , C' und C'' , die sich gerade begegnen. Dem Ereignis der Begegnung kann man in allen drei Systemen die Koordinate $x = x' = x'' = 0$ und die Zeit $t = t' = t'' = 0$ zuordnen. Für das Begegnungsereignis sind damit alle (!) Uhren in allen (!) Galileisystemen auf den gleichen Zeitpunkt $t = t' = t'' = 0$ synchronisiert.

Die mit den in S vorliegenden Linealen und Uhren gemessene Geschwindigkeit beliebiger (und damit aller) in S' ruhender Punkte sei $v = \Delta x / \Delta t$. Verkürzt sagt man dazu, dass sich das Bezugssystem S' mit der konstanten Geschwindigkeit v relativ zu S bewegt. Da es wegen der Isotropie des Raumes keinen Unterschied ausmachen darf, ob sich etwas in der positiven oder negativen Richtung bewegt und zwei Galileisysteme physikalisch gleich sind, folgt für die Geschwindigkeit $v' = \Delta x' / \Delta t'$ der Bewegung von S relativ zu S' , dass $v' = -v$ ist. Jeder in S' ruhende Punkt bewegt sich von S aus gesehen also mit der Geschwindigkeit v parallel zur x -Achse und jeder in S ruhende Punkt von S' aus gesehen mit $-v$ entlang der x' -Achse. Alle bezüglich eines dritten Galileisystems S'' ruhenden Punkte mögen sich relativ zu S mit einer konstanten Geschwindigkeit u und bezüglich S' mit einer konstanten Geschwindigkeit u' bewegen.

Die zu klärende Frage ist, wie sich die Koordinaten (t, x) bzw. (t', x') zwischen S und S' transformieren und wie sich die Geschwindigkeitskoordinaten u bzw. u' transformieren, wenn die Relativgeschwindigkeit von S' relativ zu S gleich v ist bzw. die Relativgeschwindigkeit von S relativ zu S' $v' = -v$ ist.

Über die Transformation

$$u' = u'(u, v) \tag{8.1}$$

der Geschwindigkeitskoordinaten kann man erst einmal nichts sagen, wohl aber etwas über die Zeit-Orts-Koordinatentransformation. In Gl. 8.1 soll mit der Schreibweise $u'(u, v)$ ausgedrückt werden, dass u' eine Funktion sowohl von u als auch von v ist.

Wie in Abb. 32 gezeigt, ist die Weltlinie eines jeden in S'' ruhenden Punktes sowohl

im Raum-Zeit-Diagramm von S als auch von S' eine Gerade, weil sie zur gleichen Galileiklasse gehören. Das schränkt die in Frage kommenden Koordinatentransformationen gewaltig ein, denn es kommen nur solche in Betracht, welche die Linearität von Koordinatenbeziehungen unangetastet lassen. Wenn man also die Forderung erhebt, dass eine Koordinatentransformation $t' = t'(t, x, v)$ bzw. $x' = x'(t, x, v)$ die Bedingung erfüllen muss, dass sie gerade Weltlinien stets in gerade Weltlinien transformiert, dann muss es sich um eine lineare Transformation in t und x handeln (s. Kap. 5.5). Für den allgemeinst möglichen Ansatz⁸

$$\begin{aligned} t' &= \alpha(v)t + \beta(v)x \\ x' &= \gamma(v)x + \delta(v)t \end{aligned} \quad (8.2)$$

der linearen Transformation werden vier Koeffizienten benötigt, die mit α, β, γ und δ bezeichnet wurden. Sie hängen von der Relativgeschwindigkeit v ab, was hier als Argument explizit kenntlich gemacht wurde.

Das Galileische Relativitätsprinzip ist ein Äquivalenzprinzip. Analog zu Kap. 7.2 müssen sich äquivalente Bezugssysteme zunächst einmal relativ zueinander in gleicher Form beschreiben. Da die Orientierung der Achsen in x -Richtung gleich gewählt wurden, geht es nur um die Lage des Ursprungs der jeweiligen Bezugssysteme. Der Ursprung von S' hat dort stets den Wert $x'=0$. Bezüglich S bewegt sich dieser Punkt mit $x = vt$. Daraus folgt:

$$\delta(v) = -v\gamma(v).$$

Von S' aus gesehen bewegt sich der Ursprung $x = 0$ des Systems S gemäß $x' = v't' = -vt'$. Aus Gl. 8.2 erhält man damit $t' = \alpha(v)t$ und $x' = \delta(v)t$ und somit

$$\alpha(v) = \gamma(v).$$

Damit hat die Koordinatentransformation die gleiche Form wie diejenige, die in Kap für relativ zueinander gedrehte Bezugssysteme abgeleitet wurde, nämlich

$$\begin{aligned} t' &= \gamma(v)t + \beta(v)x \\ x' &= \gamma(v)(x - vt) \end{aligned} \quad (8.3)$$

Man betrachtet nun, wie in Abb. 32c gezeigt, ein drittes Galileisystem, das sich relativ zu S mit u und relativ zu S' mit u' bewegt. Abgesehen davon, dass diese Geschwindigkeiten als Argumente für γ und β einzusetzen sind, müssen die Koordinatentransformationen wegen der Äquivalenz der Galileisysteme die gleiche Form wie Gl. 8.3 haben. Sie lauten daher:

$$\begin{aligned} t'' &= \gamma(u)t + \beta(u)x \\ x'' &= \gamma(u)(x - ut) \end{aligned} \quad (8.4)$$

⁸Ein ANSATZ ist gewissermaßen ein mathematisch ausformulierter „Testballon“. Man bezeichnet damit den Ausgangspunkt eines Lösungsversuchs. In einem Ansatz sind meist noch Konstanten oder Funktionen zu bestimmen. Wenn dies widerspruchsfrei gelingt, wird aus dem Ansatz eine Lösung. Ansonsten muss der Ansatz als missglückter „Testballon“ verworfen werden.

und

$$\begin{aligned} t'' &= \gamma(u')t' + \beta(u')x' \\ x'' &= \gamma(u')(x' - u't') \end{aligned} \quad (8.5)$$

Setzt man Gl. 8.3 in Gl. 8.5 ein und vergleicht mit Gl. 8.4, so erhält man

$$\begin{aligned} \gamma(u)t + \beta(u)x &= \gamma(u')[\gamma(v)t + \beta(v)x] + \beta(u')\gamma(v)(x - vt) \\ \gamma(u)(x - ut) &= \gamma(u')(\gamma(v)[x - vt] - u'[\gamma(v)t + \beta(v)x]) \end{aligned}$$

Durch Koeffizientenvergleich ergeben sich vier gleichberechtigte Formulierungen für ein Gesetz der relativistischen Geschwindigkeitstransformation, wobei hier nur drei davon angegeben sind:

$$\gamma(u) = \gamma(u')\gamma(v) - v\beta(u')\gamma(v) \quad (8.6)$$

$$\gamma(u) = \gamma(u')\gamma(v) - u'\beta(v)\gamma(u') \quad (8.7)$$

$$u\gamma(u) = (v + u')\gamma(u')\gamma(v) \quad (8.8)$$

Aus Gl. 8.6 und Gl. 8.7 folgt

$$\frac{\beta(v)}{v\gamma(v)} = \frac{\beta(u')}{u'\gamma(u')}$$

Da dies für beliebige Geschwindigkeiten v und u' zu gelten hat, müssen diese Ausdrücke konstant sein. Man führt eine Konstante K ein und schreibt

$$\beta(v) = Kv\gamma(v).$$

Die Größe K hat also für jedes Galileisystem den gleichen Wert. Man bezeichnet eine solche Größen als INVARIANTE. Setzt man das in Gl. 8.7 ein, so erhält man

$$\gamma(u) = \gamma(u')\gamma(v)(1 - Ku'v)$$

und mit Gl. 8.8 folgt, dass sich die Geschwindigkeiten gemäß

$$u = \frac{v + u'}{1 - Ku'v} \quad (8.9)$$

transformieren und die Koordinaten gemäß

$$t' = \gamma(v)(t + Kvx) \quad (8.10)$$

$$x' = \gamma(v)(x - vt) \quad (8.11)$$

Für $v = 0$ folgt daraus $\gamma(0) = 1$. Nun sei eine im System S bei $x = 0$ ruhende Uhr C betrachtet, die mit fortschreitender Zeit an Uhren des Systems S' vorbeilaufe. Wenn die Uhr C bei der Uhr C' am Ort x' in S' eintrifft, wird dort gemäß Gl. 8.10 die Zeit

$$t'(x') = \gamma(v)t(x = 0) \quad (8.12)$$

angezeigt. Da die Uhrzeit in S' wegen der Isotropie des Raumes nicht von der Bewegungsrichtung abhängen kann, muss gelten:

$$\gamma(-v) = \gamma(v). \quad (8.13)$$

Setzt man speziell $u = 0$ und $u' = -v$ in Gl. 8.7 ein, so erhält man

$$\gamma(v) = \frac{1}{\sqrt{1 + Kv^2}}. \quad (8.14)$$

8.3. Die Raum/Zeit-Konstante c

In die Herleitung der mathematischen Form der Lorentzgleichungen gehen nur die Homogenität und Isotropie des Raumes, die Homogenität der Zeit sowie das Relativitätsprinzip ein. Das waren dann auch schon alle Zutaten. Zweifelt man die Einsteinsche Relativitätstheorie an, so ist das gleichbedeutend damit, dass man Zweifel an einem dieser drei axiomatischen Eckpfeiler der Physik hat, aus denen die Relativitätstheorie logisch folgt. Die Frage des Vorzeichens und des Zahlenwerts von K ist allerdings keine mathematisch zu beantwortende Frage, sondern eine physikalische.

Hier soll davon ausgegangen werden, dass nicht nur ein Maß für die Einheit Sekunde definiert ist, sondern auch, wie es zu Lebzeiten Einsteins noch war, durch das Pariser Urmeter ein Maß für die Länge. Unter dieser Voraussetzung kann man die Frage stellen, wie man die Konstante K messen kann. Der Schlüssel ist das Transformationsgesetz der Geschwindigkeit (Gl. 8.9). Durch Messung der Geschwindigkeiten u und u' eines Teilchen in zwei Bezugssystemen, die eine Relativgeschwindigkeit v zueinander haben, folgt

$$K = -\frac{v + u' - u}{vuu'}. \quad (8.15)$$

Gl. 8.15 ist die Grundlage aller experimentellen Methoden, durch die K sich am genauesten bestimmen lässt. Man zieht zur Messung Teilchen heran, die man auf möglichst hohe Geschwindigkeiten bringen kann, beispielsweise Elementarteilchen, die aus hochenergetischen Zerfällen stammen, d.h. aus Zerfällen, bei denen die freiwerdende Energie die Masse der Elementarteilchen bei weitem übersteigt. Am einfachsten geht das mit Photonen, d.i. den Teilchen, aus denen sich die Lichtstrahlung zusammensetzt. Man kann aber auch geladene Elementarteilchen verwenden, die in Hochenergie-Beschleunigern auf sehr hohe Geschwindigkeiten gebracht werden können. Im Rahmen der Messunsicherheit findet man in allen Experimenten mit egal welchen beweglichen Objekten größenordnungsmäßig

$$K \approx -10^{-17} \text{ m}^2/\text{s}^2.$$

Es hat sich eingebürgert, statt K eine Konstante c einzuführen, so dass man die negative Konstante K gemäß

$$K = -1/c^2$$

durch eine positiv definierte Konstante c ausdrückt. Während K die physikalische Dimension eines reziproken Geschwindigkeitsquadrats hat, ist die RAUM/ZEIT-KONSTANTE c

ein Skalar mit der physikalischen Dimension einer Geschwindigkeit. Sie ist jedoch keine Geschwindigkeit! Das ist klar, denn Geschwindigkeiten sind Vektoren, während K und somit auch c einfache Skalare sind. Die Sachlage ist ähnlich wie mit der quantenmechanischen Strukturkonstante \hbar für die Unschärfe. Auch wenn diese die gleiche physikalische Dimension wie ein Drehimpuls hat, handelt es sich selbstverständlich nicht um einen Drehimpuls, denn letzterer ist ein Pseudovektor, während \hbar ein Skalar ist. Die Strukturkonstante c der Raumzeit ist eine der universellen skalaren Konstanten der Physik. Ihr Zahlenwert

$$c \approx 3.00 \times 10^8 \text{ m/s}.$$

bestimmt gewissermaßen die Geometrie der Raumzeit. Man bestimmt c über Gl. 8.9 bzw. Gl. 8.16, indem man die Geschwindigkeiten u und u' eines Teilchens bezüglich zweier Bezugssysteme misst, welche sich mit einer Geschwindigkeit v relativ zueinander bewegen. Aus der Existenz von c folgt daher nicht, dass es ein Teilchen geben müsste, das sich mit dem Geschwindigkeitsbetrag c bewegt. Um bewusst zu machen, dass sich c aus rein kinematischen Messungen ergibt, die man mit beliebigen Teilchen durchführen kann, wird die Größe in diesem Buch nicht als Vakuumlichtgeschwindigkeit, sondern als Raum/Zeit-Konstante bezeichnet.

Hier sollen noch einmal alle wichtigen Gleichungen in der üblichen Form angegeben werden, bei der K durch die Konstante c ausgedrückt ist:

1. Das Gesetz der Geschwindigkeitstransformation:

$$u = \frac{v + u'}{1 + u'v/c^2}. \quad (8.16)$$

2. Die LORENTZGLEICHUNGEN

$$t' = \gamma(v)(t - vx/c^2) \quad (8.17)$$

$$x' = \gamma(v)(x - vt) \quad (8.18)$$

mit

$$\gamma(v) = \frac{1}{\sqrt{1 - v^2/c^2}}. \quad (8.19)$$

Sie wurden 1905 von ALBERT EINSTEIN als grundlegend für die Mechanik erkannt. Diese Leistung wurde von der Bayrischen Staatsregierung durch Aufstellung einer Büste Einsteins in Walhalla gewürdigt (Abb. 33). Walhalla ist die Ruhmeshalle des Deutschen Volkes und bezeichnet gemäß der altgermanischen Mythologie den himmlischen Königssaal Odins..

3. Indem man mit $x_0 = ct$ und $x_1 = x$ Koordinaten gleicher physikalischer Dimension einführt, lassen sich die Lorentzgleichungen in einer symmetrischen Form hinschreiben, die man sich leicht merken kann:

$$\begin{aligned} x'_0 &= \gamma(x_0 - vx_1/c) \\ x'_1 &= \gamma(x_1 - vx_0/c) \end{aligned} \quad (8.20)$$

4. Die Formeln für die Umkehrtransformation, d.h. für die Berechnung der Koordinaten in S aus jenen in S' , lauten:

$$\begin{aligned} t &= \gamma(v)[t' + \frac{v}{c^2}x'] \\ x &= \gamma(v)[x' + vt'] \end{aligned} \quad (8.21)$$

bzw.

$$\begin{aligned} x_0 &= \gamma(v)[x'_0 + vx'_1/c] \\ x_1 &= \gamma(v)[x'_1 + vx'_0/c] \end{aligned} \quad (8.22)$$

8.4. Definition der Längeneinheit Meter

Man hat sich inzwischen darauf geeinigt hat, den Zahlenwert der Strukturkonstanten der Raumzeit nicht ausgehend vom Urmeter zu messen, sondern durch eine Definition festzulegen:

$$c = 299792458 \text{m/s} \quad (8.23)$$

Die Längeneinheit METER

$$1 \text{ Meter} = 1 \text{ m}$$

ergibt sich dann aus diesem definitorischen Wert der Raum/Zeit-Konstanten c und der in allen Galileisystemen identisch definierten Zeiteinheit Sekunde. Zu ihrer Darstellung zieht man das Gesetz der Geschwindigkeitstransformation, Gl. 8.15, heran. Wie man tatsächlich in den Eichbüros experimentell vorgeht, hängt von der Frage ab, mit welchem Messverfahren und mit welchen Teilchen man die kleinste Messunsicherheit erzielen kann. Die Teilchen sollten eine möglichst hohe Geschwindigkeit haben und leicht handhabbar sein, und da sind Photonen selbstverständlich die beste Wahl.

8.5. Invarianz der Ereignisintervalle

Es seien A und B zwei Ereignisse, deren Zeitabstand in S durch $\Delta t = t_B - t_A$ gegeben ist und in einem relativ dazu bewegten System S' durch $\Delta t' = t'_B - t'_A$. Die örtlichen Abstände dieser Ereignisse seien entsprechend $\Delta x = x_B - x_A$ in S und $\Delta x' = x'_B - x'_A$ in S' . Die durch

$$\tau^2 = \Delta t^2 + K \Delta x^2 \quad \text{bzw.} \quad \tau'^2 = \Delta t'^2 + K \Delta x'^2 \quad (8.24)$$

definierten Größen τ bzw. τ' bezeichnet man als EREIGNISINTERVALLE. Mit Gl. 8.10, Gl. 8.11 und Gl. 8.14 kann man leicht nachrechnen, dass

$$\tau = \tau'. \quad (8.25)$$

Das bedeutet, dass Ereignisintervalle in allen Galileisystemen den gleichen Wert haben. Sie ändern sich durch die Transformation nicht, d.h. sie sind invariante bzw. unveränderliche Größen (Invarianten) der Transformation. Diese Feststellung gilt unabhängig davon, welches Vorzeichen oder welchen Wert die Konstante K hat.

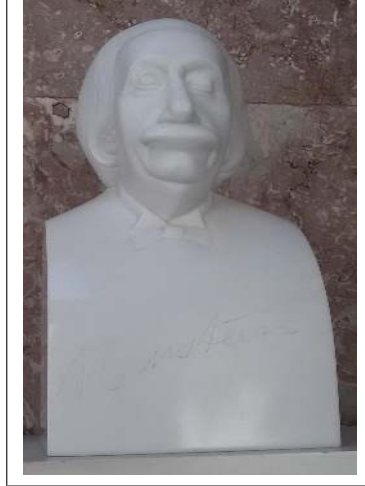


Abbildung 33: Büste Albert Einsteins in der Walhalla bei Regensburg.

8.6. Zeitdilatation

Betrachtet werden zwei Paare von Uhren, die in ihren jeweiligen Ruhesystemen nach dem Einsteinverfahren synchronisiert sind:

1. C_A und C_B seien zwei Uhren, die relativ zueinander ruhen. In ihrem Ruhesystem S sollen sie sich an den Stellen x_A und x_B im Abstand $\ell = x_B - x_A$ befinden.
2. C'_A und C'_B seien zwei Uhren, die relativ zueinander ruhen. In ihrem Ruhesystem S' sollen sie sich an den Stellen x'_A und x'_B im Abstand $\ell' = x'_B - x'_A$ befinden.

Die gestrichenen Uhren mögen sich mit der Geschwindigkeit v relativ zu den ungestrichenen bewegen. Das kann man auch so ausdrücken, dass man sagt, dass sich das Bezugssystem S' mit der Geschwindigkeit v relativ zu S bewegt.

Das Ereignis A bestehe darin, dass C_A und C'_A aufeinandertreffen und das Ereignis B darin, dass C_B und C'_B aufeinandertreffen. Die Zeitdifferenz, die zwischen den Ereignissen A und B festgestellt wird, betrage $\Delta t(\ell)$ bzw. $\Delta t'(\ell')$. Wegen der Invarianz der Ereignisintervalle (Gl. 8.25) gilt mit $K = -c^2$:

$$\Delta t'^2(\ell') - c^2 \ell'^2 = \Delta t(\ell)^2 - c^2 \ell^2 \quad (8.26)$$

Ein Spezialfall von Gl. 8.26 ist derjenige, bei dem die Zeitdifferenz in einem der beiden Bezugssysteme, beispielsweise in S' , auf derselben Uhr $C' = C'_A = C'_B$ abgelesen wird. In Gl. 8.26 ist dann $\ell' = 0$ und $\ell = v\Delta t(\ell)$ einzusetzen, weil die Uhr C' während der Zeitspanne $\Delta t(\ell)$ die Strecke ℓ von C_A zu C_B mit der Geschwindigkeit v durchlaufen muss, damit die Ereignisse in der geschilderten Weise stattfinden können. Man erhält

$$\Delta t(\ell) = \gamma(v)\Delta t'(0). \quad (8.27)$$

Auf der bewegten Uhr laufen Vorgänge bzw. läuft die Zeit langsamer ab als Bezugssystemen, wo die Zeit an den beiden synchronisierten Uhren abgelesen wird: Die Zeit auf der

bewegten Uhr ist dilatiert. Das Phänomen heißt ZEITDILATATION.
 Die Symmetrie ist offensichtlich, d.h. es gilt auch umgekehrt

$$\Delta t'(\ell') = \gamma(v)\Delta t(0). \quad (8.28)$$

Wie der durch die Symmetrie der Zeitdilatation aufgeworfene scheinbare Widerspruch zu lösen ist (Zwillingsparadoxon), soll erst später, nämlich in Kap. C geklärt werden.

Zum besseren Verständnis der Zeitdilatation soll der oben betrachtete Ereignisablauf noch einmal in zwei etwas anderen Formulierungen wiedergegeben werden. Der erste wieder mit Uhren (C =clock):

- Die Uhren C_A und C_B haben einen Abstand ℓ voneinander und sind relativ zueinander in Ruhe. Die Uhr C' bewegt sich mit der Geschwindigkeit v zu C_A bzw. C_B .
- Das Ereignis A bestehe darin, dass die Uhren C_A, C_B und C' gleichzeitig abgelesen werden, wenn auf C_A der Zeitpunkt t_A und auf C' der Zeitpunkt t'_A angezeigt wird. Gleichzeitigkeit für alle drei Uhren lässt sich nur dadurch sicherstellen, dass C_B mit C_A nach dem Einsteinverfahren synchronisiert worden ist und C' sich gerade am gleichen Ort befindet wie C_A .
- Das Ereignis B bestehe darin, dass die drei Uhren wieder gleichzeitig abgelesen werden. Das geht nur, wenn die Uhr C' gerade bei der Kerze C_B eingetroffen ist. Dann zeigt die im Abstand ℓ befindliche Uhr C_B die zu C_A synchronisierte Zeit $t_B(\ell)$ an und C' die Zeit $t'_B(0)$, weil die Ablesung auf Uhren erfolgte, welche den Abstand null voneinander haben, denn es handelt sich um die gleiche Uhr.

Bei der nächsten Formulierung spielen Kerzen die Rolle der Uhren (C =candle):

- Die Kerzen C_A und C_B haben einen Abstand ℓ voneinander und sind relativ zueinander in Ruhe. Die Kerze C' bewegt sich mit der Geschwindigkeit v zu C_A bzw. C_B .
- Das Ereignis A bestehe darin, dass die Kerzen C_A, C_B und C' gleichzeitig angezündet werden. Dies lässt sich realisieren, indem man die Zeitangabe auf Uhren am Ort x_A und x_B zuvor nach dem Einsteinverfahren synchronisiert hat und C' gerade C_A gerade passiert, sich also gerade am gleichen Ort befindet wie C_A .
- Das Ereignis B bestehe darin, dass C' gerade erlischt, wenn es bei der Kerze C_B eintrifft. Nach Gl. 8.27 sind die Kerzen C_A und C_B bereits vorher erloschen, d.h. die transportierte Kerze „lebt länger“ als die ruhenden Kerzen.

9. Analysis (M1)

Bis hier hin wurden lineare Zusammenhänge betrachtet, wie zum Beispiel die geradlinigen gleichförmigen Bewegungen in Kap. 8. Will man darüber hinaus gehen, so muss man sich mit allgemeineren Funktionen des Ortes bzw. der Zeit befassen. In diesem Kapitel werden einige mathematische Hilfsmittel vorgestellt, die man dabei benötigt.

9.1. Ortsableitung

Die ABLEITUNG

$$f'(x_0) = \frac{df}{dx}(x_0) = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \quad (9.1)$$

beschreibt die Änderung $df = f'(x)dx$ einer von der Koordinate x abhängigen Funktion $f(x)$ im Punkt x_0 . Die erste Ableitung nach dem Ort kennzeichnet man in der Physik i.A. durch einen Strich „ $'$ “. Wonach abgeleitet wird, geht gewöhnlich aus dem explizit angegebenen Argument hervor. Wenn jedem Punkt $x = x_0$ der Funktionswert $f'(x_0)$ zugeordnet wird, liegt die Ableitungsfunktion $f'(x)$ vor. Wenn man vom DIFFERENZIEREN EINER FUNKTION spricht, dann meint man, dass man die ABLEITUNG EINER FUNKTION bildet. Man kann leicht nachrechnen, dass

$$(\alpha f(x) + g(x))' = \alpha f'(x) + g'(x), \quad (9.2)$$

wobei α hier eine Konstante sein soll. Für das Produkt zweier Funktionen gilt⁹

$$(fg)' = f'g + fg' \quad (\text{Produktregel}) \quad (9.3)$$

und für die Abhängigkeit einer Funktion f von einer Funktion g gilt¹⁰

$$(f(g(x)))' = f'(g(x))g'(x) \quad (\text{Kettenregel}). \quad (9.4)$$

Beispiel 1: Für die Funktion $f(x) = x^n$ mit einer natürlichen Zahl n ergibt die Produktregel $(x^n)' = (x^{n-1})'x + x^{n-1} = (x^{n-2})'x^2 + x^{n-1} + x^{n-1} = \dots$ Nach n ganzzahligen Schritten erreicht man $(x^0)' = 0$. Folglich gilt $(x^n)' = nx^{n-1}$. Für jede rationale Zahl q gibt es eine ganze Zahl m , so dass qm ganzzahlig ist. Dann folgt durch Anwendung der Kettenregel: $((x^q)^m)' = m(x^q)^{m-1}(x^q)'$ und andererseits ist $((x^q)^m)' = qmx^{qm-1}$. Daraus folgt für jede rationale Zahl q :

$$(x^q)' = qx^{q-1} \quad (9.5)$$

⁹ $(fg)' = \lim_{\Delta x \rightarrow 0} \frac{f(x+\Delta x)g(x+\Delta x) - f(x)g(x)}{\Delta x}$. Wegen $f(x+\Delta x)g(x+\Delta x) - f(x)g(x) = f(x+\Delta x)g(x+\Delta x) - f(x+\Delta x)g(x) + f(x+\Delta x)g(x) - f(x)g(x) = f(x+\Delta x)[g(x+\Delta x) - g(x)] + [f(x+\Delta x) - f(x)]g(x)$ folgt dann $(fg)' = f'g + fg'$.

¹⁰

$$\begin{aligned} f(g(x))' &= \lim_{\Delta x \rightarrow 0} \frac{f(g(x+\Delta x)) - f(g(x))}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \left(\frac{f(g(x+\Delta x)) - f(g(x))}{g(x+\Delta x) - g(x)} \frac{g(x+\Delta x) - g(x)}{\Delta x} \right) = f'(g(x))g'(x) \end{aligned}$$

Funktion	Ableitung	Funktion	Ableitung
$y = \exp(x) = e^x$	$y' = \exp(x)$	$y = \ln x$	$y' = 1/x$
$y = \sin x$	$y' = \cos x$	$y = \cos x$	$y' = -\sin x$
$y = \sinh x$	$y' = \cosh x$	$y = \cosh x$	$y' = \sinh x$

Tabelle 3: Ableitungen der wichtigsten Funktionen

Beispiel 2: In der Gleichung $f/f=1$ kann man die linke und die rechte Gleichungsseite differenzieren. Dies nennt man oft auch IMPLIZITES DIFFERENZIEREN. Mit der Produktregel folgt $(f/f)' = f'(1/f) + f(1/f)' = 0$ und somit

$$\left(\frac{1}{f}\right)' = -\frac{f'}{f^2} \quad (9.6)$$

Beispiel 3: Die Umkehrfunktion der Funktion $\sin x$ wird mit $\sin^{-1} x$ bzw. $\arcsin x$ bezeichnet (Den Kehrwert der Funktion $\cos x$ bezeichnet man hingegen mit $(\cos x)^{-1}$). Es ist also $\sin(\arcsin x) = x$. Durch beider Gleichungsseiten nach der Kettenregel ergibt sich $\cos(\arcsin x) \arcsin' x = 1$. Wegen $\cos(\arcsin x) = \sqrt{1 - \sin^2(\arcsin x)}$ folgt für die Ableitung der arcsin-Funktion:

$$\arcsin'(x) = \frac{1}{\sqrt{1 - x^2}} \quad (9.7)$$

Weitere wichtige Ableitungen sind in Tab. 3 tabelliert. Die Funktion $\sinh x = (e^x - e^{-x})/2$ ist der Sinus Hyperbolicus, die Funktion $\cosh x = (e^x + e^{-x})/2$ der Cosinus Hyperbolicus.

9.2. Zeitableitung

Im Rahmen der KINETIK interessiert man sich für das zeitliche Verhalten einer physikalischen Größe $f(t)$. Wenn die zeitlichen Änderungen nicht gleichförmig erfolgen, benötigt man die Ableitung

$$\dot{f}(t) = \frac{df}{dt} \quad (9.8)$$

Sie beschreibt die Änderung $df = \dot{f}(t)dt$ der Funktion $f(t)$ im Zeitpunkt t . Die erste Ableitungen nach der Zeit kennzeichnet man in der Physik i.A. durch einen übergestellten Punkt „ $\dot{}$ “.

9.3. Stammfunktion

Kennt man die Funktion $f(t)$, so kann man die Ableitung $\dot{f}(t)$ gemäß Gl. 9.8 berechnen. Manchmal kennt man aber die Ableitung $\dot{f}(t)$ und möchte die STAMMFUNKTION

$$f(t) = \int \dot{f} dt + C \quad (9.9)$$

wissen, d.h. jene Funktion $f(t)$, von der die Ableitung $\dot{f}(t)$ stammt. Wenn man genug Erfahrung mit Ableitungen hat, kann man die Stammfunktion manchmal erraten, und wenn nicht, dann schlägt man sie in einer Integralsammlung nach. Die Lösung ist nicht eindeutig: Die freie Integrationskonstante C kann dazu genutzt werden, eine zusätzliche ANFANGSBEDINGUNG festzulegen. In Gl. 9.9 schreibt man für die wegen der freien Konstanten noch unbestimmte Stammfunktion das Symbol $\int \dot{f} dt$ für das unbestimmte Integral.

10. Geschwindigkeit und Beschleunigung

Es werden geradlinige eindimensionale Bewegungen entlang der x -Achse betrachtet. Jedem Zeitpunkt t ist genau ein Ortspunkt x zugeordnet: $t \mapsto x$. Der Bewegungsvorgang wird durch eine Funktion $x(t)$ beschrieben.

10.1. Momentangeschwindigkeit

Die Bahn eines Teilchens ist durch die Angabe $x(t)$ seines Ortes zu jedem Zeitpunkt t beschrieben. Die zeitliche Änderung des Ortes, d.h. seine Geschwindigkeit bzw. MOMENTANGESCHWINDIGKEIT

$$v(t) = \dot{x}(t) = \frac{dx}{dt} \quad (10.1)$$

ergibt sich durch Differentiation nach der Zeit. Manchmal ist die Ableitung $v(t) = \dot{x}(t)$ bekannt und der Weg $x(t)$ als Funktion der Zeit zu bestimmen. Dazu ist die Stammfunktion aufzusuchen. Nehmen wir als Beispiel den Fall einer gleichförmigen Bewegung mit

$$v(t) = v_0, \quad (10.2)$$

d.h. einer Bewegung mit konstanter Geschwindigkeit v_0 . Wenn man die frei wählbare Konstante C in Gl. 9.9 dadurch festlegt, dass man einen bestimmten Ort x_0 zur Zeit $t = 0$ vorschreibt, also $x(0) = x_0$ als Anfangsbedingung verlangt, dann ergibt sich

$$x(t) = \int v dt + C = v_0 t + x_0. \quad (10.3)$$

10.2. Geradlinige beschleunigte Bewegung

Für eine geradlinige Bewegung entlang der x -Achse kann man das zeitliche Verhalten der Geschwindigkeit durch ein Geschwindigkeits-Zeit-Diagramm (bzw. v-t-Diagramm) veranschaulichen. In Abb. 34 sind drei Fälle gezeigt. Der Fall A zeigt eine geradlinig-gleichförmige Bewegung, der Fall B eine gleichförmig beschleunigte Bewegung und der Fall C eine allgemeine beschleunigte Bewegung. Die Größe

$$a(t) = \dot{v}(t) = \ddot{x}(t) \quad (10.4)$$

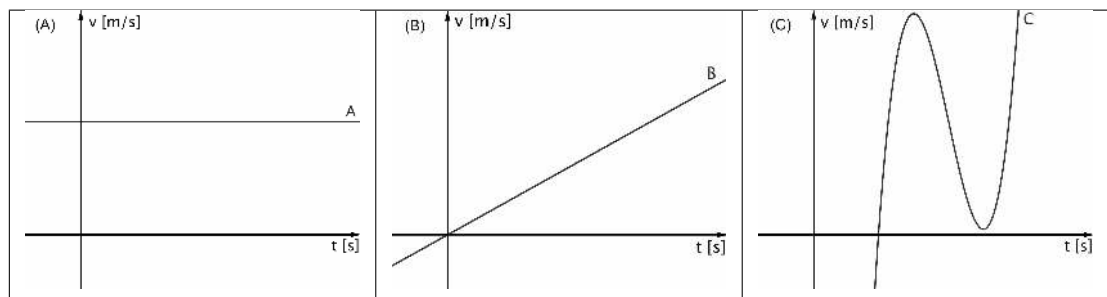


Abbildung 34: Graphische Darstellung von eindimensionalen Bewegungen im v-t-Diagramm: (A) Gleichförmige Bewegung. (B) Gleichförmig beschleunigte Bewegung. (C) Allgemeine beschleunigte Bewegung mit Phasen des Abbremsens und Beschleunigens

nennt man BESCHLEUNIGUNG. Sie ist gleich der ersten Ableitung der Geschwindigkeit und somit gleich der zweiten Ableitung des Ortes nach Zeit. Letzteres wird durch zwei über die Ortskoordinate geschriebene Punkte „ $\ddot{}$ “ kenntlich gemacht. Wenn $a(t)$ konstant ist, spricht man von einer GLEICHFÖRMIGEN BESCHLEUNIGUNG. Dabei nimmt die Geschwindigkeit in gleichen Zeitabschnitten in gleichem Maße zu (Abb. 34B).

10.3. Relativistische Transformation der Beschleunigung

„It used to be argued that it would be necessary to pass to Einstein’s general relativity in order to handle acceleration, but this is completely wrong.“ (Roger Penrose: „The Road to Reality“ (2004))

Zu den kaum auszurottenden Mythen der Physik gehört die Behauptung, dass man die relativistische Beschleunigung erst im Rahmen der Allgemeinen Relativitätstheorie ableiten kann. Der falsche Eindruck hat vermutlich folgenden Grund: Die Allgemeine Relativitätstheorie ist eine Theorie der Gravitation bzw. der Trägheit. Ihre Ausgangsidee ist die Äquivalenz der Wirkung von Gravitation und Beschleunigung. Daher muss man sich dort intensiv mit beschleunigten Bezugssystemen beschäftigen. Wie gleich gezeigt wird, kann man die Gleichung für die Transformation der relativistischen Beschleunigung ohne weiteres im Rahmen der Speziellen Relativitätstheorie herleiten.

Das System S' bewege sich mit der Geschwindigkeit v relativ zu S . In S' bestimmt man die Beschleunigung $a' = \frac{du'}{dt'}$ eines Teilchens durch Differentiation seiner Geschwindigkeit u' nach der Systemzeit t' . Wenn man Gl. 8.16 nach t' differenziert, erhält man für konstante Relativgeschwindigkeit v :

$$u = \frac{v + u'}{1 + u'v/c^2}$$

$$\frac{du}{dt'} = \frac{1 + u'v/c^2 - (v + u')v/c^2}{(1 + u'v/c^2)^2} a' \quad (10.5)$$

Die Uhr in S' ist die relativ zu S bewegte Uhr und daher ist $dt' = dt/\gamma(v)$. Es sei nun S' gerade das momentane Ruhesystem des Teilchens, d.h. es ist $u' = 0$ und das Teilchen bewegt sich mit der Geschwindigkeit $u = v$ relativ zu S' . Setzt man das in Gl. 10.5 ein, so folgt:

$$\gamma(v)a = (1 - v^2/c^2)a'$$

bzw.

$$a(v) = a'(0)/\gamma^3(v). \quad (10.6)$$

Die Beschleunigung eines $a(v)$ in einem Galileisystem S , in dem das Teilchen sich mit der Geschwindigkeit v bewegt, ist um den Faktor $1/\gamma^3(v)$ kleiner als die Beschleunigung $a'(0)$, die es in seinem momentanen Ruhesystem hat, d.h. in dem Bezugssystem, in dem es die Geschwindigkeit 0 hat. Die Beschleunigung fällt um so kleiner aus, je schneller sich ein Teilchen bewegt.

Tatsächlich beobachtet man, dass die Beschleunigung eines Teilchens mit zunehmender Geschwindigkeit abnimmt. Es gibt leider Physikbücher, die dieses experimentelle Faktum vom Newtonschen Standpunkt aus interpretieren. Sie führen die abnehmende Beschleunigung auf eine mit der Geschwindigkeit zunehmende Trägheit bzw. eine dynamische Masse zurück. Das ist falsch. Die Masse ist ein Begriff der Dynamik (s. Teil III). Von Dynamik ist hier noch nicht die Rede. Wie die Herleitung zeigt, ist die Abnahme der Beschleunigung für Bezugssysteme, in denen sich ein Teilchen mit der Geschwindigkeit v bewegt, ein lupenreiner kinematischer Effekt.

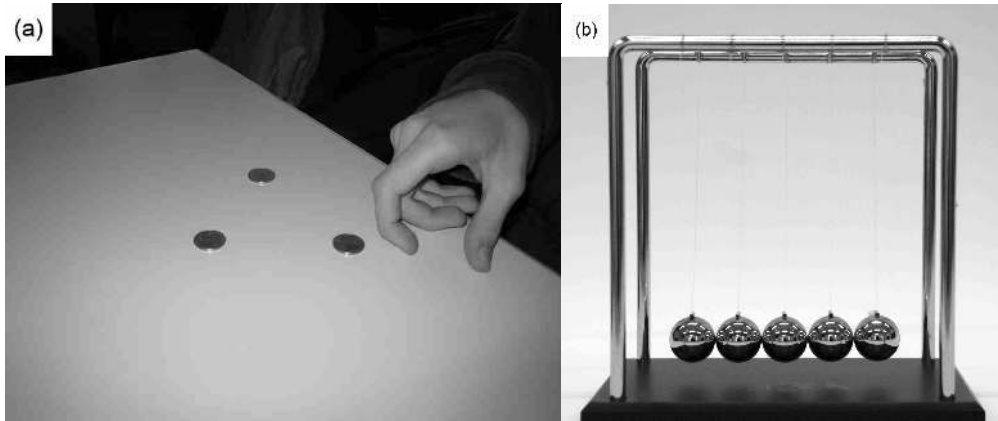


Abbildung 35: (a) Münzen schnippen. (b) Kugelstoßpendel

Teil III. Dynamik

Die Dynamik ist das zweite große Teilgebiet der Mechanik. Sie befasst sich mit der Klärung der Ursachen von Bewegungsänderungen. Um die dabei wesentlichen Konzepte klar herausarbeiten zu können, beschränken wir uns in der Theorie hier zunächst meist auf eindimensionale Bewegungen. Die systematische Erweiterung der theoretischen Konzepte auf Bewegungen im dreidimensionalen Raum ist zwar nicht schwierig, soll aber erst ab Teil V erfolgen.

11. Grundbegriffe der Dynamik

Schnippt man eine Münze gegen eine andere, so dass eine bewegte Münze gegen eine ruhende stößt, so beobachtet man, dass sich der Bewegungszustand der ursprünglich ruhenden Münze beim Stoß ändert, der der ursprünglich bewegten Münze aber auch (Abb. 35). Bei so einem Spiel mit zwei Münzen wird man eine unüberschaubare Vielzahl von unterschiedlichen Vorgängen beobachten, je nachdem, wie groß die Münzen sind, wo sie sich beim Stoß treffen oder welche Geschwindigkeit ihnen mitgegeben wurde. Es scheint aussichtslos, diese unendliche Vielfalt der in der Natur realisierten Fälle durch nur ein, zwei oder drei Grundsätze vollständig beschreiben zu wollen. Und dennoch – es geht!

Wie kommt man zu den genial einfachen Gesetzmäßigkeiten, mit denen man solche Stoßexperimente verstehen kann? Das ist nicht ganz einfach, und es führt auch kein geradliniger Weg zu solchen Gesetzen. Es ist vielmehr ein kreativer Prozess, der von vielerlei experimenteller Vorerfahrungen ausgeht. Es können Erfahrungen mit Stoßprozessen auf einer Luftkissenbahn, beim Billardspielen oder mit dem Kugelstoßpendel (Abb. 35b) sein,

die schließlich die erfolgreichen Ideen und Begriffe der Dynamik reifen lassen. Wir wollen nun einige Begriffe definieren, die man aus den erwähnten spielerischen Experimenten extrahieren kann und die in den folgenden Kapiteln eine Rolle spielen.

Substanz und Zustand. Durch die Experimente macht man zunächst einmal die Erfahrung, dass es etwas gibt, das sich – jedenfalls nach Augenschein – nicht ändert: die Münzen erscheinen uns vor und nach dem Stoß gleich, die Billardkugeln und die Körper auf der Luftkissenbahn auch. Das, was uns bei einem Stoß unverändert zu bleiben erscheint, fasst man unter den Begriff der SUBSTANZ. Und dann gibt es etwas, das sich nach unserer Wahrnehmung ändert: die Objekte bewegen sich nach dem Stoß anders als vorher. Das was nicht substanzuell und unveränderlich ist, sich also ändern kann, fällt unter den Begriff des ZUSTANDS. Zur Charakterisierung des Zustands (in den geschilderten Experimenten hier: des Bewegungszustands) kann man erst einmal die Geschwindigkeit heranziehen. Die Geschwindigkeitsänderungen der Objekte bzw. Substanzen stellen Zustandsänderungen dar. Indem man einem Objekt bzw. einer Substanz eine Geschwindigkeit als Etikett anheften, ist der Zustand der Substanz charakterisiert (jedenfalls was seine Bewegung anbetrifft). Die Geschwindigkeit ist hier ein Beispiel für eine Zustandsvariable bzw. ZUSTANDSGRÖSSE.

Wechselwirkung. Wenn man eine Münze A auf eine zunächst relativ zur Tischplatte ruhende Münze B schnippt, so bleibt B so lange in Ruhe, bis die beiden Münzen miteinander in Kontakt kommen: Sie bleibt in Ruhe, bis eine Ursache hinzukommt, welche die Geschwindigkeit ändert, nämlich der Stoß. Beim Stoß ändert sich aber auch die Geschwindigkeit von A. Das habe ich bis hier hin aus der Perspektive eines Bezugssystems betrachtet, in dem B ursprünglich ruhte bzw. die Tischplatte ruht.

Man kann sich den Vorgang aber genauso gut von einem Bezugssystem aus betrachtet denken, in dem A ruhen würde, wenn es seine Bahn ohne einen Zusammenstoß mit B verfolgen würde. Relativ zu diesem Bezugssystem bewegt sich B auf A zu. Münze A ändert so lange ihre Geschwindigkeit nicht, bis die Münzen in Kontakt kommen. Dann ändern wiederum beide Münzen ihre Geschwindigkeit.

Die Beobachtungen kann man zu folgender Aussage verallgemeinern: Wenn ein Objekt ein anderes beeinflusst, so wird es auch umgekehrt vom anderen Objekt beeinflusst. Objekte üben also eine wechselseitige Wirkung bzw. WECHSELWIRKUNG aufeinander aus. So lange keine Wechselwirkung eintritt, bleibt die Geschwindigkeit eines Objektes relativ zum gewählten Bezugssystem unverändert.

Man mag einwenden, dass sich aus der ersten Perspektive die Bewegung der Münze A ja bereits vorher ständig ändert. Die Geschwindigkeit von A nimmt relativ zur Tischplatte ja bereits vor dem Stoß ständig ab. Das lässt sich aber leicht dadurch erklären, dass es bereits ständige Wechselwirkungen gibt, nämlich zwischen Münze A und Tischplatte, Luft usw., welche durch die Fokussierung unserer Betrachtung auf ein Modell, in dem wir nur die beiden Münzen alleine im Blick hatten, nicht berücksichtigt sind.

Wenn man also tatsächlich eine experimentelle Situation herbeiführen möchte, bei dem ausschließlich die zwei Münzen miteinander wechselwirken, so muss man also erst einmal

dafür sorgen, dass alle anderen Wechselwirkungsmöglichkeiten ausgeschlossen werden, was unter irdischen Laborbedingungen i.A. schwierig ist. Man müsste das Experiment z.B. im Vakuum ablaufen lassen, d.h. in einem Gebiet, in dem sich sonst keine Materie befindet, und zweitens wäre es ganz hübsch, wenn man sich nicht mit den Komplikationen durch langreichweitige Wechselwirkungen (gravitativ, elektrisch, magnetisch) herumplagen muss. Im Gegensatz zur landläufigen Meinung haben Physiker jedoch sehr viel Phantasie. Sie können sich vorstellen, dass es nicht wirklich nötig ist, den idealen Fall tatsächlich zu realisieren, um zu einer allgemeinen Erkenntnis vorzustoßen. Es genügen ihnen Experimente, bei denen man die Wechselwirkung eines Körpers mit seiner Umgebung immer weiter reduziert. Die Extrapolation der experimentellen Erfahrung läuft dabei auf folgende Aussage zu:

Ein Körper, der keinerlei Wechselwirkung erfährt, ändert seine Geschwindigkeit nicht.

Was Physiker in der Überzeugung bestärkt, dass sie damit ein allgemeingültiges physikalisches Gesetz erkannt haben, ist nicht zuletzt folgende Schlussfolgerung: Wenn diese Aussage zur Dynamik in irgendeinem Galileisystem gilt, dann gilt sie offensichtlich auch in jedem Galileisystem der gleichen Klasse. Sie ist im Rahmen der Relativitätstheorie somit widerspruchsfrei.

Stoß. Im Folgenden soll die grundsätzliche Physik der Wechselwirkungsvorgänge anhand von Stoßprozessen herausgearbeitet werden. Unter einem STOSS versteht man eine Wechselwirkung, die zeitlich und räumlich begrenzt stattfindet. Im idealen gedanklichen Modell geht sie instantan und lokal vonstatten. Wenn man sich zwei Münzen immer kleiner denkt und schließlich auf einen Punkt schrumpfen lässt sowie die Wechselwirkung mit der Umgebung wegdenkt, so dass es genügt, sich ausschließlich mit der Wechselwirkung der beiden Münzen zu befassen, dann liegt jenes Modell für den idealen Stoß vor, das im Folgenden theoretisch weiter analysiert werden soll. Bezüglich irgendeines Bezugssystems bewegen sich beide Münzen vor der Wechselwirkung mit einer konstanten Geschwindigkeit und nach dem Stoßprozess ebenfalls. Die Geschwindigkeiten sind nach der Wechselwirkung jedoch andere als davor.

Teilchen und physikalische Systeme. Ein System ist ein aus Einzelteilen bestehendes Gesamtes. Ein Bezugssystem besteht beispielsweise aus seinen Bezugselementen und ein optisches System aus optischen Komponenten. Ein PHYSIKALISCHES SYSTEM bezeichnet entsprechend ein physikalisches Objekt, das sich aus Einzelteilen zusammensetzt. Im Folgenden werden wir diese oft als als TEILCHEN bezeichnen. Jedes Teilchen kann wiederum selbst teilbar sein, d.h. selbst ein physikalisches System darstellen, das sich seinerseits aus („kleineren“) Teilchen zusammensetzt. Umgekehrt kann ein physikalisches System selbst wiederum Teilchen eines („größeren“) physikalischen Systems sein.

Wenn man Objekte untersucht, welche miteinander wechselwirken, z.B. durch einen Stoß, dann haben sie unweigerlich miteinander zu tun. Das sagt schon der Begriff der Wechselwirkung aus. Wenn man die Natur der Wechselwirkung analysieren will, so muss man alle wechselwirkenden Objekte als Teile bzw. Teilchen eines gemeinsamen physikalischen Systems betrachten.

11.1. Extensive Zustandsgrößen und Erhaltungsgrößen

Es liege ein physikalisches System G vor, das aus N zunächst unabhängigen Teilchen bestehe, die hier mit $j = 1, \dots, N$ durchnummeriert sein sollen. Die physikalische Größe \mathcal{B} sei eine Zustandsgröße des Systems, d.h. eine physikalische Größe, welche die Zustände aller N Teilchen charakterisiert. Die Werte, die man für die N unabhängigen Teilchen misst, seien $\mathcal{B}^{(1)}, \dots, \mathcal{B}^{(N)}$. Beispielsweise können die relativ zu einem Galileisystem gemessenen Geschwindigkeiten $v^{(1)}, \dots, v^{(N)}$ der Teilchen als Charakteristika ihres Zustands angesehen werden. Damit ist die Geschwindigkeit v der Teilchen ein Beispiel für eine Zustandsvariable. Wie sich im Folgenden zeigen wird, gibt es auch andere Größen, die man als Zustandsvariable heranziehen kann. Einige haben Eigenschaften, welche sie für die physikalische Beschreibung physikalischer Systeme besonders wichtig machen. In diesem Kapitel wollen wir die Eigenschaften solcher Zustandsvariablen besprechen und zunächst die relevanten Begriffe und Bezeichnungen klarstellen.

Extensive (additive) Zustandsgrößen. Welcher Wert $\mathcal{B}^{(G)}$ ist dem Gesamtsystem G zuzuordnen? Wenn also beispielsweise die Geschwindigkeit der einzelnen Teilchen $v^{(1)}, \dots, v^{(N)}$ ist, mit welcher Geschwindigkeit $v^{(G)}$ bewegt sich dann das Gesamtsystem? Letzteres können wir erst einmal nur beantworten, wenn sich alle Teilchen mit der gleichen Geschwindigkeit bewegen, denn dann ist die Geschwindigkeit des physikalischen Systems trivialerweise $v^{(G)} = v^{(1)} = \dots = v^{(N)}$. Für den allgemeinen Fall, wo die einzelnen unterschiedlichen Objekte (Teilchen) des physikalischen Systems sich mit unterschiedlichen Geschwindigkeiten bewegen, kann man erst einmal gar keine Aussagen machen.

Aus diesem Grund kommt den EXTENSIVEN ZUSTANDSGRÖSSEN eine besondere Bedeutung zu. Sie sind durch die Eigenschaft gekennzeichnet, dass der Wert $\mathcal{B}^{(G)}$ des Gesamtsystems gleich der Summe seiner Teilsysteme ist, also

$$\mathcal{B}^{(G)} = \sum_{j=1}^N \mathcal{B}^{(j)}. \quad (11.1)$$

Gl. 11.1 definiert, was eine extensive (bzw. additive) Zustandsvariable ist. Sie definiert den Begriff, sagt aber nicht, ob eine solche Zustandsvariable existiert oder wie sie gemessen werden kann. Der Vorteil extensiver Zustandsvariablen ist aber offensichtlich, denn man kann stets sagen, welcher Wert dem Gesamtsystem zukommt, wenn man die Werte der Teilsysteme kennt. Für die Geschwindigkeit ist das, wie gerade besprochen, nicht der Fall. Folglich ist sie keine extensive Zustandsvariable.

Wenn die Teilchen miteinander wechselwirken, werden die Teilchen anschließend andere Zustände einnehmen und folglich werden sich die Werte $\mathcal{B}^{(1)}, \dots, \mathcal{B}^{(N)}$ der Zustandsvariablen \mathcal{B} i.A. ändern. Ist die Wechselwirkung ein Stoß, so sind die Teilchen nach dem Stoß wieder unabhängig voneinander und können wieder durch die Werte der Zustandsgröße \mathcal{B} charakterisiert werden. Um die Zustandsgrößen vor und nach dem Stoß zu unterscheiden, werden im Folgenden die Größen nach dem Stoß durch eine übergestellte Tilde ($\tilde{}$) gekennzeichnet. Handelt es sich um eine extensive Zustandsvariable, so gilt nach dem Stoß also per Definitionem für den Wert $\tilde{\mathcal{B}}^{(G)}$ des Gesamtsystems G

$$\tilde{\mathcal{B}}^{(G)} = \sum_{j=1}^N \tilde{\mathcal{B}}^{(j)}. \quad (11.2)$$

Erhaltungsgrößen. Eine ERHALTUNGSGRÖSSE liegt dann vor, wenn bei einer Wechselwirkung stets

$$\mathcal{B}^{(G)} = \tilde{\mathcal{B}}^{(G)} \quad (11.3)$$

gilt, d.h. wenn die das Gesamtsystem beschreibende Zustandsgröße vor und nach einem Stoß stets gleich bleibt. Gl. 11.3 wird auch als ERHALTUNGSAUSSAGE bezeichnet.

Additive (extensive) Erhaltungsgrößen. Gl. 11.3 macht nur dann einen Sinn, wenn man weiß, wie man aus den Zustandswerten der Teilchen des Systems diejenige des Gesamtsystems vor und nach dem Stoß ermitteln kann. Für eine additive bzw. extensive Erhaltungsgröße ist das allerdings klar: Es ist die Summe. Für eine additive Erhaltungsgröße muss daher die Bilanzgleichung

$$\mathcal{B}^{(G)} = \sum_{j=1}^N \mathcal{B}^{(j)} = \sum_{j=1}^N \tilde{\mathcal{B}}^{(j)} = \tilde{\mathcal{B}}^{(G)} \quad (11.4)$$

gelten.

Die Frage ist, ob physikalische Gesetze in Form von Bilanzgleichungen wie Gl. 11.4 formuliert werden können, ohne in einen logischen Widerspruch zu den Relativitätsprinzipien zu geraten. Das läuft auf folgende Forderung hinaus: Wenn ein solches Gesetz in irgendeinem Bezugssystem formuliert werden kann, so muss es auch in jedem physikalisch äquivalenten Bezugssystem in gleicher Form aufgestellt werden können. Wir wollen nun die Konsequenzen analysieren, die sich daraus ergeben:

Nehmen wir an, S sei ein Bezugssystem, in dem Gl. 11.4 gilt und S' ein dazu äquivalentes Bezugssystem. Ferner sei T ein Gesetz bzw. eine Transformation, durch welche eine Größe \mathcal{B} beschrieben im Bezugssystem S in die Größe \mathcal{B}' in einem zu S äquivalenten Bezugssystem S' transformiert bzw. umrechnet werden kann. Für jeden Beitrag in Gl. 11.4 soll also gelten $\mathcal{B}^{(j)'} = T(\mathcal{B}^{(j)})$ und auch für die Summe soll gelten $\mathcal{B}^{(G)'} = T(\mathcal{B}^{(G)})$,

und zwar mit dem gleichen (!) Transformationsgesetz T . Die entscheidende Forderung ist dann, dass in S' eine Bilanzgleichung

$$\mathcal{B}^{(G)'} = \sum_{j=1}^N \mathcal{B}^{(j)'} = \sum_{j=1}^N \tilde{\mathcal{B}}^{(j)'} = \tilde{\mathcal{B}}^{(G)'} \quad (11.5)$$

gleicher Form herauskommen muss. Die Forderung läuft also darauf hinaus, dass die Transformationsgesetz (von S nach S') für additive Erhaltungsgrößen die Gleichung

$$T(\mathcal{B}^{(G)}) = T\left(\sum_{j=1}^N \mathcal{B}^{(j)}\right) = \sum_{j=1}^N T(\mathcal{B}^{(j)}) = \sum_{j=1}^N T(\tilde{\mathcal{B}}^{(j)}) = T\left(\sum_{j=1}^N \tilde{\mathcal{B}}^{(j)}\right) = T(\tilde{\mathcal{B}}^{(G)}) \quad (11.6)$$

erfüllen muss. Das ist tatsächlich möglich, schränkt die Transformation T aber auf homogene lineare Transformation ein (s. Kap. 38.3).

Eine Bilanzgleichung kann nur dann ein allgemeines physikalisches Gesetz sein, wenn sich die Erhaltungsgrößen beim Übergang zwischen äquivalenten Bezugssystemen linear transformieren!

Unsere Überlegungen stellen damit einen engen Zusammenhang zwischen Transformationsgesetzen für äquivalente Bezugssysteme und Erhaltungsgrößen her.

In der Praxis treten manchmal Probleme auf, die ich Ihnen nicht verschweigen möchte: Zum Beispiel kann es sein, dass sich einige an einer Wechselwirkung beteiligten Teilsysteme oder Teilchen der Beobachtung entzogen haben. Dann sieht das so aus als ob die Erhaltungsaussage verletzt wäre (d.h. nicht zuträfe). Diese Situation trat historisch z.B. beim β -Zerfall auf. Bei der Analyse des β -Spektrums sah es so aus, als ob die Energieerhaltung verletzt wäre. Wolfgang Pauli schlug zur Rettung der Energieerhaltung vor, dass ein damals noch nie beobachtetes unbekanntes Teilchen existieren müsse. Dieses „Geisterteilchen“ wird heute Neutrino genannt. Er postulierte, dass dieses Teilchen allen damaligen Nachweismethoden der Physiker entgangen sei und wettete sogar um eine Kiste Sekt, dass es nie gelingen würde, dieses Teilchen nachzuweisen. 26 Jahre später verlor er die Wette [6]. Sie können einen Erhaltungssatz in der Form von Gl. 11.4 daher nur dann auswerten, wenn alle an der Wechselwirkung beteiligten Partner bekannt bzw. identifizierbar sind. Man sagt dann, dass ein ABGESCHLOSSENES SYSTEM vorliegt. Nur für ein abgeschlossenes System ist zu erwarten, dass ein Erhaltungssatz wie Gl. 11.4 gilt. Bei einem OFFENEN SYSTEM ist die Bilanzgleichung nun einmal leider nicht vollständig. Machen wir uns die Konsequenzen an einigen einfachen Beispielen klar:

1. Wenn in der Bilanzgleichung, Gl. 11.4, der N -te Summand fehlt, weil man ihn nicht ermitteln kann, dann gilt im allgemeinen

$$\sum_{j=1}^{N-1} \mathcal{B}^{(j)} \neq \sum_{j=1}^{N-1} \tilde{\mathcal{B}}^{(j)},$$

denn es ist i.A. ja auch

$$\mathcal{B}^{(N)} \neq \tilde{\mathcal{B}}^{(N)}$$

2. Wenn bei einer Wechselwirkung

$$\mathcal{B}^{(N)} = \tilde{\mathcal{B}}^{(N)}$$

gelten würde, dann wäre das N -te Teilchen an der Wechselwirkung nicht beteiligt, eben weil die Größe $\tilde{\mathcal{B}}^{(N)}$ am Ende den gleichen Wert wie am Anfang hätte. Folglich wäre es zulässig gewesen, auch das aus $N - 1$ Teilchen bestehende System als abgeschlossen zu betrachten. Das tun Physiker dann auch und streben der besseren Durchschaubarkeit wegen an, ein abgeschlossenes System allein auf die wirklich beteiligten und somit relevanten „Spielpartner“ zu reduzieren.

Der Sachverhalt ist eigentlich ganz einfach zu verstehen: Eine vollständige Fußballmannschaft besteht aus 11 Spielern. Das wäre das abgeschlossene System. Wenn einer der Spieler sich entschließt, sich lieber an den Spielfeldrand zu setzen, um mit Bekannten zu plaudern, dann genügt es, die 10 spielenden Spieler zu betrachten. Wenn einer von den 11 Spielern so schlecht spielt, dass er in 90 Minuten den Ball nicht auf den Fuß kriegt, dann trägt er näherungsweise nichts bei, und es ist ungefähr so, wie wenn effektiv 10 Spieler gespielt hätten. Diese 10 Spieler bilden dann näherungsweise das abgeschlossene System, sind also näherungsweise bereits die vollständige Mannschaft.

3. Wenn eine Wechselwirkung unendliche Reichweite hat (wie z.B. die Gravitation), dann kann man streng genommen nie davon sprechen, dass ein abgeschlossenes System vorliegt. Man kann sich beispielsweise streng genommen nicht darauf beschränken, alleine die Sonne und ihre Planeten als ein vollständiges System zu betrachten, weil selbst irgendein Staubteilchen der entferntesten Galaxie in Wechselwirkung mit jedem dieser Himmelskörper steht. In diesem Falle hilft die Einsicht, dass der Einfluss $\mathcal{B}^{(N)}$ eines sehr weit entfernten Objekts nur sehr gering ist und man ihn im Rahmen der Messunsicherheit vernachlässigen darf. Es liegt dann ein quasi-abgeschlossenes System vor, bei dem man von

$$\sum_{j=1}^{N-1} \mathcal{B}^{(j)} \approx \sum_{j=1}^{N-1} \tilde{\mathcal{B}}^{(j)}$$

auf

$$\sum_{j=1}^{N-1} \mathcal{B}^{(j)} = \sum_{j=1}^{N-1} \tilde{\mathcal{B}}^{(j)} \tag{11.7}$$

idealisieren darf.

Die Methode der Naturbeschreibung durch Erhaltungsaussagen (bzw. Erhaltungssätze) ist genau deshalb so erfolgreich, weil es gewöhnlich genügt, nur eine beschränkte Zahl von Objekten und Beiträgen zu betrachten, die zusammen ein quasi-abgeschlossenes System darstellen, für welches Gl. 11.7 gilt.¹¹ Als Physiker muss man lernen zu erkennen,

¹¹Dazu muss für alle „externen“ Beiträge, die hier durch den Beitrag $\mathcal{B}^{(N)}$ zusammengefasst sind, gelten, dass $\mathcal{B}^{(N)} \ll \mathcal{B}^{(j)}$ für alle $j = 1, \dots, N - 1$.

ob ein Beitrag wichtig oder unwichtig ist. Kurzum, man muss ein Gefühl dafür entwickeln, wann man nähern darf und wann nicht oder wie man einen komplexen Vorgang so weit vereinfachen darf, dass er mathematisch-logisch aufgeklärt werden kann, ohne dass interessierende bzw. wesentlichen Aspekte verloren gehen. Das ist eher eine Kunst als ein Handwerk, und daher macht, wie beim Klavierspielen, erst die Übung den Meister. Manchmal geht es jedoch nicht um die Reduktion der Bilanzgleichung, sondern es geht umgekehrt um die Vervollständigung eines Systems und damit um die Vervollständigung der Bilanzgleichung durch summarische Hinzunahme von Beiträgen wie das später beispielsweise für die Thermodynamik vorgeführt wird (Teil IV).

11.2. Konstruktion der grundlegenden Erhaltungsgröße der Mechanik

Für Galileisysteme sind die linearen Transformationen, mit denen sich die Ereigniskoordinaten $(x_0, x_1) = (ct, x_1)$ transformieren, bereits bekannt. Man dreht nun gewissermaßen den Spieß um, geht von diesen Transformationen aus und konstruiert die dazu passenden Erhaltungsgrößen: Man postuliert, dass es zwei physikalische Größen p_0 und p_1 gibt, von denen erstere genauso wie x_0 von einer Rauminversion nicht betroffen ist, also ein Skalar ist, und p_1 wie x_1 bei einer Rauminversion sein Vorzeichen wechselt (also $p_1 \rightarrow -p_1$ bei einer Inversion) und somit eine vektorielle Größe ist. Ferner verlangt man, dass sich p_0 und p_1 gemäß

$$p'_0 = \gamma(p_0 - vp_1/c) \quad (11.8)$$

$$p'_1 = \gamma(p_1 - vp_0/c) \quad (11.9)$$

transformieren, wenn man von einem System S zu einem System S' umrechnet, das sich mit der Geschwindigkeit v relativ zu S bewegt. Für die Umkehrtransformation gilt

$$\begin{aligned} p_0 &= \gamma(p'_0 + vp'_1/c) \\ p_1 &= \gamma(p'_1 + vp'_0/c) \end{aligned} \quad (11.10)$$

Gl. 11.8 stimmt mit Gl. 8.20 formal überein: die Transformation von (p_0, p_1) ist wie gewünscht die gleiche wie diejenige der Ereigniskoordinaten. Damit erfüllt die Größe (p_0, p_1) die Forderung, die mit Gl. 11.6 erhoben wurde. Folglich ist sie eine additive (bzw. extensive) Erhaltungsgröße der Physik.

Analog zu Gl. 8.24 kann man nachrechnen, dass

$$p_0^2 - p_1^2 = p_0'^2 - p_1'^2 = P^2 \quad (11.11)$$

nicht nur eine Invariante der Rauminversion ist (was die Skalar-Vektor-Kombination überhaupt möglich macht), sondern auch eine Invariante der Lorentztransformation. Der invariante Skalar P^2 ist ein vom Bezugssystem unabhängiges Charakteristikum des betrachteten physikalischen Systems bzw. des physikalischen Objekts.

Gemäß Gl. 11.11 muss $|p_0| \geq |p_1|$ gelten. Für den Fall $|p_0| = |p_1|$ ist $P = 0$. Dieser Grenzfall soll erst später diskutiert werden.

Für den Fall $|p_0| > |p_1|$ kann aus Gl. 11.9 geschlossen werden, dass es einen Wert $-1 < v/c < 1$ gibt, für den $p'_1 = 0$ gilt. Das bedeutet, dass es ein Galileisystem S' gibt, wo

$$p'_0 = P \quad (11.12)$$

$$p'_1 = 0 \quad (11.13)$$

In einem Galileisystem S , das sich mit der Geschwindigkeit v relativ zu diesem Bezugssystem S' entlang der x -Achse bewegt, ergibt sich gemäß Gl. 11.10

$$p_0 = P \gamma(v) \quad (11.14)$$

$$p_1 = (P/c) v \gamma(v) \quad (11.15)$$

Geschwindigkeit ist ein relativer Begriff. Ist in einem Bezugssystem S die Geschwindigkeit $v = 0$, dann ruht das physikalische Objekt relativ zu S . Bewegt es sich relativ zu S , so ist seine Geschwindigkeit gemäß den Gln. 11.12 und 11.13 durch

$$v/c = p_1/p_0 \quad (11.16)$$

gegeben.

11.3. Einheitenkonversionen mit Naturkonstanten

Die Raum/Zeit-Konstante c , die Elementarladung e oder die Dirac-Konstante \hbar sind Naturkonstanten. Sie sind natürliche Invarianten, d.h. sie haben in jedem äquivalenten Bezugssystem (bei gleicher physikalischer Einheit) den gleichen Zahlenwert. Daher kann man sie genauso gut von Anfang an mit einem Zahlenwert und einer Einheit definieren, wie das bei der Raum/Zeit-Konstanten ja auch gemacht wird (s. Kap. 8.4). Naturkonstanten stellen das natürliche Einheitensystem der Physik dar.

Man kann jede physikalische Größe mit beliebigen Produkten oder Potenzen dieser Naturkonstanten multiplizieren und so Größen definieren, die zwar andere Einheiten und Zahlenwerte haben, aber im Prinzip physikalisch gleichbedeutend sind mit den ursprünglichen Größen. In bestimmten Fällen gibt man den in andere Einheiten konvertierten Größen (oder auch den anderen Einheiten selbst) eigene Namen. Ich möchte solche Einheitenkonversionen nun an zwei Beispielen erläutern.

1. Den in Gl. 11.15 auftretenden Quotienten

$$m = P/c \quad (11.17)$$

der Invarianten P und der Naturkonstanten c bezeichnet man als die MASSE. Diese ist folglich genauso wie P ein invarianter Skalar. Invariant bedeutet hier lorentzinvariant, d.h. die Masse hat in allen Galileisystemen der gleichen Galileiklasse den gleichen Wert. Skalar meint hier, dass die Masse kein Vektor und auch keine Komponente eines Vektors ist. Auch bei Inversionen oder Drehungen des Raumes hat sie stets den gleichen Wert. Ihre Einheit, das KILOGRAMM,

$$[m] = 1 \text{ Kilogramm} = 1 \text{ kg},$$

als eine der Basiseinheiten des SI-Systems ausgewählt. Wie man diese Einheit darstellt und die Masse korrekt misst, wird im nächsten Abschnitt 11.4 erläutert.

2. Wenn man das im System S gegebene $p_0(v)$ mit c multipliziert, dann bezeichnet man die Größe

$$E(v) = cp_0(v) \quad (11.18)$$

als ENERGIE. Sie wird in der SI-Einheit JOULE angegeben:

$$[E] = 1 \text{ Joule} = 1 \text{ J} = 1 \text{ kg m}^2/\text{s}^2$$

Mit den neuen Bezeichnungen kann man Gl. 11.14 und Gl. 11.15 in der Form

$$E = cp_0(v) = mc^2 \gamma(v) = \frac{mc^2}{\sqrt{1 - v^2/c^2}} \quad (\text{Energie}) \quad (11.19)$$

$$p = p_1 = mv \gamma(v) = \frac{mv}{\sqrt{1 - v^2/c^2}} \quad (\text{Impuls}) \quad (11.20)$$

aufschreiben. Durch Einsetzen in Gl. 11.11 erhält man mit

$$E^2 - p^2 c^2 = m^2 c^4 \quad (11.21)$$

bzw.

$$E(p) = \sqrt{m^2 c^4 + p^2 c^2}. \quad (11.22)$$

einen allgemeinen Zusammenhang zwischen Energie, Impuls und Masse eines Objekts.

11.4. Wie misst man die Masse?

Bis hier hin wurden einige Beziehungen und Definitionen für hypothetische Erhaltungsgrößen dargestellt. Das ist noch keine Physik. Physik wird es erst, wenn man aufgrund der bis hier hin entwickelten Theorie ein Messverfahren für die Erhaltungsgrößen angeben kann. Das soll jetzt geschehen, indem angegeben wird, wie man den einzigen noch unbestimmten Parameter der Theorie messen kann, nämlich die systemcharakteristische Größe P bzw. die Masse $m = P/c$. Um etwas zu messen, braucht man prinzipiell zweierlei: Erstens die Festlegung einer Einheit und zweitens eines Vergleichsverfahrens (s. Kap. 1.3):

1. Die Einheit der Masse wird definiert, indem sie durch einen Prototypen dargestellt wird, dem URKILOGRAMM. Dieser Prototyp wird im Tresor des Internationalen Büros für Maß und Gewicht bei Paris verwahrt, damit diese Realisierung der Masseneinheit auch für künftige Generationen zur Verfügung steht. Das Kilogramm ist übrigens die letzte SI-Maßeinheit, die man noch durch einen solchen Prototypen darstellen muss. Aber es gibt bereits intensive Diskussionen darüber, wie man auch für die Masseneinheit zu einer fundamentalen Definition übergehen könnte (S. Anhang B).

2. Die Herstellung von Kopien des Urkilogramms ist kein prinzipielles Problem. Man braucht nur einen Körper aus dem gleichen Material mit den gleichen Abmessungen nach dem gleichen Herstellungsverfahren zu realisieren. Dann sind die Massen der Kopien selbstverständlich gleich dem des Urkilogramms. Das Problem, wie man zwei Massen miteinander vergleichen kann, stellt sich erst, wenn die Massen zweier Körper verglichen werden müssen, die aus unterschiedlichem Material bestehen bzw. aus unterschiedlicher Herstellung bzw. Formgebung stammen. Sie sind dann nicht gleich und folglich muss auch ihre Masse nicht gleich sein. Daher muss für die Messung der Masse ein Vergleichsverfahren definiert werden. Im Folgenden geht es nicht um die technische Frage, wie der Massenvergleich tatsächlich durchgeführt wird, sondern um die prinzipielle Definition des Vergleichsverfahrens, mit dem die Masse gemessen wird. Das Messprinzip muss gewährleisten, dass man für jedes äquivalente Galileisystem die gleiche Masse herausbekommt, denn die Masse ist gemäß der Relativitätstheorie eine invariante Größe. Das leistet eine Messung nach folgendem Vergleichsverfahren:

Wir lassen einen Körpers A mit der Masse m_A , der die Maß-Darstellung der Masse repräsentiert (Urkilogramm), mit einem Körper B wechselwirken, dessen Masse m_B ermittelt werden soll. Die Wechselwirkung sei ein elastischer Stoßprozess (s. Kap. 12.2), bei dem die Massen der beiden Körper sich durch den Stoß nicht ändern. Aus der Erhaltungsgleichung für die Energie folgt in diesem Fall:

$$m_A \gamma(v_A) + m_B \gamma(v_B) = m_A \gamma(\tilde{v}_A) + m_B \gamma(\tilde{v}_B).$$

Wenn die Massen gerade gleich groß sind (also $m_A = m_B$), folgt daraus

$$\gamma(v_A) + \gamma(v_B) = \gamma(\tilde{v}_A) + \gamma(\tilde{v}_B).$$

Das hat zwei allgemeine Lösungen, nämlich $v_A = \tilde{v}_A$ und $v_B = \tilde{v}_B$ oder $v_A = \tilde{v}_B$ und $v_B = \tilde{v}_A$. Der erste Fall ist uninteressant, denn da hat keine Wechselwirkung stattgefunden. Da i.A. $v_A \neq v_B$, kommt man für den zweiten Fall zum Schluss:

Zwei Körper A und B besitzen genau dann gleiche Masse, wenn für einen elastischen Zweikörperstoß gilt: $v_A = \tilde{v}_B$ und $v_B = \tilde{v}_A$.

Die Voraussetzung, dass die Wechselwirkung elastisch ist, dass also die Massen vor und nach dem Stoß gleich sind, bringt keine prinzipielle Schwierigkeit mit sich, denn ob sie erfüllt ist, lässt sich prüfen. Das Messverfahren lässt sich messtechnisch i.A. so gestalten, dass der Einfluss eines beim Stoß auftretenden Massendefekts (s. Kap. 12.2) vernachlässigt werden kann.

Der wichtige Punkt der Definition ist, dass die Massenbestimmung unabhängig vom verwendeten Galileisystem (derselben Klasse) ist, denn sie beruht nicht auf dem Zahlenwert der Geschwindigkeiten, sondern nur auf der Feststellung der Gleichheit von Geschwindigkeiten. Die so definierte Masse hängt somit tatsächlich nicht von der Geschwindigkeit ab und ist folglich keine relative (vom gewählten Galileisystem abhängige) Größe, sondern eine absolute, d.h. eine Invariante.

Wenn man beispielsweise sagt, ein Elektron hat eine Masse von rund $1 \cdot 10^{-30} \text{kg}$, dann hat diese Aussage in jedem Galileisystem der gleichen Klasse einen Sinn; bei dieser Mitteilung ist es schlicht egal, auf welches Galileisystem sich der Sprecher oder der Hörer der Mitteilung bezieht. Wenn man hört, dass der Planet eines fernen Sonnensystems eine so und so große Masse hat, dann brauchen man auch nicht rückfragen, auf welches Datum sich diese Information bezieht (weil sich die Relativgeschwindigkeit zum fernen Planeten im Laufe eines Erdenjahrs ändert). Wenn man hingegen sagt, dass sich etwas mit 5km/h nach rechts bewegt, dann macht diese Angabe erst dann einen Sinn, wenn dem Hörer klar ist, auf welches Bezugssystem sich diese Angabe bezieht, also wie das Koordinatensystem des Sprechers relativ zum Koordinatensystem des Hörers orientiert ist bzw. wie sich die Galileisysteme, auf welche sich Sprecher und Hörer beziehen, relativ zueinander bewegen. Trotzdem hält sich noch heute in einigen physikalischen Lehrbüchern hartnäckig das Gerücht, dass es eine „geschwindigkeitsabhängige Masse“ bzw. „dynamische Masse“ gäbe. Der Begriff „geschwindigkeitsabhängige Masse“ rührt in diesen Lehrbüchern von einer falschen Betrachtung her, die sich an den Vorstellungen der Newtonschen Physik orientiert (s. Kap. 10.3). In der Relativitätstheorie gibt es keine „geschwindigkeitsabhängige Masse“.

Wie schon angedeutet, misst man die Masse nicht wirklich so, wie ich das hier skizziert habe. Aber alle Methoden der Massenbestimmung beruhen auf dem hier geschilderten Prinzip eines Vergleichsverfahrens mit dem Urkilogramm, das die Invarianz der Masse für alle Galileisysteme sicherstellt. Die Masse des Urkilogramms wurde als positiv definiert, nämlich mit dem Wert 1kg . Damit sind auch alle anderen Vergleichsmassen positiv, die mit elastischen (!) Stoßprozessen charakterisiert werden können. Alle physikalischen Objekte mit positiver Masse zählen zur MATERIE. Gemäß Gl. 11.14 ist damit auch p_0 positiv und wegen Gl. 11.18 folgt, dass die Energie $E = cp_0$ von Materie stets positiv ist. Ist die Masse eines Körpers bekannt und kennt man auch seine Geschwindigkeit relativ zu einem Galileisystem, so sind in diesem Galileisystem Energie und Impuls des Körpers bekannt (Gl. 11.19 und Gl. 11.20).

11.5. Inertialsysteme

Bis hier hin gelten alle abgeleiteten Resultate für beliebige Galileiklassen. Für abgeschlossene physikalische Systeme, also Systeme, die keine weitere Wechselwirkung mit der Außenwelt haben, kann man die Frage stellen, ob es eine spezielle Galileiklasse gibt, für welche die Dynamik besonders einfach formuliert werden kann. Das ist in der Tat der Fall, nämlich für die Galileiklasse, in der Gesamtimpuls und seine Gesamtenergie ZEITUNABHÄNGIG (!) denselben Wert haben. Die Galileisysteme dieser Klasse heißen INERTIALSYSTEME. Relativ dazu beschleunigte Bezugssysteme sind Nicht-Inertialsysteme. Da sich alle Galileisysteme derselben Klasse mit konstanter Relativgeschwindigkeit bewegen, ist auch der Impuls (und damit auch die Energie) eines abgeschlossenen physikalischen Systems in allen Inertialsystemen konstant, d.h. die Definition ist konsistent, auch wenn der Impuls für unterschiedliche Inertialsysteme nicht identisch ist. Der wesentliche Punkt der Definition eines Inertialsystems ist, dass sich der Impuls eines abgeschlossenen

Systems nicht mit der Zeit ändert.

Ruhsystem und innere Energie Sei G ein abgeschlossenes physikalisches System, das der Einfachheit halber nur aus zwei Teilchen A und B bestehen soll. Bezüglich irgendeines Inertialsystems mögen sie die Impulse $p_A^{(A)}$ bzw. $p^{(B)}$ haben. Der Impuls des Gesamtsystems ist dann $p = p^{(A)} + p^{(B)}$. Die Energie

$$E(p) = \sqrt{m^2 c^4 + c^2 p^2} = E_A(p^{(A)}) + E_B(p^{(B)}) \quad (11.23)$$

des Gesamtsystems G ist konstant und gleich der Summe der beiden Teilchenenergien $E_A = \sqrt{m_A^2 c^4 + c^2 (p^{(A)})^2}$ und $E_B = \sqrt{m_B^2 c^4 + c^2 (p^{(B)})^2}$. Sie hängt von der Wahl des Inertialsystems ab und ist somit weitgehend unphysikalisch. Man kann diese vom Bezugssystem verursachte Äußerlichkeit „wegtransformieren“, indem man das RUHESYSTEM von G als Bezugssystem wählt. Das Ruhesystem ist ein Spezialfall der Inertialsysteme. Hier wird zusätzlich $p = 0$ gefordert und somit gilt für die Geschwindigkeit des Gesamtsystems $v = 0$. Im Ruhesystem besteht die Energie nur noch aus der zur Masse proportionalen INNEREN ENERGIE

$$\mathcal{U} = E(0) = mc^2$$

Da die innere Energie der Masse proportional ist, ist sie ebenfalls eine Invariante der Lorentztransformation. Die in einem beliebigen anderen Inertialsystem auftretende Differenz

$$E(p) - \mathcal{U} = \sqrt{m^2 c^4 + c^2 p^2} - mc^2$$

heißt ÄUSSERE ENERGIE des Systems. Die äußere Energie des abgeschlossenen Gesamtsystems G ist ein Artefakt, das durch das gewählte Bezugssystem zustande kommt und dem keine eigentliche physikalische Bedeutung zukommt.

Inertialsysteme kann man auch ausgehend vom Ruhesystem eines abgeschlossenen Systems definieren. Die Klasse, welche das Ruhesystem enthält, ist die Klasse der Inertialsysteme. Aufgrund des Galileischen Relativitätsprinzips bewegen sich abgeschlossene physikalische Systeme in einem Inertialsystem geradlinig und gleichförmig.

11.6. Die Axiome der Dynamik

Dass eine physikalische Größe extensiv ist, ist keine Selbstverständlichkeit. Extensive Größen gestatten uns, ein aus Teilchen aufgebautes Gesamtsystem auf die Eigenschaften seiner Teilchen zurückzuführen. Masse, Energie und Impuls sind extensive Größen.

Die besondere Bedeutung der extensiven Erhaltungsgrößen liegt darin, dass man mit ihnen Wechselwirkungsprozesse durch Bilanzgleichungen beschreiben kann. Die Vision, die in Kap. 11.1 formuliert wurde, lässt sich mit der skalaren Größe E , der Energie, und der vektoriellen Größe p , dem Impuls, tatsächlich widerspruchsfrei umsetzen: Für jedes Teilchen eines Systems lassen sich diese beiden Größen aufstellen. Wechselwirkungen laufen so ab, dass für das Gesamtsystem sowohl die Summe der Energien als auch

die Summe der Impulse aller Teilchen erhalten bleibt. Wenn man die Galileiklasse der INERTIALSYSTEME als als Beschreibungsbasis heranzieht, dann lässt sich die Dynamik auf die folgenden Axiome gründen:

- 1. Axiom: Energieerhaltung.** Alle physikalischen Vorgänge laufen so ab, dass die Gesamtenergie E eines abgeschlossenen Systems erhalten bleibt. Die Gesamtenergie ist eine zeitunabhängige extensive Erhaltungsgröße.
- 2. Axiom: Impulserhaltung.** Alle physikalischen Vorgänge laufen so ab, dass der Gesamtimpuls p eines abgeschlossenen Systems erhalten bleibt. Der Gesamtimpuls ist eine zeitunabhängige extensive Erhaltungsgröße.
- 3. Axiom: Masseninvarianz.** Die Masse $m = U/c^2$ eines abgeschlossenen Systems ist eine Invariante. Sie ist eine charakteristische extensive Größe des Systems und hat in jedem Inertialsystem denselben Wert.

Jedes mechanische Problem kann auf dieser Grundlage gelöst werden. Auch die drei Axiome der Dynamik sind Hypothesen und in gewissem Sinn metaphysisch, d.h. im Rahmen der Physik nicht mehr weiter begründbar. So lange man keinen Anlass hat, an ihnen zu zweifeln und so lange keine experimentelle Erfahrung ihnen widerspricht, hält man an ihnen fest. Zweifel an diesen drei Sätzen der Dynamik hat es in der Physikgeschichte immer wieder einmal gegeben. Dies zeigt, dass die Sätze eben aus empirischer Erfahrung induziert, d.h. extrapoliert wurden und deshalb immer den Vorbehalt tragen „Soweit uns die Natur nicht eines Besseren belehrt ...“ So schreibt beispielsweise Philosoph Wladimir Iljewitsch Lenin zum Stand der Physik seiner Zeit [7]:

„Diese Krise (der Physik) erschöpfe sich nicht darin, dass der große Revolutionär Radium das Prinzip der Erhaltung der Energie in Frage stelle. Auch alle anderen Prinzipien sind in Gefahr. Zum Beispiel das Lavoisiersche Prinzip oder das Prinzip der Erhaltung der Masse sei durch die Elektronentheorie der Materie untergraben. Nach dieser Theorie werden Atome von kleinsten mit positiver oder negativer Elektrizität geladenen Teilchen (Elektronen) gebildet, die in eine Umgebung getaucht sind, die wir Äther nennen. Die Experimente der Physiker liefern das Material, um die Bewegungsgeschwindigkeit der Elektronen und ihre Masse zu berechnen. Die Bewegungsgeschwindigkeit ist der Lichtgeschwindigkeit (300000 Kilometer pro Sekunde) vergleichbar, sie erreicht beispielsweise ein Drittel dieser Geschwindigkeit. Unter diesen Umständen muss eine zweifache Masse des Elektrons in Betracht gezogen werden.“

In der Tat schien es kurz nach der 19. Jahrhundertwende so, als ob das neu entdeckte Radium eine gewaltige Energiemenge aus dem Nichts zu produzieren schien, denn weder konnte man irgendwelche chemischen Umsätze ausfindig machen, denen man die Energieproduktion hätte zuschreiben können, noch schien eine messbare Änderung der Masse aufzutreten. Das Postulat der Energieerhaltung war damit in Frage gestellt. Dieses Problem sollte sich später beispielsweise auch beim β -Zerfall wieder stellen, bis es durch die Neutrino-Hypothese von Wolfgang Pauli gelöst werden konnte.

Bei der scheinbaren Massenzunahme des Elektrons mit der Geschwindigkeit spricht Lenin die 1901-1905 von Walter Kaufmann durchgeführten Experimente an. Heute weiß

Tabelle 4: Das Axiomensystem der Mechanik

Kinematik	Dynamik
Prinzip der Homogenität der Zeit	Energieerhaltung
Prinzip der Homogenität und Isotropie des Raums	Impulserhaltung
Relativitätsprinzip	Masseninvarianz

man, dass es diese sogenannte „relativistische Massenzunahme“ nicht gibt. Die Masse ist invariant, hat also bei jeder Geschwindigkeit den gleichen Wert.

Anders als der Energieerhaltungssatz gilt der Massenerhaltungssatz heute aber tatsächlich als widerlegt. Die von Lenin geschilderten Probleme der Physik um die Jahrhundertwende werden beispielsweise von Ludwig Boltzmann im Rahmen seiner Auseinandersetzung mit dem Kantianismus und der Philosophie Schopenhauers angesprochen [8]: „*Ebenso folgert Schopenhauer aus dem Satz vom zureichenden Grunde, dass das Gesetz der Massenerhaltung a priori klar wäre. Gerade über dieses Gesetz hat Landoldt Versuche angestellt, welche es anfangs zu widerlegen schienen. Heute scheint es freilich wahrscheinlicher, dass sie diesem Gesetz nichts werden anhaben können. Allein es handelt sich nicht um das Resultat der Versuche, vielmehr bloß darum, ob überhaupt Versuche die Macht hätten, das Gesetz zu widerlegen oder ob die Logik dem Zeiger der Waage Landoldts seinen Weg vorschreiben kann. Zum zweiten sind Zweifel an der Richtigkeit dieses Gesetzes gelegentlich beim Radium aufgetaucht. Ich bin überzeugt, dass auch diese Versuche das Gesetz bestätigen werden. Aber es ist das ein Beweis, dass es kein apriorisches Gesetz ist. Wenn es nicht gelten würde, könnten wir vom logischen Standpunkte nichts entgegenen.*“

Die Energieerhaltung hat die Revolution der Einsteinschen Relativitätstheorie überlebt, die Massenerhaltung hingegen nicht. Unser heutiges Verständnis lässt sich prägnant so zusammenfassen:

- Die Energie ist *erhalten*, aber *nicht invariant*.
- Die Masse ist *invariant*, aber *nicht erhalten*.

Mit den drei Axiomen der Kinematik und den drei Axiomen der Dynamik ist das Axiomensystem der Theorie der Mechanik vollständig (Tab. 4). Die tiefere Ursache für die Beziehung zwischen der Zeit und der Energie einerseits sowie dem Raum und dem Impuls andererseits liegt aus relativistischer Sicht darin, dass die Größen den gleichen Äquivalenz-Transformationen bzw. Symmetrien unterliegen.

12. Mechanische Energien

Zur inneren Energie \mathcal{U} eines abgeschlossenen Systems können in unterschiedlichen Formen der kinetischen bzw. potentiellen Energie beitragen. Sie werden durch Zustandsvariable charakterisiert. Der kinetischen Energie kommt dabei eine Schlüsselrolle zu, weil sie sich leicht durch Geschwindigkeitsmessungen bestimmen lässt. Von ihr ausgehend kann man den Zusammenhang der Energieformen der potentiellen Energie mit ihren beschreibenden Zustandsvariablen erschließen. Die prinzipielle Vorgangsweise wird am Beispiel der elastischen Energie und der Energie der Schwere vorgestellt.

12.1. Kinetische Energie

Als KINETISCHE ENERGIE eines freien Teilchens der Masse m mit Impuls p bzw. Geschwindigkeit v bezeichnet man die Differenz seiner Gesamtenergie E und seiner inneren Energie \mathcal{U} :

$$E_{kin} = E - \mathcal{U} \quad (12.1)$$

Als beschreibende Zustandsvariable kann man entweder den Impuls oder die Geschwindigkeit wählen:

$$E_{kin}(p) = \sqrt{m^2 c^4 + c^2 p^2} - mc^2 \quad (12.2)$$

$$E_{kin}(v) = [\gamma(v) - 1]mc^2. \quad (12.3)$$

Wenn ein Teilchen ruht, sind $v = 0$, $p = 0$ und $E_{kin} = 0$. Die kinetische Energie eines Teilchens kann man stets aus Gl. 12.3 berechnen, wenn man seine Masse m und seine Geschwindigkeit v ermittelt hat.

Als einfaches Beispiel soll nun ein Gesamtsystem G diskutiert werden, das nur aus zwei Teilchen A und B besteht. Zur Gesamtenergie $E_G = E_{kin,G} + \mathcal{U}_G$ trägt die äußere Energie bei, die stets gleich seiner kinetischen Energie $E_{kin,G}$ ist. Diese kann man wegtransformieren, indem man sich ins Ruhesystem von G begibt. Das ist immer möglich. Jedoch ist es i.A. unmöglich, ein Bezugssystem zu finden, in dem die kinetischen Energien der beiden Teilchen simultan verschwinden. Im Ruhesystem von G besteht die Energie des Gesamtsystems nur aus seiner inneren Energie

$$\mathcal{U}_G = m_G c^2 = E_A + E_B = E_{kin,A} + \mathcal{U}_A + E_{kin,B} + \mathcal{U}_B, \quad (12.4)$$

die sich seinerseits aus der inneren Energie $\mathcal{U}_A = m_A c^2$ von Teilchen A und seiner kinetischen Energie

$$E_{kin,A} = E_A - \mathcal{U}_A = \sqrt{m_A^2 c^4 + c^2 p_A^2} - m_A c^2 \quad (12.5)$$

zusammensetzt sowie den entsprechenden beiden Beiträgen für Teilchen B . In jedem anderen Inertialsystem hat die Energie des Gesamtsystems fünf Beiträge

$$E_G = E_{kin,G} + E_{kin,A} + \mathcal{U}_A + E_{kin,B} + \mathcal{U}_B,$$

weil noch die kinetische Energie $E_{kin,G}$ des Gesamtsystems hinzukommt.

- Nur die im Ruhesystem auftretenden kinetischen Energiebeiträge der Teilchen eines Gesamtsystems zählen zu seiner inneren Energie.
- Nur die Geschwindigkeiten der Teilchen bezüglich des Ruhesystems sind Zustandsvariable des Gesamtsystems.

Im Ruhesystem von G gilt für die Energie- und Impulserhaltung daher

$$E_{kin,A} + E_{kin,B} = (m_G - m_A - m_B)c^2 \quad (12.6)$$

$$p_A + p_B = 0 \quad (12.7)$$

Die Vorteile des Ruhesystems als Bezugssystem für die theoretische Analyse sind offensichtlich, denn man muss sich nicht mit der „unphysikalischen“ äußeren Energie $E_{kin,G}$ herumschlagen. Wenn man aber Experimente im Labor durchführt, ist das Ruhesystem zu Beginn des Experiments in der Regel unbekannt oder für die Beobachtungen unzugänglich. Daher werden die experimentellen Daten zunächst relativ zu einem anderen Inertialsystem gemessen. Das vom Experimentator gewählte Inertialsystem bezeichnet man als das LABORSYSTEM. Nach Abschluss der Datenaufnahme rechnet man die Daten zwecks theoretischer Analyse üblicherweise mittels einer Koordinatentransformation in das Ruhesystem um. In der Praxis erfolgen also die Beobachtungen bezüglich eines willkürlich gewählten Laborsystems und die Analysen bezüglich des Ruhesystems.

12.2. Stoßprozesse

Wenn man von Stoß spricht, dann hat man dabei ein Modell für den Wechselwirkungsprozess vor Augen, bei dem die Teilchen vor und nach dem Stoß wechselwirkungsfrei sind. Die Wechselwirkung, d.h. der Stoß selbst, spielt sich dabei in einem begrenzten räumlichen und zeitlichen Bereich ab. Im Rahmen des Modells „Stoß“ darf man die Energien bzw. Impulse der Teilchen vor und nach dem Stoß einfach addieren, um die Energie bzw. den Impuls des Gesamtsystems G zu berechnen. Der Stoß selbst ist eine kurzzeitige Wechselwirkung durch welche die Anfangs-Energien und -Impulse in die End-Energien und -Impulse übergehen. Der Erhaltungssatz im Ruhesystem lautet

$$E_A + E_B = \tilde{E}_A + \tilde{E}_B \quad (12.8)$$

$$p_A + p_B = 0 = \tilde{p}_A + \tilde{p}_B \quad (12.9)$$

Gl. 12.8 lässt sich noch nach kinetischer und innerer Energie weiter aufschlüsseln:

$$E_{kin,A} + m_A c^2 + E_{kin,B} + m_B c^2 = \tilde{E}_{kin,A} + \tilde{m}_A c^2 + \tilde{E}_{kin,B} + \tilde{m}_B c^2. \quad (12.10)$$

12.2.1. Elastischer Stoß

Wenn die Massen der Teilchen unverändert bleiben, also $\tilde{m}_A = m_A$ und $\tilde{m}_B = m_B$, dann liegt ein ELASTISCHER STOß vor. Gl. 12.10 reduziert sich dann auf Änderungen

der kinetischen Energien:

$$E_{kin,A} + E_{kin,B} = \tilde{E}_{kin,A} + \tilde{E}_{kin,B} \quad (12.11)$$

$$p_A + p_B = 0 = \tilde{p}_A + \tilde{p}_B \quad (12.12)$$

Beim elastischen Stoß sind die Summe der kinetischen Energien und die Summe der Impulse vor und nach dem Stoß gleich. Da der Impuls in die kinetische Energie

$$E_{kin} = \sqrt{m^2 c^4 + c^2 p^2} - mc^2,$$

quadratisch eingeht, erkennt man für den Sonderfall gleicher Massen $m_A = m_B$, dass Gl. 12.11 und Gl. 12.12 durch $p_A = -p_B$ gelöst werden, d.h. dass die Impulse vor dem Stoß betragsmäßig exakt gleich groß, aber entgegengesetzt gerichtet sind. Nach dem Stoß haben sie mit $\tilde{p}_A = -p_A$ bzw. $\tilde{p}_B = -p_B$ jeweils gerade das entgegengesetzte Vorzeichen wie vor dem Stoß. Macht man die Beobachtungen in einem Laborsystem S' , wo der Impuls p'_G des Gesamtsystems nicht null ist, ändert sich an der Erhaltung der kinetischen Energie (Gl. 12.11) nichts, aber anstatt Gl. 12.11 hat man nun:

$$p'_A + p'_B = p'_G = \tilde{p}'_A + \tilde{p}'_B. \quad (12.13)$$

12.2.2. Inelastischer Stoß

Es kann sein, dass die Teilchen A und B eine innere Struktur besitzen und selbst wieder Systeme von Teilchen sind. Neben ihrer kinetischen Energie können sich dann auch ihre inneren Energien $\mathcal{U}_A = m_A c^2$ bzw. $\mathcal{U}_B = m_B c^2$ und somit ihre Massen bei einem Stoß ändern. Die Energiebilanz zwischen der kinetischen Energie und den inneren Energien der Teilchen lautet wegen Gl. 12.10 nun:

$$\Delta m c^2 = (m_A + m_B - \tilde{m}_A - \tilde{m}_B) c^2 = \quad (12.14)$$

$$= - \left[E_{kin,A} + E_{kin,B} - \tilde{E}_{kin,A} - \tilde{E}_{kin,B} \right] = -\Delta E_{kin} \quad (12.15)$$

Dem hier auftretenden MASSENDEFEKT $\Delta m = m_A + m_B - \tilde{m}_A - \tilde{m}_B$ entspricht eine Änderung der inneren Energie

$$\Delta \mathcal{U}_{AB} = \mathcal{U}_A + \mathcal{U}_B - \tilde{\mathcal{U}}_A - \tilde{\mathcal{U}}_B = \Delta m c^2.$$

Der Grenzfall $\Delta \mathcal{U}_{AB} = 0$ ist der elastische Stoß. Wenn die Änderung $\Delta \mathcal{U}_{AB}$ der inneren Energien der beiden Teilchen sehr klein ausfällt, bezeichnet man den inelastischen Stoß auch als QUASI-ELASTISCHEN STOSS. Dieser Begriff ist Physiker-Slang und nicht sehr scharf definiert.

Ist $\Delta \mathcal{U}_{AB} < 0$, so nehmen die inneren Energien der beiden Teilchen ab und ihre kinetischen Energien zu. Das stellt eine EXOTHERME REAKTION dar. Der Fall $\Delta \mathcal{U}_{AB} > 0$ ist eine ENDOTHERME REAKTION, bei der kinetische Energie in die innere Energie absorbiert wird.

Der endotherme inelastische Stoß ist z.B. für die Teilchenphysik wichtig: Wenn man neue Elementarteilchen produzieren will, dann schießt man mit dem Projektil A auf ein Targetteilchen B. Dabei stehen der Energieinhalt der inneren Energie beider Ausgangsteilchen und die kinetische Energie von Projektil und Targetteilchen im Ruhesystem (des von A und B gebildete Gesamtsystems G) für die Produktion neuer Teilchen zur Verfügung. Die Massen der neuen Teilchen können sich bei Reaktionen der Hochenergiephysik signifikant von denen der Ausgangsteilchen unterscheiden, d.h. der Massendefekt kann dabei enorm groß ausfallen. Das ist auch der Grund, warum man sie als neue, von den alten Teilchen verschiedene Teilchen ansieht. Aber wenn der Massendefekt sehr klein gegen die Masse der Teilchen ist, insbesondere wenn er im Rahmen der Messunsicherheit nicht nachweisbar ist, betrachtet man die Teilchen nach dem Stoß als die alten Teilchen, die man mit der Energiezufuhr angeregt hat. Die angeregten Teilchen werden hierbei als die gleichen Objekte wie zuvor aufgefasst, deren innere Energie sich geändert hat. Die Umwandlung von kinetischer Energie in die innere Energie dieser Teilchen bezeichnet man als Anregung der Teilchen. Beides, die Produktion eines neuen Teilchens oder die Anregung eines alten Teilchens, sind zulässige Sichtweisen für endotherme Vorgänge.

12.2.3. Zerfall eines physikalischen Systems

Ein Gesamtsystem zerfalle gemäß der Zerfallsreaktion $G \rightarrow A+B$ in zwei Teilchen. Im Ruhesystem von G besteht die Energie nur aus seiner inneren Energie. Die Erhaltungsgleichungen lauten:

$$\begin{aligned} \mathcal{U}_G = m_G c^2 &= \tilde{E}_A + \tilde{E}_B \\ 0 &= \tilde{p}_A + \tilde{p}_B \end{aligned}$$

Die erste Zeile ist die Bilanzgleichung für die Energie:

$$\mathcal{U}_G = \tilde{\mathcal{U}}_A + \tilde{E}_{kin,A} + \tilde{\mathcal{U}}_B + \tilde{E}_{kin,B} \quad (12.16)$$

bzw.

$$\Delta \mathcal{U} = \Delta m c^2 = E_{kin,A} + E_{kin,B} \quad (12.17)$$

Hier ist $\Delta \mathcal{U} = \mathcal{U}_G - (\tilde{\mathcal{U}}_A + \tilde{\mathcal{U}}_B)$ die Differenz der inneren Energien vor und nach dem Prozess. Ihr entspricht ein MASSENDEFEKT von $\Delta m = m_G - (\tilde{m}_A + \tilde{m}_B)$, d.h. die Masse bleibt bei einer Zerfallsreaktion i.A. nicht erhalten. Die Summe der Massen der Reaktionsprodukte ist i.A. kleiner als die des Ausgangsstoffs G. Die gesamte kinetische Energie der beiden aus dem Zerfall resultierenden Bruchstücke wird mit der inneren Energie von G bezahlt, d.h. mit einem Teil seiner Masse (Gl. 12.17). Egal ob in der Chemie beim exothermen Zerfall eines Moleküls oder in der Atomphysik bei einem atomaren Übergang oder in der Kernphysik bei der Kernspaltung, stets tritt ein Massendefekt Δm auf, so dass die Summe der kinetischen Energien der Bruchstücke gleich $\Delta m c^2$ ist, d.h. gleich der Änderung der inneren Energie der Bruchstücke ist. Die innere Energie des Gesamtsystems, die ja aus den kinetischen Energien und den inneren Energien der Bruchstücke besteht, bleibt selbstverständlich dabei unverändert.

Beispiel: Ein Körper G zerfalle in zwei Teile A und B . Im Ruhesystem des Körpers habe der erste Teil eine kinetische Energie von 200 J und der zweite eine kinetische Energie von 1.6 kJ . Wie groß ist der Massendefekt? Es ist $\Delta m c^2 = 1.8 \text{ kJ}$ und mit $c^2 \approx 9 \cdot 10^{16} \text{ m}^2/\text{s}^2$ folgt, dass der Massendefekt $\Delta m \approx 2 \cdot 10^{-14} \text{ kg}$ beträgt.

12.3. Die mechanischen Grenzfälle

In diesem Abschnitt sollen einige Gleichungen für zwei wichtige Grenzfälle einander gegenübergestellt werden. Für den Bereich sehr großer Geschwindigkeiten verwendet man den RELATIVISTISCHEN GRENZFALL. Er ist beispielsweise für die Elektrodynamik, die Astrophysik oder die Teilchenphysik relevant. Der NICHTRELATIVISTISCHE GRENZFALL bzw. der NEWTONSCHE GRENZFALL beschreibt den Bereich kleiner Geschwindigkeiten und hat sich insbesondere für den Bereich der Alltagsphysik bewährt.

12.3.1. Relativistischer Grenzfall

Die relativistische Näherung verwendet man, wenn die Geschwindigkeit v in die Größenordnung von c kommt, d.h. wenn

$$v \approx c. \quad (12.18)$$

Als Konsequenz kann die innere Energie $\mathcal{U} = mc^2$ in Gl. 11.22 vernachlässigt werden. Energie und Betrag des Impulses werden dann näherungsweise proportional zueinander

$$E \approx c|p|. \quad (12.19)$$

Der relativistische Grenzfall ist beispielsweise für Photonen wichtig.

12.3.2. Newtonscher Grenzfall

Die Alltagserfahrung spielt sich in einem Geschwindigkeitsbereich ab, der klein gegen die Raum/Zeit-Konstante ist:

$$|v| \ll c \quad (12.20)$$

Wenn man nicht gerade über hochpräzise Messinstrumente verfügt, kann man denn keine Zeitdilatation feststellen, d.h. es gilt

$$t' \approx t \quad (12.21)$$

und das Gesetz der Geschwindigkeitstransformation läuft näherungsweise auf eine einfache Addition der Geschwindigkeiten hinaus:

$$u \approx v + u'. \quad (12.22)$$

In nichtrelativistischer Näherung gilt für den Impuls

$$p \approx mv \quad (12.23)$$

und wegen $[\gamma - 1]c^2 \approx v^2/2$ folgt aus Gl. 12.3 für die kinetische Energie näherungsweise (s. 1.4)

$$E_{kin} \approx \frac{1}{2}mv^2 = \frac{p^2}{2m} \quad (12.24)$$

bzw. für die Gesamtenergie

$$E \approx \frac{1}{2}mv^2 + \mathcal{U}. \quad (12.25)$$

Bei der Wechselwirkung von Teilchen wird der „MASSENERHALTUNGSSATZ“ in nichtrelativistischer Näherung befolgt. Wenn also beispielsweise m_A bzw. m_B die Massen zweier Teilchen A und B vor einem Stoß sind und \tilde{m}_A bzw. \tilde{m}_B die Massen nach dem Stoß, so gilt

$$m_A + m_B \approx \tilde{m}_A + \tilde{m}_B. \quad (12.26)$$

Alle Näherungen der Newtonschen Physik sind für die Alltagsphysik fantastisch genau, so dass man im Rahmen der Messunsicherheit ohne schlechtes Gewissen in den Gln. 12.21-12.26 das Ungefährzeichen durch ein Gleichheitszeichen ersetzen darf. In den nun folgenden Kapiteln werden Mechanik, Kontinuumsphysik und Thermodynamik meist in dieser Newtonschen Näherung präsentiert.

12.4. Potentielle Energie

Wenn sich die innere Energie eines Teilchens ändert, weil es Energie nach außen abgibt, dann kann man die Energieänderung im Prinzip durch die Massenänderung charakterisieren. Im Newtonschen Limes bleibt das aber ein frommer Wunsch, weil die Massenänderung so klein ist, dass sie mit geläufigen Messinstrumenten und Messmethoden nicht zugänglich ist.

Um die Größenordnung des Massendefekts bei mechanischen Wechselwirkungen abzuschätzen, kann man beispielsweise zwei Wagen A und B mit einer Masse von $m \approx 0.1$ kg betrachten, die sich mit einer gespannten Feder abstoßen. Wenn man sie loslässt, so „zerfällt“ dieses zusammengesetzte System G in seine beiden Teile A und B, die mit zunehmender Geschwindigkeit in entgegengesetzter Richtung auseinander streben, während sich die Feder mit zunehmender Längenänderung entspannt. Am Ende des Vorgangs werde von jedem Wagen eine Geschwindigkeit von größenordnungsmäßig 1 m/s erreicht. Die kinetische Energie der beiden Wagen stammt aus der inneren Energie des Systems G. Setzt man in Gl. 12.17 die Summe der beiden kinetischen Energien $E_{kin,A} + E_{kin,B} = 0.1$ J ein, so ergibt sich ein Massendefekt von $\Delta m = 0.1 \text{ J} / 9 \times 10^{16} \text{ m}^2 \text{ s}^{-2} \approx 1 \times 10^{-15} \text{ g}$. Auch mit den besten Messmethoden ist es unmöglich, eine so winzig kleine Massenänderung an einem makroskopischen Objekt nachzuweisen, das eine um den Faktor 10^{17} größere Masse hat. Der ernüchternde Schluss ist, dass man mit dem Massendefekt zwar im Prinzip

ein Maß für Energieänderungen hat, aber man kann es weder messen noch beobachten, und das ist nun einmal nicht sehr praktisch.

Beobachtbar und damit zugänglich, ist aber die Formänderung der Feder und schon mit wenigen Experimenten lässt sich einsehen, dass es einen Zusammenhang mit der freigesetzten Energie gibt. Man kann daher auf den Gedanken kommen, die Auslenkung x der Feder als ein Maß der im ursprünglichen System G gespeicherten Energie zu verwenden, die anschließend in kinetische Energie überführt wird. Diese Idee lässt sich verallgemeinern: Wenn sich die kinetische Energie in einem abgeschlossenen physikalischen System ändert, dann trachtet man danach, eine bestimmte beobachtbare Größe zu finden, die man als phänomenologisches Maß für die Änderung der inneren Energie heranziehen kann. Oft eignen sich Positions- oder Lagekoordinate des Systems G oder seiner Teile. Jede beobachtbare Größe q , die für eine Charakterisierung von Beiträgen zur inneren Energie geeignet ist, bezeichnet man als mechanische ZUSTANDSVARIABLE. Der ihr zugeordnete Energiebeitrag wird als POTENTIELLER ENERGIEBEITRAG $E_{pot}(q)$ bezeichnet.

Die potentielle Energie $E_{pot}(q)$ charakterisiert einen Beitrag zur inneren Energie, der durch irgendeine beobachtbare allgemeine Größe q parametrisiert werden kann. Die Größe q stellt die Zustandsvariable des entsprechenden potentiellen Energiebeitrags dar. Potentielle Energien stellen Anteile der inneren Energie dar, die sich in kinetische Energie umsetzen und dadurch bestimmen lassen.

Das Konzept der potentiellen Energie ist recht abstrakt. In den nachfolgenden Abschnitten soll es daher an einigen Beispielen erläutert werden. Einige Vorbemerkungen sollen die Grenzen und Möglichkeiten des Konzepts kurz erläutern:

1. Die potentielle Energie kann von mehreren Zustandskoordinaten abhängen. Ist die Zuordnung zwischen Zustandskoordinate und potentieller Energie nicht immer eindeutig, so strebt man an, einen minimalen Satz an Zustandsvariablen zu identifizieren, der die Eindeutigkeit herstellt. Die Anzahl der eindeutigen Zustandsvariablen stellen die energetischen Freiheitsgrade des energetischen Modells eines physikalischen Systems dar.
2. Man kann unterschiedliche unabhängige Zustandsvariable wählen. Diese lassen sich eindeutig ineinander transformieren. Das eröffnet die Möglichkeit, sich für solche zu entscheiden, die eine Rechnung möglichst vereinfachen.
3. Der Zusammenhang zwischen Zustandsvariablen und potentieller Energie ist meist nicht von vornherein bekannt. Man muss ihn dann durch empirische Untersuchungen herausfinden.
4. Neben der Zustandsvariablen q kann auch ihre zeitliche Änderung \dot{q} betrachtet und für die Charakterisierung kinetischer Energiebeiträge herangezogen werden. Energetische Modelle sind i.A. eine Funktion des Koordinatenpaares (q, \dot{q}) .
5. Man kann nicht erwarten, dass energetische Modelle, die so konstruiert werden, alle energetischen Freiheitsgrade eines makroskopischen Systems erfassen. Ein Kontinuum hätte unendlich viele Freiheitsgrade und selbst das atomistische Modell eines

makroskopischen Systems hat eine ungeheuer große Zahl von Freiheitsgraden, die so groß ist, wie die Zahl der atomistischen Einheiten. Das kann in der Größenordnung von 10^{23} Freiheitsgraden liegen, um einmal eine Zahl hier zu nennen. Diese alle erfassen zu wollen, wäre auch nicht sinnvoll. Man beschränkt sich darauf, die energetischen Hauptbeiträge durch einige wenige bequem zu messende Zustandsvariable zu charakterisieren. Wenn deren Anzahl f ist und zwischen 1 und 100 liegt, dann hat man $10^{23} - f$ Energiebeiträge nicht erfasst, also fast alle. Man geht optimistisch davon aus, dass das trotzdem genügt, um zumindest eine Beschreibung der wesentlichen physikalischen Vorgänge zu gewinnen.

Potentielle Energien werden i.A. durch rein empirische Untersuchungen gewonnen. Über den funktionalen Zusammenhang $E_{pot}(q) = f(q)$ zwischen der Zustandsvariablen und der potentiellen Energie weiß man in der Regel erst einmal gar nichts. Wenn man aus der Theorie nicht einmal einen Tipp bekommen kann, von welcher Art die Funktion $f(q)$ ist, welche die empirischen Daten beschreiben soll, beschreibt man sie meist durch eine POTENZREIHE

$$f(q) = a_0 + a_1q + a_2q^2 + \dots + a_nq^n \quad (12.27)$$

mit konstanten Koeffizienten a_1, \dots, a_n . Beliebter ist die Beschreibung empirischer Daten durch die STANDARDPOTENZREIHE

$$f(q) = k_0 + \frac{1}{1!}k_1q + \frac{1}{2!}k_2q^2 + \dots + \frac{1}{n!}k_nq^n \quad (12.28)$$

mit den konstanten Koeffizienten k_1, \dots, k_n . Diese hängen nämlich mit den Ableitungen der Funktion f an der Stelle $q = 0$ auf folgende einfache Weise zusammen:

$$k_0 = f(0), \quad k_1 = f'(0), \quad k_2 = f''(0), \quad \dots, \quad k_n = f^{(n)}(0),$$

wobei hier $f^{(n)}(0)$ die n -te Ableitung der Funktion f bedeutet, bei der anschließend $q = 0$ eingesetzt wird.

12.4.1. Das Hookesche Gesetz

Wir staten das Ende einer waagrecht einjustierten Bahn mit einer Feder aus. Diese drückt man um eine Strecke x zusammen. Auf die Bahn wird ein Wagen gesetzt, der mit der Anordnung in Kontakt gebracht wird (Abb. 36a). Indem man die Feder zusammendrückt, erhöht man ihre innere Energie. Dieser Anteil ist die in Form eines Beitrags zur elastischen Energie gespeicherte potentielle Energie $E_{pot}(x)$ der Feder. Sie wird durch den Parameter x charakterisiert. Die Auslenkung x der Feder ist die Zustandsvariable für den elastischen Energiebeitrag.

Wird die Feder losgelassen, gibt sie diese Energie wieder ab. Das zeigt sich darin, dass sich der Wagen nach dem Loslassen der Feder in Bewegung setzt. Experimentell findet man für die Endgeschwindigkeit v des Wagens, dass $v \propto x$ (Abb. 36b). Unter der Annahme, dass sich (näherungsweise) die potentielle Energie vollständig in kinetische Energie $E_{kin} = \frac{1}{2}mv^2$ des Wagens umsetzt, gilt

$$E_{kin}(v) = \frac{1}{2}mv^2 = E_{pot}(x) \propto x^2.$$



Abbildung 36: Hookesches Gesetz: (a) Eine Feder wird um eine Strecke x ausgelenkt und losgelassen. Auf einen Wagen, der sich auf einer horizontalen Bahn befindet, wird Energie übertragen. Die dem Wagen erteilte Endgeschwindigkeit wird gemessen. (b) Trägt man Geschwindigkeit gegen Federauslenkung auf, erkennt man im Rahmen der (durch Fehlerbalken angegebenen) Messunsicherheit einen linearen Zusammenhang. Daraus folgt das Hookesche Gesetz (Gl. 12.29).

Aus dem experimentellen Befund folgt dann durch Anpassung an die Standardpotenzreihe (Gl. 12.28) das HOOKESCHE GESETZ

$$E_{pot}(x) = \frac{1}{2} k x^2. \quad (12.29)$$

mit einer Konstanten k , der FEDERKONSTANTEN, welche ein Charakteristikum für die Steifigkeit der Feder ist. Die Endgeschwindigkeit

$$v = \sqrt{k/m} x \quad (12.30)$$

des Wagens nimmt bei gleicher Auslenkung x mit der Steifigkeit k der Feder zu und mit seiner Masse m ab.

Man sollte immer im Auge behalten, dass die Modellbildung erstens auf der optimistischen Annahme beruht, dass bereits eine einzige unabhängige Zustandsvariable von größenordnungsmäßig 10^{23} genügt, um Änderungen der inneren Energie des aus Wagen und Feder bestehenden makroskopischen Systems ausreichend genau zu erfassen und zweitens, dass das System in guter Näherung als abgeschlossen betrachtet werden kann. So erstaunlich dieser Berufsoptimismus der Physiker auch sein mag, der Erfolg gibt ihnen meist näherungsweise recht.

Das Hookesche Gesetz, Gl. 12.29, ist ein mathematisches Modell für die Abhängigkeit der potentiellen Energie von der Auslenkung x , das aus empirischen Daten herausgelesen wurde. Da mag man sich fragen, ob es das einzig mögliche mathematische Modell ist, mit dem die empirischen Daten beschrieben werden können. Die Antwort ist: nein! Man hätte die Messdaten genauso gut durch das mathematische Modell

$$E_{pot}(x) = \frac{1}{2} k_2 x^2 + \frac{1}{4!} k_4 x^4 \quad (12.31)$$

mit zwei empirischen Konstanten k_2 und k_4 beschreiben, vielleicht sogar viel besser. Aber so lange das einfachere mathematische Modell (mit $k_2 \neq 0$ und $k_4 = 0$) die Messdaten im Rahmen der streuenden Messergebnisse (d.h. in den Schranken der Unsicherheit) beschreibt, gibt es keinen Anlass, ein komplizierteres mathematisches Modell anzunehmen. Wenn man aber nun einmal das kompliziertere Gesetz für die Beschreibung zugrunde legen würden, dann ließe sich aus den Messdaten eine obere Schranke für k_4 herauslesen. Durch Verfeinerung der Messung mit dem Ziel, die Messunsicherheit zu verringern, kann man diese Schranke im Prinzip immer weiter herabsetzen. Irgendwann könnte man dabei auf einen Wert für k_4 stoßen, der im Rahmen der Unsicherheit einen von Null verschiedenen Wert hat. Dann wäre Gl. 12.31 das korrekte empirische Gesetz und nicht Gl. 12.29. Trotzdem würden man dann immer noch für viele Überlegungen von Gl. 12.29 ausgehen, also von einer Hookeschen Gesetzmäßigkeit, weil sie immer noch als Näherung brauchbar ist und uns physikalische Probleme bequemer oder schneller durchdenken lässt als das genauere (und eigentlich korrektere) Gesetz gemäß Gl. 12.31.

Wenn mehrere mathematische Modelle die Messdaten beschreiben und kein logisches Argument eines davon ausschließt, dann sind alle gleich richtig und die Wahl des mathematischen Modells liegt im Ermessen des Physikers. Gewöhnlich wird er eine denkökonomische Wahl treffen, sich also für das Modell entscheiden, das ihm am einfachsten beim Durchdenken der möglichen Konsequenzen des erschlossenen Zusammenhangs weiterhilft. Beispielsweise entscheidet man sich tatsächlich meist für das Hookesche Energiemodell, Gl. 12.29. Hookesches Verhalten ist kein allgemeines, kein prinzipielles Gesetz der Physik, sondern eine empirische Beschreibung von Federn, die nur innerhalb eines bestimmten Wertebereichs gültig ist. Es ist ja offensichtlich, dass das Gesetz nicht für beliebig große x gelten kann und von Verarbeitung, Form und Material der Feder abhängig sein kann. Selbst wenn sich tatsächlich herausstellt, dass die Feder nur mit einem komplizierteren als dem Hookeschen Gesetz beschrieben werden muss, verwendet man oft das Hookesche Gesetz als mathematisches Näherungsmodell erster Ordnung, ähnlich wie es mit der Optik erster Ordnung war, wo die Abbildung realer Linsen durch die ideale geometrisch-projektive Abbildung beschrieben wurde, die hier durch den Parameter f und die Lage der Hauptebene gekennzeichnet ist. Das Hookesche Gesetz ist zugegebenermaßen kein ganz fairer Vergleich, denn es gibt eigentlich überhaupt keine theoretische Basis für das Hookesche Gesetz, während man mit der idealen Abbildung eine Theorie hat und die empirischen Daten an theoretische Parameter anpassen kann.

12.4.2. Das Schweregesetz

Ist ein Wagen auf einer horizontal ausgerichteten Bahn zur Ruhe gekommen, dann bleibt er in Ruhe. Erst wenn man die Bahn neigt, ändert sich das (Abb. 37a). Der Wagen beginnt, sich mit zunehmender Geschwindigkeit zu bewegen, also beschleunigt. Trägt man das Quadrat der Geschwindigkeit des Wagens gegen die „durchfallene“ Höhe h aus, so erhält man im Rahmen der Messunsicherheit einen linearen Zusammenhang (Abb. 37b).

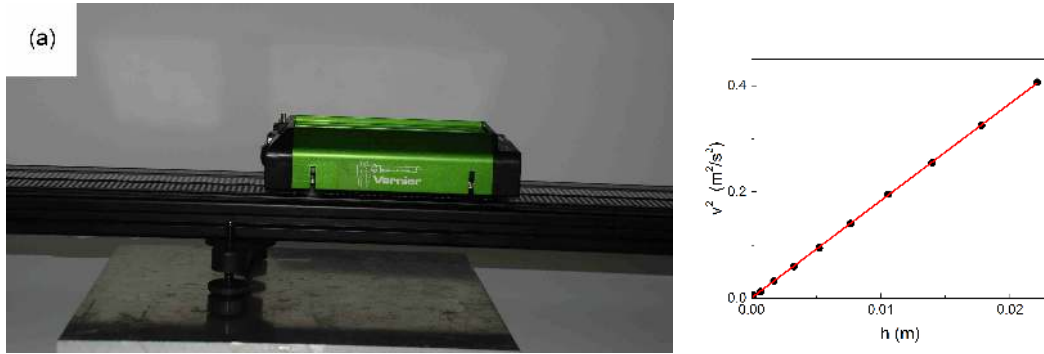


Abbildung 37: Schweregesetz. (a) Ein Wagen läuft eine gegen die Horizontale geneigte Bahn hinab. Gemessen wird die Geschwindigkeit v eines Wagens als Funktion der „durchfallenen“ Höhen h . (b) Trägt man v^2 gegen h auf, ergibt sich empirisch ein linearer Zusammenhang. Es sind keine Fehlerbalken im Diagramm angegeben, weil die Messunsicherheit kleiner als die Größe der für die Messdaten eingezeichneten Kreise ausfiel.

Wenn man für die Proportionalitätskonstante $2g_{eff}$ schreibt,¹² ergibt sich das empirische Gesetz:

$$\frac{1}{2}v^2 = g_{eff}h. \quad (12.32)$$

Aus Abb. 37b entnimmt man beispielsweise $g_{eff} \approx 9 \text{ m/s}^2$. Multipliziert man diese Gleichung mit der Masse m des Wagens aus, so stellt die linke Seite die kinetische Energie dar. Ihre Zunahme entspricht eine Abnahme der potentiellen Energie um

$$\Delta E_{pot} = mg_{eff}h. \quad (12.33)$$

Je mehr Mühe man darauf verwendet, dissipative Energieverluste zu vermeiden, desto mehr nähert sich die Konstante g_{eff} im Vakuum¹³ dem Wert des SCHWEREPARAMETERS

$$g \approx 10 \text{ m/s}^2$$

an. Im Folgenden wird das mit

$$g_{eff} \rightarrow g \quad (12.34)$$

sich ergebende Idealgesetz

$$E_{pot}(h) = mgh \quad (12.35)$$

für den verlustfreien (dissipationsfreien) Grenzfall im Vakuum verwendet.

¹²Der Faktor 2 tut nichts zur Sache. Er wurde hier nur gewählt, damit die späteren Formeln so aussehen, wie sie in Lehrbüchern der Mechanik üblich sind.

¹³Bei einer genauen Analyse muss auch das Fluid berücksichtigt werden, in dem sich der Wagen bewegt, s. Kap 18.3.

12.5. Dissipationskinetik

Jedes makroskopische System besteht aus extrem vielen Teilen, die miteinander wechselwirken. Der in Abb. 37 gezeigte Wagen, der mit der Schiene, auf der er sich bewegt, und der sonstigen Umgebung ein abgeschlossenes System darstellen würde, hat Räder, Achsen usw., die an dem untersuchten Vorgang beteiligt sind. Er wechselwirkt mit der umgebenden Luft, produziert Schall usw., und seine Schiene sowie Umgebung ist mit seiner Energie letztlich gekoppelt, ja jedes der kleinsten Teile von Wagen, Schiene und Umgebung kann sich gegen andere Teile bewegen und Energie in verschiedenster Form speichern oder abgeben. Wenn man also einen Wagen durch so reduktionistische Energiemodell beschreibt, wie es in Kap. 12.4 geschah, so muss man sich der Ehrlichkeit halber eingestehen, dass man von der ungeheuer großen Zahl an Energiebeiträgen, welche da zusammenwirken, praktisch alle (!) Beiträge nicht berücksichtigt hat, bis auf die ein, zwei, drei, ..., wenigen Ausnahmen, für die man geeignete phänomenologische Variable zur Beschreibung der potentiellen Energie identifizieren hat. Das sind die wenigen „sichtbaren Variablen“ bzw. Zustandsvariablen des physikalischen Systems, denen nahezu unendlich viele „verborgene Variable“ des physikalischen Systems gegenüberstehen. Deren Beiträge sind zwar jeweils eventuell nur zu ganz geringem Teil am Energieumsatz beteiligt, aber aufgrund ihrer ungeheuer großen Anzahl kann sich im Laufe der Zeit dennoch ein makroskopischer Effekt einstellen.

Wenn sich ein makroskopisches Teilsystem A (z.B. ein Wagen, eine Münze oder ein Ball) relativ zu einem anderen Teilsystem B (z.B. die umgebende Luft) bewegt, mit dem es wechselwirken kann, so stellt man in der Regel fest, dass die Anfangsgeschwindigkeit von A relativ zu B abnimmt, ohne dass man, abgesehen von der Geschwindigkeitsänderung selbst, an den Objekten A oder B irgendwelche Änderungen ausmachen kann. Wenn man B nur groß genug wählt, kann man das aus A und B bestehende Gesamtsystem G bezüglich der Wechselwirkung zwischen A und B als abgeschlossen betrachten. Treten für die Summe aller „sichtbaren“ Energiebeiträge Änderungen auf, so kann man aufgrund der Energieerhaltung für abgeschlossene Systems den sicheren Schluss ziehen, dass noch mindestens ein „verborgener“ Energiebeitrag vorhanden ist, wenn nicht sehr viel mehr. Diese Energiebeiträge, für die keine phänomenologisch beobachtbaren Indikatoren zur Verfügung stehen, sind „unsichtbar“. Nur in Ausnahmefällen kann man sie „sichtbar“ machen, beispielsweise weil man, wie bei der Energieabgabe radioaktiver Stoffe, den Massendefekt nachweisen kann.

Phänomene, bei denen im obigen Sinne „unsichtbare“ Energie unbekannter Mitspieler B eine Rolle spielt und bei denen stets die aus den „sichtbaren“ Energiebeiträgen A von den unbekannt Mitspielern B nicht zurückerstattet wird, werden als DISSIPATION bezeichnet. Dissipation liegt genau dann vor, wenn man einen Energietransfer $A \rightarrow B$ beobachtet, nicht aber den umgekehrten Energietransfer $B \rightarrow A$. Dazu gehören insbesondere jene Phänomene, bei denen die Geschwindigkeit eines makroskopischen Objekts A relativ zu einem damit in Kontakt stehenden Teilsystems B abnimmt, bei denen aber nicht beobachtet wird, dass die Geschwindigkeit des makroskopischen Objekts A wieder zunimmt, weil die energetischen „Mitspieler“ des Teilsystems B wieder Energie zurückgeben. Die Bezeichnung „Dissipation“ ist für Vorgänge reserviert, die bei einer Beobachtung



Abbildung 38: Dissipationsgesetze. (a) Wagen mit einer Filzplatte, die in einstellbarer Weise auf der Schiene schleift. (b) Wagen mit Ausleger. Der Ball bewegt sich in einer mit Glyzerin gefüllten Wanne. (c) Im ersten Fall nimmt die Geschwindigkeit linear mit der Zeit ab, im zweiten Fall nichtlinear.

auf einer ausreichend makroskopischen Skala als IRREVERSIBEL erscheinen.

Um zumindest die Dissipationskinetik quantitativ zu beschreiben, kann man beispielsweise so vorgehen, wie das in Abb. 38 am Beispiel eines Wagens demonstriert wird, der auf einer Schiene läuft. Man erteilt ihm eine Anfangsgeschwindigkeit und misst die Rate \dot{v} , mit der Geschwindigkeit v als Funktion der Zeit t abnimmt. Das wurde hier für zwei Modellsituationen durchgeführt: Im ersten Fall (Abb. 38a) schleift dabei eine am Wagen befestigte einstellbare Filzplatte über den Boden. Im zweiten Fall (38b) wurde ein Ball über einen Ausleger am Wagen befestigt. Der Ball bewegte sich in einer mit Glyzerin gefüllten Wanne. Abb. 38c zeigt, dass die Geschwindigkeit im ersten Fall linear mit der Zeit abnimmt, im zweiten Fall nichtlinear. Der Verlauf dieser Kurven ist für entgegengesetzte Bewegungsrichtungen gleich. Daher hängt das Phänomen nur vom Betrag $|v|$ der Relativgeschwindigkeit (!) des Wagens zu jenen Teilen des Systems ab, gegen die er sich bewegt. Anders als beim Stoß, bei dem man meist vom Ruhesystem ausgeht, wählt man bei Dissipationsphänomenen ein Bezugssystem, bei dem die Umgebung ruht und z.B. der Wagen sich relativ zu diesem Bezugssystem und somit zur Umgebung mit der Geschwindigkeit v bewegt.

Durch Kurvenanpassung an die Messdaten von Abb. 38c und Differentiation nach der Zeit erhält man $|\dot{v}|$ als Funktion von $|v|$. Im ersten Fall folgt daraus das Gesetz der KINETIK NULLTER ORDNUNG,

$$\frac{d|v|}{dt} = -a_0, \quad (12.36)$$

und im zweiten Fall das Gesetz der KINETIK ERSTER ORDNUNG,

$$\frac{d|v|}{dt} = -a_1 |v|. \quad (12.37)$$

Allgemeiner gesprochen erweist sich die Rate der Geschwindigkeitsabnahme als eine Funktion $f(|v|)$ des Geschwindigkeitsbetrags des Systems A relativ zur Umgebung B:

$$|\dot{v}| = \frac{d|v|}{dt} = -f(|v|).$$

Indem hier das negative Vorzeichen herausgezogen wurde, ist $f(|v|)$ eine positive Funktion, für die man eine Potenzreihe

$$f(|v|) = a_0 + a_1 |v| + a_2 v^2 + \dots, \quad (12.38)$$

ansetzen kann. Die Koeffizienten a_0, a_1, a_2, \dots müssen aus empirischen Daten ermittelt werden. Manchmal dominiert einer der Beiträge. Gl. 12.36 folgt aus dem allgemeinen Modellgesetz, Gl. 12.38, für den Grenzfall, bei dem der Koeffizient a_0 dominiert und alle anderen Koeffizienten vernachlässigbar klein sind, und Gl. 12.37 entsprechend, wenn der Koeffizient a_1 dominiert.

Wenn man die „sichtbare“ Energie eines Systems mit E bezeichnet, so ist die dissipative Abnahme \dot{E} gleich der Abnahme des kinetischen Energiebeitrags, weil nur die auf Relativbewegung beruhende Energie dissipiert wird. Potentielle Energie eines Teilsystems A wird durch Wandlung in kinetische Energie und seine anschließende Dissipation an die „verborgene“ Energie des abgeschlossenen Systems abgegeben. Die DISSIPATIVE VERLUSTLEISTUNG

$$P_v = -\frac{dE}{dt} = -\frac{dE_{kin}}{dt} = -\frac{m}{2} \frac{d}{dt} (v^2) = -mv \frac{dv}{dt},$$

erhält man durch Einsetzen von Gl. 12.36 bzw. Gl. 12.37. Für Dissipation nullter Ordnung ergibt sich damit

$$P_v = ma_0 |v|$$

und für Dissipation erster Ordnung

$$P_v = a_1 m v^2. \quad (12.39)$$

Nur für den Sonderfall, dass die Dissipationskinetik hinsichtlich der Geschwindigkeit von erster Ordnung ist, ist sie es auch hinsichtlich der kinetischen Energie:

$$\frac{dE_{kin}}{dt} = -2a_1 E_{kin}. \quad (12.40)$$

Für ein System, das aus mehreren Teilchen (=Teilsystemen) besteht, redet man bei der kinetischen Energie und bei der Dissipation i.A. nicht über die gleiche Geschwindigkeit. Wenn man die kinetische Energie als Teilbeitrag zur inneren Energie des Gesamtsystems aller Teilchen betrachtet, dann ist das Bezugssystem das Ruhesystem des Teilchensystems, also das Bezugssystem, in dem der Gesamtimpuls des Teilchensystems verschwindet. Die Geschwindigkeiten v_i der einzelnen Teilchen (hier des i -ten) haben einen zugehörigen bzw. konjugierten Impuls p_i , und beide sind Zustandsvariable des Systems. Bei der Dissipation geht es hingegen um die Geschwindigkeit eines makroskopischen Systems relativ zur Geschwindigkeit eines anderen makroskopischen Systems. Diese Geschwindigkeitsdifferenz ist die Geschwindigkeit v , welche in die phänomenologische Beschreibung der Dissipation eingeht. So lange die Relativgeschwindigkeit null ist, also beide Systeme relativ zueinander ruhen, gibt es auch keine Dissipation, d.h. weder das eine noch das andere System „verliert“ Energie.

12.6. Leistung und Energiefluss bei offenen Systemen

Ein System heißt offen, wenn es nach außen Energie abgeben kann oder ihm von außen Energie zugeführt werden kann. Die Rate der Energieabgabe nennt man die vom System nach außen abgegebene LEISTUNG P . Für die gesamte innere Energie \mathcal{U} des Systems gilt

$$\frac{d\mathcal{U}}{dt} + P(t) = 0 \quad (12.41)$$

Die Leistung wird auf den Energielieferanten bezogen. P ist positiv, wenn das betrachtete System die Leistung nach außen abgibt. Seine innere Energie nimmt infolgedessen ab. Die Leistung wird in der Einheit WATT gemessen:

$$[P] = 1 \text{ Watt} = 1 \text{ W} = 1 \text{ J/s}.$$

Umgekehrt kann man die Energieeinheit Joule dann auch durch die Leistungseinheit Watt ausdrücken:

$$1 \text{ Joule} = 1 \text{ Wattsekunde} = 1 \text{ Ws} \quad (12.42)$$

Wenn bei der „Stromrechnung“ in Kilowattstunden ($1 \text{ kWh} = 3.6 \text{ Millionen Joule}$) abgerechnet wird, so zahlt man nicht für den elektrischen Strom, sondern für den Zustrom an nutzbarer Energie, die das Elektrizitätswerk liefert. Die Stromrechnung ist sozusagen eine Energiestrom-Abrechnung. Wenn allgemein die innere Energie \mathcal{U} eines physikalischen Gesamtsystems abnimmt, so entspricht dem ein ENERGIEFLUSS bzw. ENERGIESTROM $I_E = P$ nach außen. Energiestroms I_E und Leistung P sind synonyme Begriffe, d.h. man kann Gl. 12.41 auch in der Form

$$\frac{d\mathcal{U}}{dt} + I_E = 0 \quad (12.43)$$

ausdrücken. Gl. 12.43 stellt eine KONTINUITÄTSGLEICHUNG dar. So lange man Dissipation vernachlässigen kann, gilt

$$\frac{d}{dt} \left(\sum E_{kin} + \sum E_{pot} \right) + \sum I_E = 0. \quad (12.44)$$

Das Summenzeichen \sum steht hier symbolisch dafür, dass alle kinetischen und potentiellen Energiebeiträge sowie die Ströme zu addieren sind.

12.7. Platons Höhlengleichnis und die Phänomenologie

„Dinge sind uns als außer uns befindliche Gegenstände unserer Sinne gegeben. Allein von dem, was sie an sich selbst sein mögen, wissen wir nichts, sondern kennen nur ihre Erscheinungen, d. i. die Vorstellungen, die sie in uns bewirken, indem sie unsere Sinne affizieren. Demnach gestehe ich allerdings, daß es außer uns Körper gibt, d.i. Dinge, die, obzwar nach dem, was sie an sich selbst sein mögen, uns gänzlich unbekannt sind, wir sie durch die Vorstellungen kennen, welche ihr Einfluß auf unsre Sinnlichkeit uns verschafft, und denen wir die Benennung eines Körpers geben. Dieses Wort bedeutet also bloß die Erscheinung jenes uns unbekanntes, aber nichtsdestoweniger wirklichen Gegenstandes.“

(Immanuel Kant, „Prolegomena zu einer jeden künftigen Metaphysik die als Wissenschaft wird auftreten können,, (1783))

Im Platonschen Höhlengleichnis leben Menschen in einer Höhle. Sie können nur auf eine Wand blicken, auf der sie die Schattenbilder wahrnehmen können, die von Vorgängen entworfen werden, die außen stattfinden. Wenn Menschen außen unterschiedliche Gegenstände hin- und hertragen, so dass sie selbst verborgen bleiben. Dann sehen die Höhlenbewohner nur das Schattentheater dieser Gegenstände. Wenn einer der Träger spricht, hallt das Echo von der Höhlenwand so zurück, als ob die Schatten sprächen. Daher meinen die Höhlenbewohner, die Schatten könnten reden. Sie betrachten die Schatten als Lebewesen und deuten alles, was geschieht, als deren Handlungen. Das, was sich auf der Wand abspielt, ist für sie die gesamte Realität. Sie entwickeln eine Wissenschaft von den Schatten und versuchen, in deren Auftreten und Bewegungen Gesetzmäßigkeiten festzustellen und daraus Prognosen abzuleiten.

Die Höhlenbewohner können zwar die Erscheinungen wahrnehmen, nicht aber die Wirklichkeit selbst. Die Erscheinungen bezeichnet man als die PHÄNOMENE und jene Wissenschaft der Höhlenbewohner, die sich an den Phänomenen, also den Erscheinungen, orientiert, ist eine PHÄNOMENOLOGISCHE WISSENSCHAFT.

Die potentielle Energie ist ein Paradebeispiel für einen Begriff aus der phänomenologischen Physik. Wenn beispielsweise in einem Physikhörsaal der Zerfall eines zunächst ruhenden Objekts in zwei gleiche Teile demonstriert wird, die sich nach dem Zerfall mit einer bestimmten Geschwindigkeit voneinander fortbewegen, ist man ganz außer Stande, die Wirklichkeit wahrzunehmen, also den Massendefekt. Wie die Höhlenbewohner aus Platons Gleichnis nimmt man erst einmal nur die Erscheinung wahr, dass die kinetische Energie vor dem Zerfall null war, nachher aber von null verschieden ist. Da der Massendefekt zu klein ist, um wahrgenommen zu werden, sieht es zunächst einmal so aus, als ob da Energie aus dem Nichts entsteht, also als ob der Energiesatz verletzt wäre. Als man die Energieproduktion des Radiums entdeckte, war das für die Physik in der Tat ein Problem dieser Art, weil man den Effekt zunächst einmal auf keine bekannte Phänomenologie der Energie zurückführen konnte.

Wenn man aber durch genauere Beobachtung gewahr wird, dass beim Zerfallsexperiment zwischen beiden Objekten eine Feder gespannt ist und durch noch genauere Beobachtung, dass diese vor und nach dem Zerfall eine Formänderung aufweist, dann kann man eine an dieser Erscheinung, an diesem PHÄNOMEN der Formänderung, orientierte Energiebeschreibung aufbauen. Ähnlich wie die Höhlenbewohner entwickelt man durch die Einführung der potentiellen Energie eine Wissenschaft von den uns zugänglichen Schattenbildern der Wirklichkeit (Diese Wirklichkeit ist die auf den Massendefekt beruhende Änderung der inneren Energie!). Die nach der Wechselwirkung auftretende kinetische Energie wird einem phänomenologischen Zusammenhang zwischen der potentiellen Energie E_{pot} und der Dehnung x der Feder oder sonst einem geeigneten, weil für den experimentalphysikalischen „Höhlenbewohner“ beobachtbaren Indikator zugeschrieben. Die Bemühungen um den phänomenologischen Begriff der potentiellen Energie dauerten viele Jahrzehnte. Erst Albert Einstein gelang es 1905, die hinter dieser Erscheinung stehende „Realität“, d.h. den Massendefekt, zu erkennen. Die physikalische Methode benötigt

aufmerksame Menschen, denen Phänomene überhaupt erst einmal auffallen und die sie kommunizieren; Experimentalphysiker, welche die Zusammenhänge zwischen den Phänomenen erforschen und quantifizieren; theoretische Physiker, welche phänomenologische Theorien für die aufgefundenen Gesetzmäßigkeiten entwerfen; und schließlich jene herausragenden Genies, denen es gelingt, von dieser *Physik der Bilder unserer Wirklichkeit* auf eine höhere Stufe der Annäherung an die Realität aufzusteigen, wie es etwa Albert Einstein, Mohammad Abdus Salam, Murray Gell-Mann, Chen Ning Yang oder Peter Higgs gelang. Von den Schatten, den Phänomenen, führt kein bruchloser logischer Weg zu den Erkenntnissen dieser Genies, sondern meist nur eine heuristische Eingebung, ein Geistesblitz. Je nachdem, ob sich die heuristische Idee bei allen späteren Prüfungen als tragfähig, widerspruchsfrei und erfolgreich erweist, wird sie in die Sammlung der Erkenntnisse der Physik über die „Realität“ aufgenommen oder verworfen. Nun muss man sich leider eingestehen, dass dabei naturgemäß nie klar ist, ob nicht auch diese „Realität“ sich eines Tages als eine nur phänomenologische erweist, also selbst wieder ein Schattenbild einer uns noch immer nicht bekannten Wirklichkeit darstellt.

13. Differentialgleichungen (M2)

Oft möchte man eine Funktion in Abhängigkeit von ihrer Variablen bestimmen, für welche, wie z.B. mit Gl. 12.32, nur ein allgemeiner Zusammenhang zwischen der Funktion und ihren Ableitungen (hier $v = \dot{h}$) vorliegt, d.h. eine Differentialgleichung.

Im einfachsten Fall einer Differentialgleichung kennt man die Ableitung und muss dafür die Stammfunktion finden, wie z.B. bei Gl. 10.2. In der für Differentialgleichungen üblichen Form lautet sie:

$$\dot{x} - v_0 = 0 \tag{13.1}$$

Ihre Lösung ist $x(t) = v_0 t + C$ mit einer freien Konstanten C . Die Anzahl der freien Konstanten entspricht der Ordnung der Differentialgleichung. Gl. 13.1 ist eine Differentialgleichung erster Ordnung und hat also genau eine freie Konstante. Da man diese Differentialgleichung durch Integration lösen kann, bedeutet es im Physikerjargon dasselbe, wenn man davon spricht, dass man eine Differentialgleichung *löst* oder sagt, dass man eine Differentialgleichung *integriert*.

Physiker widmen sich dem Ratespiel „Differentialgleichung lösen“ mit der gleichen Inbrunst und Leidenschaft wie Normalsterbliche, die über die vier Grundrechenarten noch nicht hinausgekommen sind, dem Sudoku. Ein wichtiger vorbereitender Schritt zum Auffinden der Lösung besteht darin, die Differentialgleichung zu klassifizieren:

Art: Differentialgleichungen von Funktionen nur einer Variablen heißen GEWÖHNLICHE DIFFERENTIALGLEICHUNGEN. Treten auch partielle Ableitungen auf (s. Kap. 19), handelt es sich um PARTIELLE DIFFERENTIALGLEICHUNGEN.

Ordnung: Die höchste auftretenden Ableitung bestimmt die ORDNUNG DER DIFFERENTIALGLEICHUNG. Die Anzahl der freien (unbestimmten) Konstanten einer Lösung der Differentialgleichung ist gleich ihrer Ordnung.

	Differentialgleichung	Lösung	Konstanten
DG1	$y' = \sqrt{1 - y^2}$	$y = \sin(x + \varphi)$	φ
DG2	$y' + y = 0$	$y = b \exp(-x)$	b
DG3	$y'' + y' = 0$	$y = a + b \exp(-x)$	a, b
DG4	$y'' - y = 0$	$y = c \exp(x) + d \exp(-x)$ $= a \cosh(\gamma x) + b \sinh(\gamma x)$	c, d a, b
DG5	$y'' + y = 0$	$y = a \cos x + b \sin x$	a, b

Tabelle 5: Wichtige Differentialgleichungen und ihre Lösungen.

Homogenität: Eine Differentialgleichung ist HOMOGEN, wenn die Nullfunktion ($y = f(x) = 0$) zu den Lösungen gehört. Andernfalls ist sie INHOMOGEN.

Linearität: Sind f und g zwei Lösungen einer Differentialgleichung und a eine skalare Konstante, so ist die Differentialgleichung LINEAR, wenn auch $af + g$ eine Lösung ist, ansonsten ist sie NICHTLINEAR. Treten Funktion und Ableitungen nur in erster Potenz auf, ist die Differentialgleichung linear.

Die Lösungen einiger für die Physik wichtiger Differentialgleichungen erster und zweiter Ordnung sind in Tab. 5 angegeben. DG1 ist eine inhomogene nichtlineare Differentialgleichung, alle anderen sind homogen und linear. Homogene Differentialgleichungen erster Ordnung haben eine Fundamentallösung und entsprechend eine freie Konstante, solche zweiter Ordnung deren zwei. Fundamentallösungen sind linear unabhängige Lösungsfunktionen, welche sich linear zur allgemeinen Lösung zusammensetzen. Man verifiziert die Lösungen, indem man einfach in die Differentialgleichung einsetzt. Die darin auftretenden freien Konstanten werden meist durch zusätzliche Bedingungen festgelegt, die an die Lösung gestellt werden.

Zu den Lösungen kann man auf folgende Weise gelangen: Formt man DG1 zu $y^2 + y'^2 = 1$ um, erkennt man die Lösung durch Vergleich mit Tab. 3 und der bekannten Beziehung $\sin^2 x + \cos^2 x = 1$. Die Lösungen von DG2–DG5 kann man ebenfalls unmittelbar durch Vergleich mit Tab. 3 erraten. Dass die konstante Funktion $y = a$ eine Lösung von DG4 ist, ist offensichtlich.

14. Bewegungsgleichungen mechanischer Systeme

In diesem Kapitel werden eindimensionale Bewegungen mechanischer betrachtet, d.h. Bewegungen, die durch eine Raumkoordinate $x(t)$ als Funktion der Zeit charakterisiert werden können. Für einige exemplarische Fälle wird gezeigt, wie die Beschreibung der Bewegung aus Energiemodellen hergeleitet werden kann. Dabei wird stillschweigend vorausgesetzt, dass die Impulserhaltung durch Wechselwirkung mit einer unendlich großen Masse erfüllt ist und somit für die Energiebilanz ignoriert werden kann (s. Kap. 15.2).

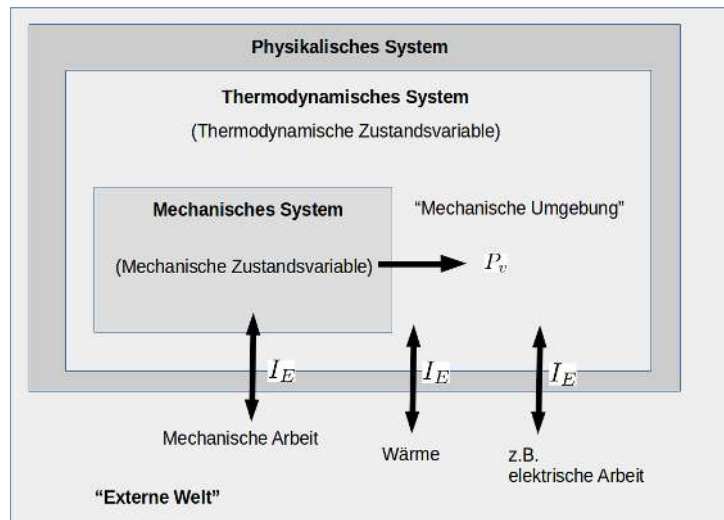


Abbildung 39: Einordnung des mechanischen System-Modells in das Modell für ein physikalisches System

Mechanische Systeme. Mechanische Systeme werden durch ein energetisches Modell definiert, das nur mechanische Energiebeiträge berücksichtigt, d.h. kinetische und potentielle Energiebeiträge. Die innere Energie \mathcal{U} eines physikalischen Systems besteht also aus seinen (wenigen) mechanischen Energiebeiträgen plus den (vielen) Energiebeiträgen, die nicht explizit durch kinetische oder potentielle Energien phänomenologisch gegeben sind (Abb. 39). Die extrem vielen nicht-mechanischen Freiheitsgrade sind und bleiben im mechanischen Modell „unsichtbar“. Sie zählen für ein mechanisches System gewissermaßen zu seiner „Umgebung“. Man kann die Wechselwirkung eines mechanischen Systems mit diesen Beiträgen durch phänomenologische kinetische Gesetze (Dissipationskinetik, P_v) beschreiben. Was die Energiebeiträge eines mechanischen Systems selbst anbetrifft, so sind alle Beiträge durch mechanische Zustandsvariable (z.B. Positionskordinaten, Winkelkoordinaten, Volumenänderungen, Geschwindigkeitskoordinaten, Drehimpulse usw.) charakterisiert. Wenn in einem abgeschlossenen mechanischen System eine Zustandsvariable und damit ein Energiebeitrag abnimmt, dann nimmt irgendeine andere Zustandsvariable und ihr Energiebeitrag zu. Das ist wegen der Energieerhaltung ein bloßes hin- und herbuchen ohne dass eine Richtung der Energieübertragung ausgeschlossen bzw. von einem Verbot betroffen wäre. Man sagt dazu, dass die Prozesse im abgeschlossenen mechanischen System reversibel sind. Abgeschlossene mechanische Systeme sind i.A. nur für eine begrenzte Zeitspanne akzeptable Modelle für reale physikalische Systeme. Der Grund ist, dass mechanische Modelle nur eine winzige Teilmenge aller tatsächlichen Freiheitsgrade eines physikalischen Systems darstellen. Der Übergang von Energie aus dem mechanischen System in die anderen energetischen Freiheitsgrade wird durch die Gesetze der irreversiblen Dissipationskinetik beschrieben. Ist das physikalische System nicht abgeschlossen, kann Energie von außen zufließen oder nach außen abfließen (I_E).

14.1. Bewegungsverlauf bei Dissipation

Dissipationskinetik nullter Ordnung. Ist die Wechselwirkung eines Körpers bzw. mechanischen Systems mit seiner Umgebung rein dissipativ und verläuft sie gemäß Gl. 12.36, liegt eine einfache Differentialgleichung erster Ordnung vor. Durch Integration erhält man für den Betrag der Geschwindigkeit das Zeitgesetz

$$|v(t)| = C - a_0 t.$$

Die freie Integrationskonstante C (s. Gl. 9.9) kann man beispielsweise dadurch festlegen, dass man zum Zeitpunkt $t = 0$ die Anfangsbedingung $|v(t = 0)| = v_0$ fordert. Damit ergibt sich

$$|v(t)| = \begin{cases} v_0 - a_0 t & \text{für } t \leq v_0/a_0 \\ 0 & \text{für } t \geq v_0/a_0 \end{cases}. \quad (14.1)$$

Sobald $|v(t)|$ auf null gesunken ist, hört die Relativbewegung auf (Beträge können nicht negativ werden) und der Körper verbleibt im Ruhezustand relativ zu seiner Umgebung. Seine anfängliche kinetische Energie wurde durch Dissipation vollständig aufgezehrt. Dieses Verhalten des Geschwindigkeitsbetrags entspricht der in Abb. 38 dargestellten Beobachtung.

Nach einer weiteren Integration erhält man aus Gl. 14.1 mit der Anfangsbedingung $x(0) = 0$ die zurückgelegte Wegstrecke

$$x(t) = \begin{cases} v_0 t - \frac{1}{2} a_0 t^2 & \text{für } t < v_0/a_0 \\ \frac{1}{2} a_0 v_0^2 = \text{const.} & \text{für } t \geq v_0/a_0 \end{cases}$$

Dissipationskinetik erster Ordnung. Gl. 12.37 ist eine lineare homogene Differentialgleichung erster Ordnung mit einem konstanten Koeffizienten. Man kann sie u.a. mit der Methode der Trennung der Variablen lösen: Man integriert $dv/v = -a_1 dt$ auf beiden Seiten der Gleichung und erhält $\ln(v/v_0) = -a_1 t$ mit einer freien positiven Konstanten v_0 bzw.

$$v(t) = v_0 \exp(-a_1 t).$$

Wenn man $\dot{x}(t) = v(t)$ noch einmal integriert, erhält man für den Weg

$$x(t) = x_\infty [1 - \exp(-a_1 t)] = 2x_\infty \exp(-\gamma_D t) \sinh(\gamma_D t), \quad (14.2)$$

wobei $\gamma_D = a_1/2$. Wenn die Bewegung bei $x(0) = 0$ mit der Geschwindigkeit $v_0 = |v(t = 0)|$ startet, ist x_∞ die für $t \rightarrow \infty$ asymptotisch erreichbare Entfernung.

14.2. Freier Fall

In den nachfolgenden Rechnungen zum freien Fall soll vom Auftrieb abgesehen werden (s. Kap. 18.3).

Ohne Dissipation. Im dissipationsfreien Grenzfall ist $g_{eff} = g$. Ausgangspunkt ist Gl. 12.32, wobei hier die Bezeichnungen $x = h$ und $\dot{x} = v$ verwendet werden. Die resultierende Differentialgleichung

$$\dot{x}^2 = 2gx. \quad (14.3)$$

soll für die Anfangsbedingungen $\dot{x}(0) = v_0$ und $x(0) = x_0$ gelöst werden. Hier sollen zwei Wege aufgezeigt werden, wie man zur Lösung gelangen kann:

1. Man schreibt Gl. 14.3 als nichtlineare homogene Differentialgleichung erster Ordnung auf:

$$\frac{dx}{dt} - \sqrt{2g} x^{1/2} = 0. \quad (14.4)$$

Auf diese kann man das Standardverfahren der Trennung der Variablen anwenden, d.h. man trennt nach den Variablen x und t ,

$$\frac{1}{\sqrt{2g}} \frac{dx}{x^{1/2}} = dt$$

und integriert beide Seiten, d.h. man ermittelt die Stammfunktion:

$$\frac{1}{\sqrt{2g}} \int x^{-1/2} dx = \frac{2}{\sqrt{2g}} x^{1/2} = \int dt = t + C$$

mit einer aus den Anfangsbedingungen zu berechnenden Integrationskonstanten C . Durch Quadrieren beider Seiten erhält man

$$x(t) = \frac{1}{2}g(t + C)^2. \quad (14.5)$$

Aus der Anfangsbedingung $x(0) = x_0$ folgt $C^2 = 2x_0/g$ und wegen Gl. 14.4 ergibt sich $v_0 = \sqrt{2gx_0}$ und somit

$$x(t) = \frac{1}{2}gt^2 + v_0t + x_0. \quad (14.6)$$

2. Differenziert man die rechte und linke Seite von Gl. 14.3 nach der Zeit, erhält man

$$2\ddot{x}\dot{x} = 2g\dot{x}$$

Die Beschleunigung

$$\ddot{x} = g \quad (14.7)$$

ist gleichförmig. Indem man beide Seiten dieser inhomogenen Differentialgleichung zweiter Ordnung integriert, erhält man

$$\dot{x} = gt + C_1$$

und durch nochmalige Integration

$$x(t) = \frac{1}{2}gt^2 + C_1t + C_2,$$

Wie für Differentialgleichungen zweiter Ordnung üblich, hängt die Lösung von zwei freien Integrationskonstanten C_1 und C_2 ab. Aus den Anfangsbedingungen $x(0) = x_0$ und $\dot{x}(0) = v_0$ folgt schließlich

$$x(t) = \frac{1}{2}gt^2 + v_0t + x_0.$$

Mit Dissipation. Bei einem realistischen freien Fall bedient die potentielle Energie der Schwere einerseits die kinetische Energie und andererseits wird ein Teil der von ihr gelieferten Energie dissipiert:

$$\begin{aligned} \frac{d}{dt}E_{pot} + \frac{d}{dt}E_{kin} + P_v &= 0 \\ -mgv + mv\dot{v} + P_v &= 0 \end{aligned}$$

Hier ist die dissipative Verlustleistung P_v einzusetzen. Die Geschwindigkeit bezieht sich auf die Erde als Ruhesystem. Diese ist hier zugleich Referenz für die relative dissipative Bewegung.

1. Für Dissipationskinetik nullter Ordnung mit $P_v = -a_0m|v|$, wie sie für die Bewegung eines Wagens auf einer schiefen Ebene in guter Näherung vorliegt, betrachtet man eine positive Bewegungsrichtung. Dann hat man die Differentialgleichung

$$\dot{v} = (g - a_0)v = g_{eff}v$$

mit $g_{eff} = g - a_0$ zu lösen. Dissipation nullter Ordnung wirkt sich so aus als ob sich der Schwereparameter von g auf g_{eff} verringert. Bei Dissipationskinetik nullter Ordnung wächst die Geschwindigkeit unbeschränkt (Abb. 37b).

2. Falls eine Dissipationskinetik erster Ordnung mit $P_v = a_1mv^2$ vorliegt, erhält man die inhomogene Differentialgleichung

$$\dot{v} + a_1v = g.$$

Sie hat die spezielle Lösung $v_i = g/a_1$. Zusammen mit der Lösung $v_h = v_0 \exp(-a_1t)$ der homogenen Differentialgleichung $\dot{v} + a_1v = 0$ mit dem freien Parameter v_0 bildet sie die allgemeine Lösung $v(t) = v_i + v_h = g/a_1 + v_0 \exp(-a_1t)$. Legt man $v(0) = 0$ als Anfangsbedingung $v(0) = 0$ fest, folgt

$$v(t) = v_\infty[1 - \exp(-a_1t)].$$

Die Geschwindigkeit wächst hier zunächst an, strebt für $t \rightarrow \infty$ aber der Grenzggeschwindigkeit $v_\infty = g/a_1$ zu. Für große Zeiten wird die Bewegung praktisch stationär mit der Grenzggeschwindigkeit v_∞ .

14.3. Schwingungen

In diesem Abschnitt wird das energetische Modell eines mechanischen Systems untersucht, dem zwei Beiträge zuerkannt werden: eine kinetische Energie und eine Hookesche potentielle Energie. Ferner sollen noch dissipative Energieverluste berücksichtigt werden:

$$\frac{d}{dt} \left(\frac{1}{2} m \dot{x}^2 + \frac{1}{2} k x^2 \right) + P_v = 0 \quad (14.8)$$

Um ein konkretes Beispiel vor Augen zu haben, kann man sich einen Wagen vorstellen, der auf einer Rollbahn zwischen zwei Federn elastisch gespannt ist. Wenn man ihn aus der Ruhelage auslenkt und dann loslässt, schwingt er periodisch um eine Gleichgewichtslage (d.i. die ursprüngliche Ruhelage). Es ist leicht einsehbar, dass man mindestens die drei genannten Energiebeiträge braucht, um das System zu beschreiben. Zur weiteren Vereinfachung sei angenommen, dass die Energieverluste eine Kinetik erster Ordnung folgen:

$$\frac{d}{dt} \left(\frac{1}{2} m \dot{x}^2 + \frac{1}{2} k x^2 \right) + a_1 m \dot{x}^2 = 0 \quad (14.9)$$

Das ist sehr reduktionistisch, denn das ganze komplexe makroskopische System wird nur durch drei Modellparameter beschrieben: eine Masse m , eine Hookesche Federkonstante k und eine Dissipationskonstante a_1 . Dennoch kann man aus dem Modell viel lernen, nämlich zuvorderst das Auftreten des neuen Bewegungstyps der SCHWINGUNGEN. Im Rahmen einer ausreichend großen Messunsicherheit kann man überraschend viele schwingungsfähige Systeme, denen man in der Praxis begegnet, damit bereits zufriedenstellend beschreiben. Ein schwingungsfähiges System nennt man ganz allgemein einen OSZILLATOR. Immer wenn man den Bewegungstyp einer Schwingung in einem Phänomen erkennt oder im Experiment beobachtet, kann man versuchen, die drei Modellparameter aus den experimentellen Daten zu extrahieren und damit eine Beschreibung des Systemverhaltens zu liefern.

14.3.1. Idealer harmonischer Oszillator

Mit abnehmender Dissipation wird die Rate, mit welcher die mechanischen Energien abnehmen, immer geringer und im Grenzfall $a_1 \rightarrow 0$ schließlich null. Der dissipationsfreie Idealfall $P_v = 0$ ist ein abgeschlossenes mechanisches System und heißt IDEALER HARMONISCHER OSZILLATOR. Gl. 14.9 lässt sich in diesem Fall sofort integrieren:

$$\frac{1}{2} m \dot{x}^2 + \frac{1}{2} k x^2 = E_0. \quad (14.10)$$

Die Integrationskonstante E_0 ist die erhalten bleibende mechanische Gesamtenergie. Beispielsweise kann man dem Gesamtsystem durch eine Anfangsauslenkung x_0 zur Zeit $t = 0$ eine bestimmte Energiemenge E_0 extern zuführen, die zuerst einmal als rein potentielle Energie vorliegt:

$$E_0 = E_{pot}(t = 0) = \frac{1}{2} k x_0^2$$

Mit einsetzender Bewegung nimmt die potentielle Energie ab, während die kinetische zunimmt, bis die potentielle Energie vollends aufgebraucht ist. Anschließend pöppelt die kinetische Energie des Systems wieder den potentiellen Energieanteil auf, bis sie wiederum aufgebraucht ist, und die ganze Energie nun als potentielle Energie vorliegt. Dann beginnt das Spiel von neuem. Gl. 14.10 ist eine inhomogene nichtlineare Differentialgleichung erster Ordnung. Man kann sie übersichtlicher machen, indem man die Konstanten $x_0 = \sqrt{2E_0/k}$ und $\omega_0 = \sqrt{k/m}$ einführt,

$$\frac{1}{\omega_0^2} \dot{x}^2 + x^2 = x_0^2,$$

und sie durch die Substitution $\xi = x/x_0$ auf die Form

$$\frac{1}{\omega_0} \frac{d\xi}{dt} = \sqrt{1 - \xi^2}$$

bringt. Man ordnet nach gleichen Variablen und sucht durch die Integration

$$\int \frac{d\xi}{\sqrt{1 - \xi^2}} = \omega_0 \int dt + \delta,$$

beider Gleichungsseiten die Stammfunktionen auf, wobei die freie Integrationskonstante hier mit δ bezeichnet wurde. Mit Gl. 9.7 erhält man

$$\arcsin \xi = \omega_0 t + \delta$$

bzw. die allgemeine Lösung

$$x(t) = x_0 \sin(\omega_0 t + \delta). \quad (14.11)$$

Der Vorfaktor der periodischen Funktion – im hier vorliegenden Spezialfall ist das die Konstante x_0 – heißt AMPLITUDE und δ wird als PHASE bezeichnet. Verlangt man als Anfangsbedingung dass $x(0) = x_0$, so ist $\delta = \pi/2$ und somit

$$x(t) = x_0 \cos \omega_0 t. \quad (14.12)$$

Dies stellt eine ungedämpfte harmonische Schwingung dar. Ihr funktionaler Verlauf ist in Abb. 40a gezeigt. Die Zeitspanne zwischen zwei gleichen Zuständen des Systems bezeichnet man als PERIODE bzw. PERIODENDAUER T . Ihr Kehrwert $\nu = 1/T$ gibt an, wie oft der gleiche Zustand pro Zeiteinheit eintritt, also die Häufigkeit bzw. FREQUENZ, mit welcher der gleiche Zustand pro Zeiteinheit eintritt. Der gleiche Zustand tritt auf, wenn das Argument der Kosinusfunktion gerade einen Winkel von 2π durchlaufen hat, wenn also $\omega_0 T = 2\pi$. Die mit dem Faktor 2π multiplizierte Frequenz nennt man die KREISFREQUENZ

$$\omega_0 = 2\pi\nu = \sqrt{k/m}. \quad (14.13)$$

Wenn keine Verwechslungsgefahr besteht und wenn aus dem Zusammenhang klar ist, dass man mit ω_0 die Kreisfrequenz meint, spricht man auch etwas schlampig von der „Frequenz“ ω_0 .

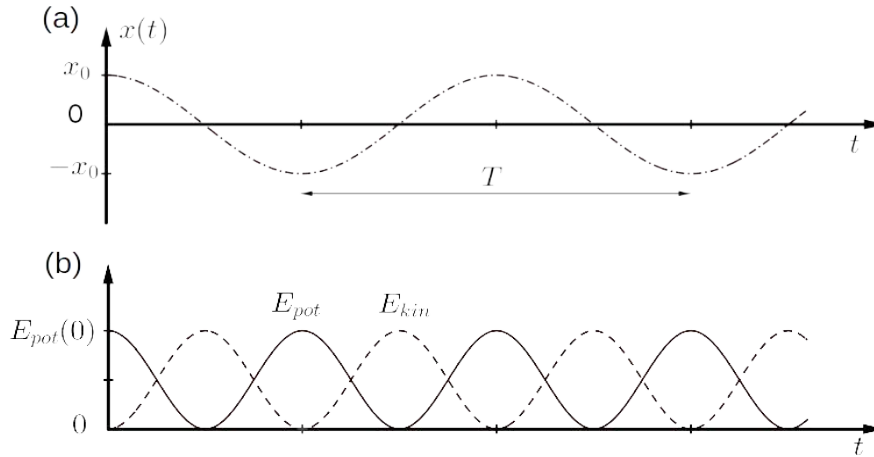


Abbildung 40: Verlauf der Auslenkung $x(t)$, der kinetischen Energie $E_{kin}(t)$ und der potentiellen Energie $E_{pot}(t)$ eines harmonischen Oszillators als Funktion der Zeit t .

Ein idealer Oszillator, der eine Schwingung der Form der Gl. 14.12 ausführt, wird als HARMONISCHER OSZILLATOR bezeichnet. Harmonisch heißt die Schwingung, weil eine Kosinus- bzw. Sinusfunktion auftritt. In Gl. 12.30 trat die Größe $\sqrt{k/m}$ als Verhältnis von Endgeschwindigkeit und Anfangsauslenkung auf. Das ist der für die Umwandlung von potentieller in kinetische Energie bestimmende Parameter. Der zeitliche Verlauf der potentiellen und der kinetischen Energie für den harmonischen Oszillator ist in Abb. 40b gezeigt. Sie oszillieren mit einer doppelt so großen Frequenz wie die Amplitude.

Als weiteres Beispiel für einen idealen Oszillator soll hier noch das dissipationsfreie Pendel diskutiert werden. Als physikalisches Modell denkt man sich einen punktförmigen Körper, der sich unter dem Einfluss der Schwere auf einem Kreis um einen Aufhängepunkt bewegen kann. Das reduktionistischste mechanische Energiemodell dazu heißt MATHEMATISCHES PENDEL. Es besteht aus der kinetischen Energie $E_{kin} = \frac{1}{2}m\dot{x}^2$ und der Schwereenergie $E_{pot} = mgh$ als potentielltem Energiebeitrag. Für die Beschreibung werden nur eine Masse m und ein Abstand l vom Aufhängepunkt als Modellparameter zugrunde gelegt, d.h. beim mathematischen Pendel wird so idealisiert, dass der Körper durch eine sich auf einem Kreis im Abstand l bewegendende Punktmasse m repräsentiert wird. Als phänomenologische Koordinate wählt man den Winkel φ der Auslenkung aus der Ruhelage, der mit der Höhe h gemäß $h = l - l \cos \varphi \approx l\varphi^2/2$ zusammenhängt. Mit der Bahngeschwindigkeit $\dot{x} = l\dot{\varphi}$ auf dem Kreis ergibt sich für die Gesamtenergie:

$$\begin{aligned} \frac{1}{2}m\dot{x}^2 + mgh &= E_0 \\ \frac{1}{2}ml^2\dot{\varphi}^2 + \frac{1}{2}mgl\varphi^2 &= E_0 \end{aligned} \quad (14.14)$$

Durch die Koordinatentransformation von h nach φ wurde die potentielle Energie von der

Höhe auf einen Winkel als Zustandskoordinate umgeschrieben und durch den Wechsel von \dot{x} nach $\dot{\varphi}$ wurde die kinetische Energie auf die zeitliche Änderung dieses Winkels bzw. die Winkelgeschwindigkeit als Zustandskoordinate umgeschrieben. Dadurch erkennt man durch einfachen Vergleich von Gl. 14.14 mit Gl. 14.10 und ohne viel Rechnung sofort die Lösung: Ein mathematisches Pendel schwingt gemäß

$$\varphi(t) = \varphi_0 \cos \omega_0 t \quad (14.15)$$

mit der Kreisfrequenz

$$\omega_0 = \sqrt{g/l}. \quad (14.16)$$

Gl. 14.16 zeigt die Möglichkeit auf, dass man den Schwereparameter g aus der Schwingungsfrequenz eines Pendels ermitteln kann.

14.3.2. Die Schwingungsgleichung

Analog zum freien Fall (Abschnitt 14.2) kann die nichtlineare Differentialgleichung erster Ordnung, Gl. 14.10, in eine lineare Differentialgleichung überführt werden, indem man sie unter Beachtung der Produktregel einmal nach der Zeit differenziert. Mit $\dot{x} \neq 0$ ergibt sich (der Fall $\dot{x} = 0$ ist trivial):

$$\begin{aligned} m\dot{x}\ddot{x} + k\dot{x}x &= \dot{E}_0 = 0 \\ \ddot{x} + (k/m)x &= 0 \end{aligned}$$

Das ist einerseits ein Schritt zurück, denn es wurde eine Differentialgleichung erster Ordnung gegen eine Differentialgleichung zweiter Ordnung eingetauscht, aber andererseits ein Schritt vorwärts, denn es wurde eine nichtlineare Differentialgleichung in eine lineare Differentialgleichung umformuliert. Man kann sich leicht davon überzeugen, dass Gl. 14.12 eine Lösung der Differentialgleichung

$$\ddot{x} + \omega_0^2 x = 0 \quad (14.17)$$

ist, wenn $\omega_0 = \sqrt{k/m}$ gewählt wird. Diese Differentialgleichung ist eine weitere sehr wichtige Modellgleichung der Physik und heißt SCHWINGUNGSGLEICHUNG. Sie beschreibt ideale Oszillatoren.

Eine lineare homogene Differentialgleichung n -ter Ordnung hat n linear unabhängige Lösungen. Gl. 14.17 hat also zwei unabhängige Lösungen. Diese sind $\cos \omega_0 t$ und $\sin \omega_0 t$. Die allgemeinste Lösung ist die Linearkombination

$$x(t) = a \cos \omega_0 t + b \sin \omega_0 t$$

mit zwei Skalaren a und b . Die Skalare werden i.A. durch Anfangsbedingungen festgelegt. Wenn man z.B. Ort und Geschwindigkeit zur Zeit $t = 0$ mit x_0 und v_0 festlegt, lautet die Lösung von Gl. 14.17

$$x(t) = x_0 \cos \omega_0 t + v_0 \sin \omega_0 t$$

Für $v_0 = 0$ ergibt sich die mit Gl. 14.12 bereits aufgefundene Lösung.

14.3.3. Gedämpfter harmonischer Oszillator

Durch Differentiation erhält man aus Gl. 14.9 die Modellgleichung des GEDÄMPFTEN HARMONISCHEN OSZILLATORS:

$$\ddot{x} + 2\gamma_D \dot{x} + \omega_0^2 x = 0 \quad (14.18)$$

Das Systemverhalten wird durch zwei Modellparameter bestimmt: die konstante (Kreis-)Frequenz $\omega_0 = \sqrt{k/m}$ des idealen Oszillators und die ABKLINGKONSTANTE $\gamma_D = a_1/2$. Gl. 14.18 ist eine homogene lineare Differentialgleichung zweiter Ordnung. Sie kann durch den PRODUKTANSATZ $x = fg$ mit zwei zu bestimmenden Funktionen $f(t)$ und $h(t)$ gelöst werden. Man setzt

$$\begin{aligned} \dot{x} &= \dot{f}g + f\dot{g} \\ \ddot{x} &= \ddot{f}g + 2\dot{f}\dot{g} + f\ddot{g} \end{aligned}$$

in Gl. 14.18 ein und erhält

$$f\ddot{g} + 2(\dot{f} + \gamma_D f)\dot{g} + (\ddot{f} + 2\gamma_D \dot{f} + \omega_0^2 f)g = 0.$$

Man verfügt über die Funktion f so, dass der mittlere Term verschwindet. Die Gleichung $\ddot{f} + \gamma_D \dot{f} = 0$ wird durch $f(t) = \exp(-\gamma_D t)$ gelöst. Damit ist für g nur noch die Differentialgleichung

$$\ddot{g} + (\omega_0^2 - \gamma_D^2)g = 0$$

zu lösen. Ist das Vorzeichen von $\omega_0^2 - \gamma_D^2$ negativ, so liegt eine Differentialgleichung vom Typ DG4 vor, ist es positiv, eine vom Typ DG5 (Tab. 5).

Der Kriechfall ($\gamma_D > \omega_0$). Mit

$$\ddot{g} - \gamma^2 g = 0 \quad \text{und} \quad \gamma = \sqrt{\gamma_D^2 - \omega_0^2} \quad (14.19)$$

folgt die allgemeine Lösung

$$x(t) = [a \exp(\gamma t) + b \exp(-\gamma t)] \exp(-\gamma_D t). \quad (14.20)$$

des sogenannten KRIECHFALLS. Wenn man als Anfangsbedingung $x(0) = x_0$ und $\dot{x}(0) = 0$ fordert, ergibt sich aus Gl. 14.20 die Lösung

$$x(t) = x_0 \left[\cosh(\gamma t) + \frac{\gamma_D}{\gamma} \sinh(-\gamma t) \right] \exp(-\gamma_D t) \quad (\text{Kriechfall}) \quad (14.21)$$

Der Schwingfall ($\omega_0 > \gamma_D$). Mit

$$\ddot{g} + \omega^2 g = 0 \quad \text{und} \quad \omega = \sqrt{\omega_0^2 - \gamma_D^2} \quad (14.22)$$

folgt die allgemeine Lösung

$$x(t) = x_0 \exp(-\gamma_D t) \sin(\omega t + \delta) \quad (14.23)$$

des SCHWINGFALLS. Es handelt sich um eine gedämpfte Schwingung, deren Amplitude im Laufe der Zeit exponentiell abnimmt.¹⁴

Die DÄMPFUNG des Oszillators geschieht mit der Abklingkonstanten γ_D .

Legt man durch die Anfangsbedingungen $x(0) = x_0$ und $\dot{x}(0) = 0$ nicht nur die Anfangsamplitude x_0 , sondern auch die Anfangsgeschwindigkeit auf null fest, folgt aus Gl. 14.23 die Lösung

$$x(t) = x_0 \left[\cos(\omega t) + \frac{\gamma_D}{\omega} \sin(\omega t) \right] \exp(-\gamma_D t) \quad (\text{Schwingfall}) \quad (14.24)$$

Der aperiodische Grenzfall ($\gamma_D = \omega_0$). Für den Grenzfall $\gamma \rightarrow 0$ bzw. $\omega \rightarrow 0$ ergibt sich Lösung

$$x(t) = x_0 [1 + \gamma_D t] \exp(-\gamma_D t) \quad (\text{aperiodischer Grenzfall})$$

des APERIODISCHEN GRENZFALLS, indem man auf Gl. 14.21 oder Gl. 14.24 die Näherungsformeln aus Tab. 2 anwendet.

In Abb. 41 werden die Lösungen für den Schwingfall ($\omega_0 > \gamma_D$), den Kriechfall ($\omega_0 < \gamma_D$) und den aperiodischen Grenzfall ($\omega_0 = \gamma_D$) für jeweils die gleichen Anfangsbedingungen $x(0) = x_0$ und $\dot{x}(0) = 0$ bei festgehaltener Frequenz ω_0 miteinander verglichen. Im aperiodischen Grenzfall nähert sich die Amplitude dem stationären Gleichgewichtswert $x = 0$ am schnellsten an. Da man für Zeigerinstrumente eine möglichst rasche Einstellung der Messwertanzeige anstrebt, bemüht man sich, die Dämpfung so einzustellen, dass man den aperiodischen Grenzfall erreicht.

14.3.4. Energiezufuhr von außen

In diesem Abschnitt sollen zwei Beispiele betrachtet werden, wo Oszillatoren eine Energiezufuhr von außen erfahren. Dazu fügt man Gl. 14.9 einen Energiezustrom I_E von außen hinzu und betrachtet das Energiemodell

$$\frac{d}{dt} \left(\frac{1}{2} m \dot{x}^2 + \frac{1}{2} k x^2 \right) + 2\gamma_D m \dot{x}^2 + I_E = 0$$

¹⁴Anmerkung zur Klassifikation: Die Funktion der gedämpften Schwingung, Gl. 14.23, ist, anders als die ungedämpfte Schwingung, selbstverständlich *keine periodische Funktion*, denn es gilt $x(t) \neq x(t+T)$, wenn $T = 2\pi/\omega$ die Periodendauer darstellt. Wenn man sich den Graphen anschaut, hat die Funktion dennoch etwas „Periodisches“, weil die Maxima periodisch aufeinander folgen. Deshalb zählt man auch die gedämpfte Schwingung zur Klasse der periodischen Bewegungsmuster.

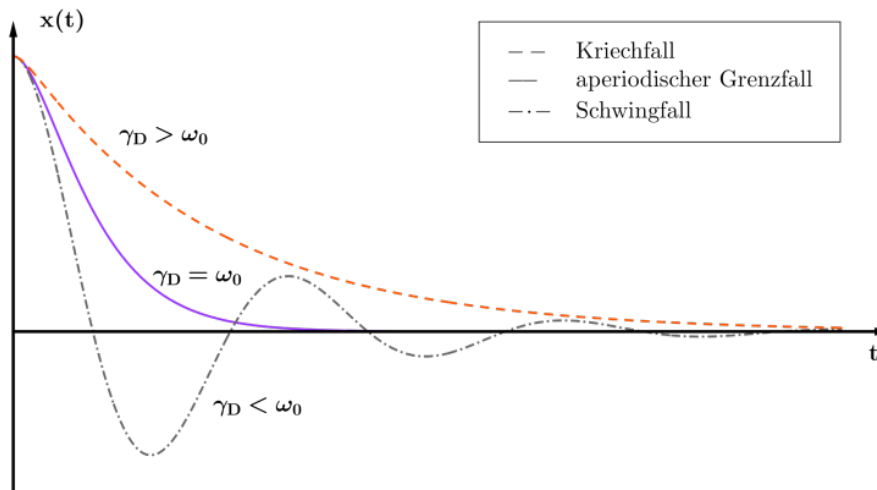


Abbildung 41: Vergleich von Kriechfall, aperiodischem Grenzfall und Schwingfall. Für alle drei Kurven sind die Anfangsbedingungen sowie die Frequenz ω_0 gleich; die Dämpfung wird variiert.

Verstärkung. Als erster soll der Fall betrachtet werden, dass der Energiezustrom durch

$$I_E = -\gamma_G m \dot{x}^2,$$

gegeben ist, d.h. dass Energie nach der gleichen Gesetzmäßigkeit zugeführt wird, wie sie dissipiert wird. Für den Schwingfall erhält man dann die Lösung

$$x(t) = x_0 e^{(\gamma_G - \gamma_D)t} \sin(\omega t + \delta). \quad (14.25)$$

Ist die Konstante γ_G positiv, wird sie als heißt GAIN bzw. VERSTÄRKUNGSKOEFFIZIENT bezeichnet. Ist sie negativ, spricht man von negativer Verstärkung bzw. Dämpfung. Der Unterschied zwischen Dämpfung und Dissipation ist subtil: Wenn eine schwingende Stimmgabel Energie durch Umwandlung in Wärme verliert, dann würde man es als Dissipation ansehen. Der Vorgang ist irreversibel. Wenn sie Schall abgibt, dann ist der Vorgang im Prinzip teilweise reversibel: Man könnte den Schall (teilweise) wieder zurückreflektieren und die Stimmgabel kann wiederum durch Aufnahme von Schallenergie (ein wenig) zum Schwingen gebracht werden.

Für $\gamma_G > \gamma_D$ wächst die Amplitude exponentiell an und für $\gamma_G < \gamma_D$ fällt sie exponentiell ab. Wenn aber $\gamma_G = \gamma_D$, so kann sich ein STATIONÄRER SCHWINGUNGSZUSTAND mit einer konstanten Schwingungsamplitude

$$x(t) = x_s \cos \omega t \quad (14.26)$$

einstellen. Die Einstellung des stationären Zustands ist heikel und labil: Ein bisschen zu hohe Verstärkung und die Amplitude wächst exponentiell, ein bisschen zu wenig und dann nimmt sie mit der Zeit ab.

Erzwungene Schwingung. Von außen werde ein periodischer Energiestrom

$$I_E = -\gamma_G m \dot{x} \cos(\omega t + \delta)$$

zugeführt. Das mag auf ersten Blick wie ein Nullsummenspiel aussehen und wäre es auch, wenn \dot{x} eine Konstante wäre. Denn dann würde eine halbe Periode lang Energie zugeführt und in der nächsten halben Periode genauso viel Energie wieder abgeführt. Aber \dot{x} ist eine Funktion der Zeit und das ändert alles. Hier soll die Frage nach der Bedingung untersucht werden, unter der sich eine stationäre Lösung der Schwingungsgleichung mit

$$x(t) = x_s \cos \omega t$$

einstellen kann. Das lässt sich leicht dadurch beantworten, indem man dies als Ansatz in die Modellgleichung

$$\ddot{x} + 2\gamma_D \dot{x} + \omega_0^2 x = \gamma_G \cos(\omega t + \delta) \quad (14.27)$$

einsetzt und untersucht, ob es für diesen Ansatz eine Lösung gibt.

Wegen $\cos(\omega t + \delta) = \cos \delta \cos \omega t - \sin \delta \sin \omega t$ (Winkeladditionstheorem) und weil Kosinus- und Sinusfunktion linear unabhängig sind, gewinnt man die beiden Gleichungen

$$\begin{aligned} (\omega_0^2 - \omega^2)x_s &= \gamma_G \cos \delta \\ 2\gamma_D \omega x_s &= \gamma_G \sin \delta \end{aligned}$$

Mit $\cos^2 \delta + \sin^2 \delta = 1$ folgt für die stationäre Amplitude

$$x_s = \frac{\gamma_G}{\sqrt{(\omega_0^2 - \omega^2)^2 + (2\gamma_D \omega)^2}}$$

und für die Phasenverschiebung

$$\tan \delta = \frac{2\gamma_D \omega}{(\omega_0^2 - \omega^2)}.$$

Es gibt also tatsächlich eine stationäre Lösung. In Abb. 42 sind einige Kurven für verschiedene Parameter γ_D/ω_0 aufgetragen, welche die Abhängigkeit der Amplitude x_s und des Phasenwinkels δ als Funktion von ω zeigen. Bei $\omega = \omega_0$ liegt der RESONANZFALL vor, bei dem die Amplitude maximal wird. Die Anregung eilt dem Oszillator stets um δ voraus bzw. der Oszillator hinkt der Anregung stets um δ hinterher. In Resonanz ist der Phasenwinkel stets 90° bzw. $\pi/2$.

Für die reversibel zugeführte Leistung erhält man

$$I_E = -\gamma_G m \dot{x} \cos(\omega t + \delta) = \gamma_G m \omega x_s \sin \omega t [\sin \delta \sin \omega t - \cos \delta \cos \omega t]$$

Im zeitlichen Mittel ist der zweite Beitrag auf der rechten Gleichungsseite ($\propto \sin \omega t \cos \omega t = \frac{1}{2} \sin 2\omega t$) null, d.h. über eine Periode gemittelt wird genauso viel Energie zugeführt wie

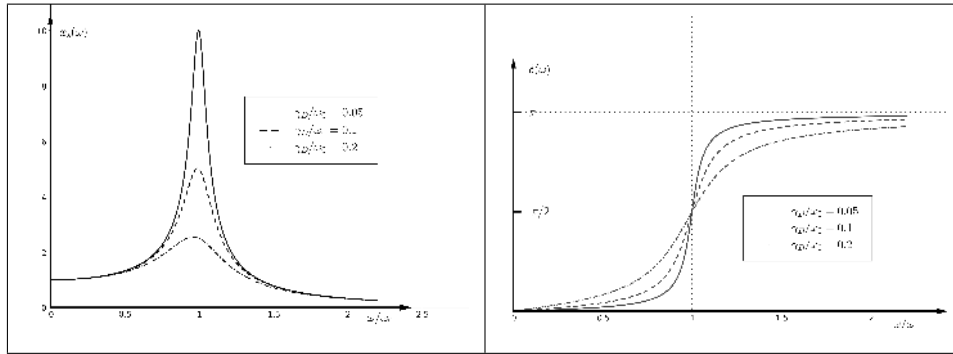


Abbildung 42: Verlauf der stationären Amplitude x_s und der Phasenverschiebung δ als Funktion der externen Frequenz ω für unterschiedliche Dämpfungswerte.

abgeführt. Dies bezeichnet man als **BLINDLEISTUNG**. Der erste Beitrag ($\propto \sin^2 \omega t = \frac{1}{2} - \frac{1}{2} \cos(2\omega t)$) ist hingegen über die Periode gemittelt ungleich null und stellt den tatsächlichen wirksamen Zustrom an Energie dar. Das ist die dem Oszillator zugeführte **WIRKLEISTUNG** P_E . Im stationären Fall stellt sich die Phasenverschiebung δ gerade so ein, dass in jeder Periode die aufgrund der Dissipation erlittene **VERLUSTLEISTUNG** P_v durch den effektiven Energiezustrom bzw. die zuströmende Wirkleistung $P_E = I_E$ gerade kompensiert wird. Am Phänomen der erzwungenen Schwingung ist bemerkenswert, dass die Einstellung der Balance selbstregulierend erfolgt und die stationäre Amplitude gegenüber Störungen stabil ist.

14.3.5. Gekoppelte Oszillatoren

Als konkretes experimentelle Beispiel für zwei **GEKOPPELTE OSZILLATOREN** seien zwei Wagen A und B gleicher Masse m betrachtet. Sie seien durch zwei Hookesche Federn gleicher Federkonstante k an ihre jeweilige Gleichgewichtsposition gebunden und über eine schwache Feder mit der viel kleineren Federkonstanten κ miteinander gekoppelt. Diese Situation kann man durch ein Energiemodell charakterisieren, in welches die kinetischen Energien der beiden Wagen und die drei potentiellen Energien der Federn eingehen. Wenn man die Dissipation vernachlässigen kann, gilt

$$E_0 = \frac{1}{2} m \dot{x}_A^2 + \frac{1}{2} m \dot{x}_B^2 + \frac{1}{2} k x_A^2 + \frac{1}{2} k x_B^2 + \frac{1}{2} \kappa (x_A - x_B)^2 \quad (14.28)$$

mit der konstanten Gesamtenergie E_0 . Hier sind x_A und x_B die Auslenkungen der beiden Wagen aus der jeweiligen Gleichgewichtsposition. Indem man die Summen- und Differenzkoordinaten

$$\begin{aligned} \xi_1 &= \frac{1}{\sqrt{2}} (x_A + x_B) \\ \xi_2 &= \frac{1}{\sqrt{2}} (x_A - x_B) \end{aligned}$$

einführt, kann man Gl. 14.28 folgendermaßen umschreiben:

$$E_0 = \frac{1}{2}m\dot{\xi}_1^2 + \frac{1}{2}k\xi_1^2 + \frac{1}{2}m\dot{\xi}_2^2 + \frac{1}{2}(k + 2\kappa)\xi_2^2$$

Durch bloßes Hinschauen erkennt man, dass dies die Summe der Energien zweier unabhängiger Oszillatoren mit den Frequenzen $\omega_1 = \sqrt{k/m}$ und $\omega_2 = \sqrt{(k + 2\kappa)/m}$ darstellt. Daher kann man die Lösung für deren EIGENSCHWINGUNGEN sofort angeben:

$$\begin{aligned}\xi_1(t) &= \xi_{10} \cos \omega_1 t \\ \xi_2(t) &= \xi_{20} \cos \omega_2 t\end{aligned}$$

Die Konstanten ξ_{10} und ξ_{20} sind dabei aus den Anfangsbedingungen zu bestimmen. Damit ergibt sich z.B.

$$x_A(t) = \frac{1}{\sqrt{2}} [\xi_1(t) + \xi_2(t)] = \frac{1}{2}(x_{A0} + x_{B0}) \cos \omega_1 t + \frac{1}{2}(x_{10} - x_{20}) \cos \omega_2 t$$

Hier sind x_{A0} und x_{B0} die Anfangsauslenkungen. Für den Fall, dass zu Beginn nur der erste Wagen mit $x_{A0} = x_0$ ausgelenkt und der zweiten in seiner Ruhelage ist, d.h. $x_{B0} = 0$, ergibt sich:

$$x_A(t) = \frac{1}{2}x_0 [\cos \omega_1 t + \cos \omega_2 t] = x_0 \cos \Delta\omega t \cos \omega t \quad (14.29)$$

$$x_B(t) = \frac{1}{2}x_0 [\cos \omega_1 t - \cos \omega_2 t] = x_0 \sin \Delta\omega t \sin \omega t \quad (14.30)$$

Hierbei ist $\Delta\omega = \frac{1}{2}(\omega_2 - \omega_1)$. Im Experiment sieht das so aus, dass der erste Wagen wie ein harmonischer Oszillator mit der mittleren Frequenz $\omega = \frac{1}{2}(\omega_1 + \omega_2)$ schwingt, wobei sich zugleich seine Amplitude periodisch ändert. Das Phänomen der periodischen Modulation der Amplitude wird als SCHWEBUNG bezeichnet. Für schwache Kopplung ($\kappa \ll k$) ändert sich die Gesamtenergie

$$E_A = \frac{1}{2}m\dot{x}_A^2 + \frac{1}{2}kx_A^2 \approx \frac{1}{2}kx_0^2 \cos^2 \Delta\omega t = k \left(\frac{x_0}{2}\right)^2 (1 + \cos \Omega t) \quad (14.31)$$

des Oszillators A periodisch mit der Differenzfrequenz $\Omega = \omega_2 - \omega_1 = 2\Delta\omega$, die man als Schwebungsfrequenz bezeichnet. Verläufe von $x_A(t)$ und $x_B(t)$ sind in Abb. 43 dargestellt.

15. Kraft

15.1. Newtonsche Bewegungsgleichung

In diesem Kapitel wird ein mechanisches System betrachtet, dessen energetisches Modell aus einer kinetischen Energie $E_{kin}(\dot{x}) = m\dot{x}^2/2$ und einer potentiellen Energie $E_{pot}(x)$ besteht. Die Zustandskoordinate x soll tatsächlich die Ortskoordinate sein (und nicht etwa eine verallgemeinerte Zustandskoordinate φ wie beim mathematischen Pendel) und

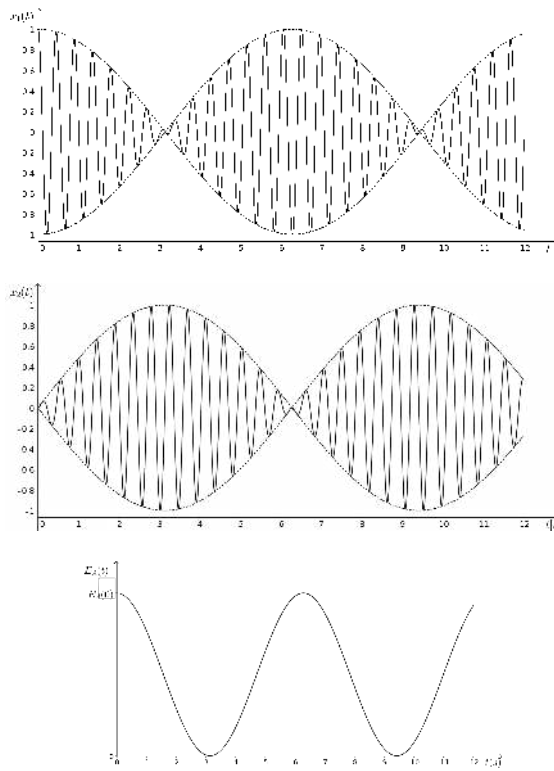


Abbildung 43: Die Amplituden $x_1(t)$ (Gl. 14.29) und $x_2(t)$ (Gl. 14.30) sowie die Energie E_A des Oszillators A (Gl. 14.31) als Funktion der Zeit t ($x_0 = 1$, $\omega = 7\text{s}^{-1}$, $\Delta\omega = 0.5\text{s}^{-1}$). Die einhüllende Schwebungskurve ist gestrichelt gezeichnet.

die Zustandskoordinate \dot{x} soll eine tatsächliche Geschwindigkeitskoordinate sein. Sie soll gleich der Relativgeschwindigkeit zum Ruhesystem des Mediums, das für die Dissipation verantwortlich ist. Das ist das hier gewählte Bezugssystem. Neben der dissipativen Verlustleistung P_v soll auch ein von außen zu- oder abfließender Energiestrom I_E berücksichtigt werden. Für dieses Newtonsche Standardmodell „Punktteilchen“ ergibt sich als Leistungsbilanz:

$$\frac{d}{dt}E_{kin}(\dot{x}) = -\frac{d}{dt}E_{pot}(x) - I_E - P_v. \quad (15.1)$$

Unter Berücksichtigung der Kettenregel für die Differentiation und mit $p = m\dot{x}$ folgt:

$$\dot{p}\dot{x} = -\frac{dE_{pot}}{dx}\dot{x} - P_v - I_E \quad (15.2)$$

Die sich auf die systeminterne Zustandsvariable x beziehende Größe

$$F_{kons}(x) = -\frac{dE_{pot}(x)}{dx} \quad (15.3)$$

wird als **KONSERVATIVE KRAFT** bezeichnet. Diese Bezeichnung hat folgenden Grund: Wenn, wie beim Beispiel des idealen harmonischen Oszillators (Kap. 14.3.1), nur konservative Kräfte auftreten, dann bleibt die aus kinetischer und potentieller Energie bestehende mechanische Energie erhalten bzw. konserviert. Durch die Beziehung

$$I_E = -F_{ext}\dot{x} \quad (15.4)$$

definiert man noch eine **EXTERNE KRAFT** und durch

$$P_v = -F_R v \quad (15.5)$$

eine **REIBUNGSKRAFT**. Für ein Dissipationsgesetz gemäß Gl. 12.39 ergibt sich die Reibungskraft beispielsweise zu

$$F_R = -ma_1 v.$$

Reibungskräfte sind stets entgegengesetzt zur Relativbewegung bezüglich für die Dissipation verantwortlichen Mediums gerichtet. Die äußere Kraft ist keine Funktionen der Zustandsvariablen x des Systems. Sie hat mit dem mechanischen System nichts zu tun. Sie ist äußere Ursache der mechanischen Energieänderungen. In der Newtonschen Mechanik nimmt man diesen Standpunkt auch für die Dissipation ein. Im statischen Fall ($\dot{x} = 0$ bzw. $v = 0$) tritt weder ein Energiestrom noch eine dissipative Verlustleistung auf – egal wie groß die externen Kräfte oder die Reibungskräfte sein mögen.

Die aus Gl. 15.2 folgende Gleichung

$$\dot{p}\dot{x} = (F_{kons} + F_R + F_{ext})\dot{x}$$

ist für alle \dot{x} erfüllbar, wenn

$$\dot{p} = F_{kons} + F_R + F_{ext} = F. \quad (15.6)$$

Die zeitliche Änderung des Impulses ist also gleich einer **KRAFT** F , welche die Summe der konservativen und nichtkonservativen (externen) Kräfte sowie der Reibungskräfte ist. Sind externe Kraft und Reibungskraft null (d.h. $P_v + I_E = 0$), dann bleibt die mechanische Energie erhalten, d.h. die Summe aus kinetischer und potentieller Energie ist dann konstant.

Die Definition der Kräfte über die Energien bringt es mit sich, dass den Kräften eine Erbschaft in den Schoß fällt: Sie erben von der Energie die Additionseigenschaft. Wenn sich also die potentielle Energie oder die zugeflossene Energie aus verschiedenen Einzelbeiträgen additiv zusammensetzen, dann setzen sich auf die Kräfte aus den Einzelbeiträgen additiv zusammen. gilt das auch für die Kräfte. Wenn k Kraftbeiträge beitragen, dann summieren sich k unterschiedliche Kraftbeiträge zu einer resultierenden Kraft F :

$$\dot{p} = \sum_{i=1}^k F_i = F \quad (15.7)$$

Gl. 15.7 ist die NEWTONSCHE BEWEGUNGSGLEICHUNG und drückt die Newtonsche Sicht der Rolle der Kraft aus: Sie ist die Ursache der Bewegungsänderung bzw. Impulsänderung. Das ist im Nachhinein der Grund für die Wahl des negativen Vorzeichens bei der Definition der konservativen Kräfte in Gl. 15.3. In ihre Herleitung gingen die nichtrelativistischen Ausdrücke für die kinetische Energie und den Impuls ein. Wenn Geschwindigkeiten auftreten, deren Betrag in der Größenordnung der Raum/Zeit-Konstante c liegt, liefert Gl. 15.7 daher falsche Resultate. Aber für die „Alltagsphysik“, d.h. für den Geschwindigkeitsbereich unserer Alltagserfahrungen, ist die die Newtonsche Bewegungsgleichung eine ausgezeichnete Näherung. Die Maßeinheit der Kraft heißt NEWTON:

$$1 \text{ Newton} = 1 \text{ N} = 1 \text{ kg m/s}^2. \quad (15.8)$$

Setzt man $p = m\dot{x}$ in Gl. 41.1 für eine konstante Masse m ein, so ergibt sich:

$$m\ddot{x} = F. \quad (15.9)$$

Erst der Umstand, dass die Beschleunigung $a = \ddot{x}$ im Newtonschen Grenzfall $v = \dot{x} \rightarrow 0$ gemäß Gl. 10.6 zu einer Invarianten wird, für zwei beliebige Inertialsysteme S und S' also näherungsweise

$$a \approx a' \quad (15.10)$$

gilt, führt dazu, dass die Gleichung „Kraft gleich Masse mal Beschleunigung“,

$$F = ma, \quad (15.11)$$

im Rahmen der Newtonschen Näherung zu einem in allen Inertialsystemen gültigen physikalischen Gesetz mit einer invarianten Kraft \vec{F} wird. Um es ganz klar zu sagen: Nur für kleine Geschwindigkeiten ist Gl. 15.11 ein (näherungsweise) allgemeines Gesetz.

Wenn man es nicht mit einem einzelnen Teilchen zu tun hat, sondern mit einem System aus N Teilchen, dann addiert man einfach die Energiebeiträge für alle Teilchen auf und erhält so die Energie des Gesamtsystems. Betrachtet man ein bestimmtes Teilchen, das mit allen übrigen Teilchen wechselwirkt, so agieren diese als externe Kräfte. Dabei wirken sich die Zu- und Abflüsse dieser „äußeren“ Energie, die sich die N Teilchen untereinander zuschieben, in Summe aber nicht aus, denn für das Teilchensystem summieren sich positive Energiezuflüsse zu einem Teilchen j mit den von den anderen Teilchen an dieses Teilchen abgegebenen Energieströme zu null. Daher gilt für das Teilchensystem:

$$\dot{p} = \sum_{j=1}^N \dot{p}_j = \sum_{j=1}^N F_j,$$

wobei F_j hier die auf das j -te Teilchen wirkende Summe aller Kräfte ist.

15.2. Actio und reactio

Es sei G ein abgeschlossenes System mit seiner inneren Energie und der aus seiner (äußeren) kinetischen Energie bestehenden äußeren Energie. Die Energieerhaltung (Kap. 11.6), würde es zulassen, dass innere Energie sich in äußere kinetische Energie umwandelt und umgekehrt. Die Impulserhaltung, verhindert diesen „Münchhausen-Trick“¹⁵, d.h. schließt einen solchen Prozess aus.

Eine Änderung der kinetischen Energie ist daher nur möglich, wenn (mindestens) zwei Systeme miteinander wechselwirken. Wenn sich das abgeschlossene System G beispielsweise aus zwei Teilsystemen A und B zusammensetzt, dann ist es möglich, dass Teilsystem A seine kinetische Energie ändert, indem die damit verbundene Impulsänderung dp_A durch eine Impulsänderung dp_B des zweiten Teilsystems B ausgeglichen wird. Wird die Impulsänderung von A gemäß der Newtonschen Bewegungsgleichung

$$\dot{p}_A = F$$

durch eine Kraft F bewirkt, so muss es ein mit A wechselwirkendes System B geben, dessen Impuls sich mit einer in entgegengesetzter Richtung angreifenden Kraft ändert:

$$\dot{p}_B = -F$$

so dass für das aus den beiden wechselwirkenden Teilsystemen bestehende abgeschlossene Gesamtsystem G die Impulserhaltung

$$\dot{p}_A + \dot{p}_B = 0$$

Newton beschrieb dieses Wechselwirkungsprinzip durch *actio gleich reactio*.

Wenn ein abgeschlossenes System aus zwei Teilsystemen A und B mit den Massen m_A und m_B besteht, ist jede Änderung der kinetischen Energie um $dE_{kin,A} = p_A dp_A / m_A$ mit einer Änderung des Impulses um dp_A verbunden ist. Das System B kommt aufgrund der Impulserhaltung daher zwangsläufig ins Spiel. Im Ruhesystem gilt stets $p_A = -p_B$ und mit $dE_{kin,B} = p_B dp_B / m_B$ folgt $dE_{kin,B} / dE_{kin,A} = m_A / m_B$. Falls die Masse m_B gegen unendlich geht, braucht die kinetische Energie von B in der Energiebilanz mit anderen Energiebeiträgen nicht berücksichtigt werden. Da System B liefert dann nur den benötigten Impuls, spielt für die Energiebilanz aber keine Rolle.

Münchhausen-Probleme. Das System A gebe einen differentiellen Anteil dE_{pot} seiner potentiellen Energie ab. Der WIRKUNGSGRAD, mit dem sie in einen Anteil $dE_{kin,A}$ eigener kinetischer Energie umwandeln kann, ist

$$\eta = \frac{dE_{kin,A}}{dE_{pot}} = \frac{dE_{kin,A}}{dE_{kin,A} + dE_{kin,B}} = \frac{1}{1 + m_A/m_B}.$$

¹⁵In einer bekannten Lügengeschichte soll sich Baron Münchhausen am eigenen Schopf aus dem Sumpf gezogen haben.

Für $m_B \rightarrow 0$, also insbesondere, wenn kein Wechselwirkungspartner B vorhanden ist, wird der Wirkungsgrad null. Daher verbietet der Impulserhaltungssatz den „Münchhausen-Trick“, obwohl der Energieerhaltungssatz ihn im Prinzip zulassen würde.

In Abschnitt 14.3.1 wurde davon ausgegangen, dass der harmonische Oszillator seine kinetische Energie periodisch in potentielle Energie und zurück umwandeln kann, also in beide Richtungen mit einem Wirkungsgrad von 100%. Das ist nur möglich, wenn der harmonische Oszillator mit einem weiteren System B wechselwirkt, dessen Masse m_B unendlich groß ist. Dann wird zwar Impuls an B abgegeben, aber keine kinetische Energie. Analog zu Abschnitt 14.3.1 wird dieser Grenzfall stillschweigend vorausgesetzt, wenn man sich auf eine rein energetische Analyse beschränken will.

CERN-Probleme. Das System A gebe einen differentiellen Anteil $dE_{kin,A}$ seiner kinetischen Energie ab. Der Wirkungsgrad, mit dem es sie in einen Anteil dE_{pot} an potentieller Energie umwandeln kann, ist

$$\eta = \frac{dE_{pot}}{dE_{kin,A}} = \frac{dE_{kin,A} - dE_{kin,B}}{dE_{kin,A}} = 1 - \frac{m_A}{m_B}.$$

Wenn man also ein Teilchen mit hoher kinetischer Energie auf ein Targetteilchen schießt, steht nur der Anteil η für eine Umwandlung in innere Energie zur Verfügung. Ein positiver Wirkungsgrad $\eta > 0$ ist nur so lange möglich, wie $m_B > m_A$ gilt.

16. Arbeit

Wenn einem mechanischen System durch einen physikalischen Prozess von außen Energie zuströmt, so bezeichnet man die zugeführte Energie

$$\Delta W = - \int_{t_i}^{t_f} I_E(t) dt \quad (16.1)$$

als (physikalische) ARBEIT. Hier ist t_i der Zeitpunkt, an dem der Prozess begann und t_f der Zeitpunkt, an dem er endete. Im Newtonschen Spezialfall, bei dem man die Bewegung von Teilchen mit ihren Orts- und Geschwindigkeitskoordinaten x und \dot{x} betrachtet, kann man Gl. 15.4 einsetzen und erhält

$$\Delta W = \int_{t_i}^{t_f} F_{ext}(t) \dot{x}(t) dt = \int_{x_i}^{x_f} F_{ext} \cdot dx, \quad (16.2)$$

wobei $x_i = x(t_i)$ die Koordinate zu Beginn und $x_f = x(t_f)$ am Ende des Prozesses ist. Die Arbeit tritt hier als Wegintegral der externen Kraft auf. Wird Arbeit zugeführt, so ändern sich Zustandsvariable des mechanischen Systems, weil die von außen zugeführte Energie von den mechanischen Energiebeiträgen aufgenommen wird. Das führt zu einer

allgemeineren Definition der Arbeit: Jede Aufnahme oder Abgabe von Energie in eine mechanische Energieform geschieht in Form von Arbeit. Das ist ohne Einschränkung in beide Richtungen möglich, also reversibel. Als Zustandsvariable, die sich dabei ändern, müssen nicht notwendigerweise Ort oder Geschwindigkeit erhalten. Jede andere mechanische Zustandsvariable tut es auch.

17. Das Prinzip des Minimums der mechanischen Energie

Wenn keine Beschränkungen es verhindern, kann potentielle Energie eines mechanischen Systems i.A. in kinetische Energie umgewandelt werden. Ist kinetische Energie im System vorhanden, so bedeutet das, dass Systemteile sich relativ zueinander bewegen und somit i.A. Energie dissipiert wird. Der Dissipationsvorgang ist eine Einbahnstraße, d.h. die dissipierte Energie ist für das mechanische System verloren gegangen. Der Vorgang kann so lange stattfinden, bis mechanisches Gleichgewicht herrscht. Damit meint man, dass der Zustand stationär wird und sich keine weiteren Zustandsänderungen mehr abspielen. Im mechanischen Gleichgewicht erreicht die mechanische Energie ihr stationäres Minimum.

Prinzip vom Minimum der mechanischen Energie

Makroskopische Systeme verringern ihre mechanische Energie durch Dissipation so lange, bis es keine erlaubte Zustandsänderung mehr gibt, mit der eine noch niedrigere innere Energie erreicht werden könnte.

Die stabile Endlage, bei welcher die mechanische Energie ein Minimum annimmt, ist der MECHANISCHE GLEICHGEWICHTSZUSTAND.

Der Gleichgewichtszustand ist durch die Werte der Zustandsvariablen im Minimum der mechanischen Energie bestimmt. Er ist gegen Störungen stabil. Wenn man ein mechanisches System durch externe Energiezufuhr in einen NICHTGLEICHGEWICHTSZUSTAND versetzt, so kehrt es durch Dissipationsprozesse nach einiger Zeit wieder in den Gleichgewichtszustand zurück. Die Dissipation führt dadurch eine Richtung der Zeit ein, die in den Grundgleichungen der Mechanik zunächst einmal so nicht auftritt: Die Vorgänge entwickeln sich im Laufe der Zeit stets so, dass das Energieminimum des mechanischen Systems erreicht wird. Manchmal kann ein System mehr als ein Minimum haben, die auch noch unterschiedlich hoch ausfallen. Dann ist das Minimum mit der höheren Energie METASTABIL, d.h. es ist nicht mehr stabil, wenn allzu große Schwankungen der Zustandsvariablen auftreten.

Beispiel 1: Ein System A sei ein Wagen der Masse m , dessen innere Energie nur aus der kinetischen Energie $mv^2/2$ seiner Relativbewegung zu einem umgebenden System B besteht. Da die kinetische Energie positiv ist, liegt das Energieminimum vor, wenn $v = 0$, d.h. wenn der Wagen relativ zu B zur Ruhe gekommen ist. Es ist eine feine Ironie, dass die Aristotelische Physik die Alltagsbeobachtungen völlig zutreffend beschreibt: Wenn sich ein Körper gleichförmig und geradlinig bewegen soll, so ist unsere Alltagserfahrung, dass dem Körper dazu ständig Energie zugeführt werden

muss bzw. eine äußere Kraft auf den Körper wirken muss. Ohne diese Kraft hört die Bewegung alsbald schon auf.

Beispiel 2: Die potentielle Energie eines Systems der Masse m liege mit dem Schwereparameter g und einer Federkonstanten k in der Form

$$E_{pot}(x) = mgx + \frac{1}{2}kx^2$$

vor. Indem x immer kleiner und schließlich negativ wird, kann die Schwere-Energie mgx immer weiter abnehmen, aber je negativer x wird, desto stärker wächst die elastische Energie der Feder an. Der Wert x_0 der Zustandsvariablen x des Systems, bei dem der Gleichgewichtszustand erreicht wird, ergibt sich durch Nullsetzen der ersten Ableitung:

$$\frac{dE_{pot}}{dx} = mg + kx = 0.$$

Daraus folgt für den Gleichgewichtszustand: $x_0 = -mg/k$. Das ist die Ruhelage, bei der keine weiteren Änderungen der Zustandsvariablen x mehr stattfindet. Die potentielle Energie im Gleichgewichtszustand beträgt $E_{pot}(x_0) = -(mg)^2/2k$ und kann durch irreversible Prozesse nicht mehr weiter vermindert werden.

18. Kontinuumsmechanik der Materie

Die Kontinuumsmechanik wird hier in nichtrelativistischer Näherung (alle auftretenden Geschwindigkeiten klein gegen c , Wechselwirkungsenergien klein gegen mc^2) und im Schwerpunktsystem (Kap. 46.1) beschrieben. Im Rahmen der Kontinuumsmechanik werden einige Grundbegriffe herausgearbeitet, welche z.B. für die Thermodynamik wichtig sind.

18.1. Materiemodelle

MAKROSKOPISCHE SYSTEME bestehen aus einer Vielzahl von Objekten bzw. Teilsystemen, die meist ihrerseits weiter teilbar sind. Ein Beispiel für ein solches physikalisches Makrosystem ist unser Sonnensystem. Es besteht aus der Sonne, der Erde, Kleinplaneten wie Ceres, Kometen, interplanetarischem Staub usw. Das sind schon ziemlich viele Objekte. Aber in jedes kann man wieder hineinzoomen und ein solches Teilsystem wiederum als das interessierende Makrosystem ansehen. Es besteht wieder aus Teilsystemen beispielsweise aus der Atmosphäre, Bergen, Meeren, Gasen, Steinen, Bären usw. Das sind auch wieder ziemlich viele Objekte. Da kann man wieder hineinzoomen. Auf irgendeiner Ebene des Zoomens wollen wir davon ausgehen, dass wir dabei auf ein Makrosystem stoßen, dessen Teile scheinbar nicht mehr diskretisiert werden können. Man modelliert es dann als KONTINUUM, d.h. als ein beliebig oft – kontinuierlich – teilbares physikalisches System. Unser Ziel wird erst einmal sein, ein einfaches Modell für ein homogenes und kontinuierliches Makrosystems zu etablieren.

Aus dem Sonnensystem kann man aber auch herauszoomen. Dann wird das Sonnensystem – so gewaltig groß es auch erscheinen mag – auf der galaktischen Ebene näherungsweise zu einem strukturlosen Punkt, erst recht wenn man auf die Ebene von Galaxienhaufen und schließlich auf den Kosmos herauszoomt. Wenn man in extremer Idealisierung unendlich weit herauszoomt, dann genügt ein Punkt als adäquates Modell des Systems. Genauer gesagt, beschreibt man das Sonnensystem und seine Bewegung im Kosmos dann durch das Modell eines MASSENPUNKTS. Die einzige Eigenschaft, die das abgeschlossene System auf dieser Ebene charakterisiert, ist seine Masse m bzw. seine innere Energie $\mathcal{U} = mc^2$.

Teilchen, deren innere Struktur nicht auflösbar ist oder nicht interessiert, bezeichnet man als PUNKTTEILCHEN. Sie werden durch ihre invarianten Eigenschaften spezifiziert. Dazu gehören z.B. Masse, elektrische Ladung (Kap. 35.3) oder Eigendrehimpuls (Kap. 46.1). Der Massenpunkt, bei dem allein die Masse zur Charakterisierung genügt, ist ein Spezialfall eines Punktteilchens. Wenn für ein Punktteilchen alleine die elektrische Ladung zur Charakterisierung genügt, spricht man von einer Punktladung.

Reale Materie wird durch Materiemodelle beschrieben.

Das Punktteilchenmodell und das Kontinuumsmodell sind die beiden Extrempole der Materiemodelle der Physik.

Dieser Teil, der Teil 18, geht vom Kontinuumsmodell der Materie aus. Materie wurde bis Ende des 19. Jahrhundert tatsächlich als etwas Kontinuierliches angesehen, das beliebig unterteilt werden konnte. Die Vorstellung, dass Materie aus diskreten Teilchen, etwa Molekülen oder Atomen, aufgebaut sein könnte, wurde damals von der großen Mehrheit der Physiker abgelehnt. Heute kommt dem KONTINUUMSMODELL DER MATERIE selbstverständlich nur der Status einer Näherung zu. Aber sie entspricht der Alltagserfahrung des Menschen und ist in vielen Anwendungsbereichen der Ingenieurwissenschaften recht brauchbar. Die Kontinuumsphysik ist PHÄNOMENOLOGISCHE PHYSIK. Sie verbleibt im Deskriptiven und versucht nicht zu ergründen, warum die Zusammenhänge so sind, wie sie sind. Die Kontinuumsphysik und auch die Thermodynamik (Teil IV) werden im Folgenden vornehmlich in nichtrelativistischer Näherung (alle auftretenden Geschwindigkeiten klein gegen c , Wechselwirkungsenergien klein gegen mc^2) und im Schwerpunktsystem (Kap. 46.1) präsentiert.

Eine erste grobe Einteilung der Materie kann man analog den vier antiken Elementen Erde, Wasser, Luft und Feuer durch die Attribute fest, flüssig, gasförmig oder plasmaartig erreichen. Die ersten drei sind Ihnen aus dem Alltagsleben geläufig. Das sind die drei KLASSISCHEN AGGREGATZUSTÄNDE. FESTKÖRPER und FLÜSSIGKEITEN fasst man zur KONDENSIERTEN MATERIE zusammen, wenn man ihre (näherungsweise) Beibehaltung des Volumens bzw. die im Vergleich zu Gasen erheblich geringere Kompressibilität betonen will. Flüssigkeiten und Gase fasst man zu den FLUIDEN zusammen, wenn man eher die Fließfähigkeit bzw. Unbeständigkeit der Form herausstreichen möchte.

$$\text{kondensierte Materie} \left\{ \begin{array}{l} \text{Festkörper} \\ \text{Flüssigkeiten} \\ \text{Gase} \end{array} \right\} \text{Fluide}$$

Um diese Einführung in die Physik einfach zu halten, wird die Materie als ISOTROP. Das Modell der ISOTROPEN MATERIE ist eine gute Beschreibung für alle Gase, die meisten Flüssigkeiten und sehr viele Festkörper.¹⁶

Meist werden wir die Materie der Einfachheit halber auch als homogen voraussetzen. Wenn man herausstreichen möchte, dass die betrachtete Materie (bzw. der betrachtete Stoff) homogen ist, spricht man auch von einer PHASE. Das Wort „Phase“ hat in Zusammenhang mit Materie also eine andere Bedeutung als in Kap. 14.3. Wenn man von einer flüssigen Phase spricht, so meint man einen homogenen Bereich eines flüssigen Stoffes.

18.2. Stoffmenge und Dichte

Stoffmenge. Materie tritt uns in Form chemisch verschiedener Stoffe entgegen. Unterschiedlichen Stoffe haben unterschiedliche stoffspezifische Eigenschaften. Um diese zu beschreiben, benötigt man erst einmal den Begriff der STOFFMENGE. Sie ist selbstverständlich eine extensive Größe. Die Stoffmenge eines homogenen Stoffes charakterisieren

¹⁶ANISOTROPE MATERIE (z.B. KRISTALLE oder FLÜSSIGKRISTALLE) wird erst im dritten Band dargestellt, denn man benötigt dafür die Tensorrechnung.

wir vorläufig durch seine Masse.¹⁷ Der Begriff der Masse tritt somit in drei unterschiedlichen Bedeutungen auf, nämlich als Maß der inneren Energie, der „gravitativen Ladung“ und der Stoffmenge. Das charakteristische Instrument zur Stoffmengenbestimmung bzw. zum Stoffmengenvergleich ist die WAAGE (Beispiel: Küchenwaage).

Dichte. Die gleiche Masse m eines Stoffes kann sehr unterschiedliche Volumina V einnehmen. Das Verhältnis

$$\rho = \frac{m}{V} \quad (18.1)$$

ist das Maß für die „Dichtigkeit“ homogener Materie und wird als SPEZIFISCHE DICHTe bezeichnet. Das Wort „spezifisch“ weist darauf hin, dass die Dichte eine stoffspezifische Eigenschaft ist. Falls ein Körper inhomogen ist, stellt das Verhältnis von Masse und Volumen seine mittlere Dichte dar.

18.3. Auftrieb

In Abschnitt 12.4.2 durchlief ein Wagen W der Masse m_W die Höhendifferenz $\Delta h = h_i - h_f$. Der Wagen repräsentiert eine bestimmte Stoffmenge, die ein Volumen $V = m/\rho_W$ einnimmt, wobei ρ_W die (mittlere) Dichte des Wagens ist. Die Änderung der potentiellen Energie wird dabei im dissipationsfreien Grenzfall in kinetische Energie umgewandelt. Nun ist es aber nicht so, dass der Wagen auf der Höhe h_i ein Loch der Größe seines eigenen Volumens V im ansonsten luftgefüllten Raum hinterlässt und das entsprechende Luftvolumen V weggehext wird, das er am Ende auf der Höhe h_f einnimmt, sondern im Gegenzug wird eine Stoffmenge Luft, welche das gleiche Volumen V wie der Wagen einnimmt und somit die Masse $m_L = \rho_L V$ hat, auf die Höhe h_i angehoben. Die potentielle Energie ändert sich daher um

$$\Delta E_{pot} = -m_W g \Delta h + m_L g \Delta h = (\rho_L - \rho_W) V g \Delta h.$$

Die Abnahme der potentiellen Energie ist also um $\rho_L g \Delta h$ geringer als sie es im luftleeren Raum wäre. Dieses Phänomen wird als AUFTRIEB bezeichnet. Die Minderung der potentiellen Schwereenergie eines Körpers mit dem Volumen V , der in ein Fluid eintaucht, um die diesem Volumen entsprechende Masse des Fluids, wird als ARCHIMEDISCHES PRINZIP bezeichnet. Es ist sogar so, dass sich die kinetische Energie eines Körpers und somit seine Geschwindigkeit sogar erhöht, wenn er in einem Fluid aufsteigt, nämlich dann, wenn seine mittlere Dichte kleiner ist als die des Fluids.

18.4. Elastizität

Ein Körper besitzt eine Gestalt, die für ihn charakteristisch und STATIONÄR ist, wenn man ihn in Ruhe lässt. Wenn man die Quietsch-Ente im Badezimmer (Abb. 44a) drückt, ändert sie ihre Gestalt, doch wenn man sie anschließend in Ruhe lässt, nimmt sie nach

¹⁷Die endgültige Definition der Stoffmenge erfolgt erst in Kap. 34.3.

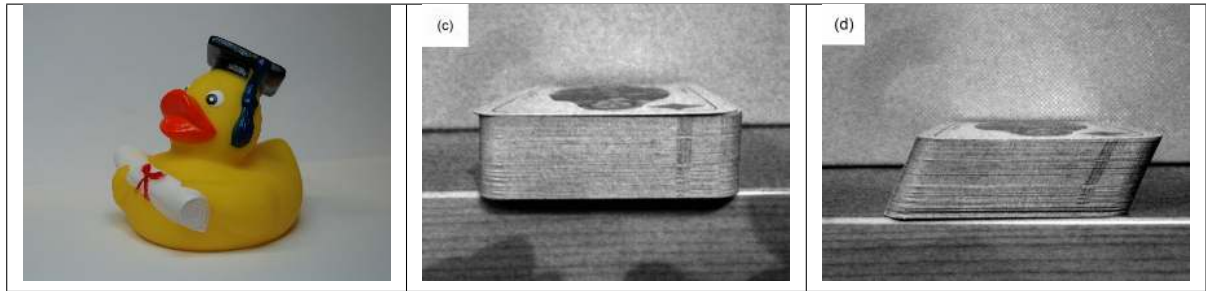


Abbildung 44: (a) Reversibel und somit elastisch deformierbare (akademische) Quietsch-Ente. (b) Ungeschertes Kartendeck mit Scherwinkel $\alpha = 0$. (c) Um einen Winkel $\alpha \approx 25^\circ$ geschertes Kartendeck. Die hier gezeigte Scherdeformation ist irreversibel und somit plastisch. Physik ist Trumpf: Könige, Damen und Buben dieser Tarok-Karten zeigen die Professoren der Fakultät für Physik der Universität Ljubljana.

einer Weile wieder die Gestalt an, die sie ursprünglich hatte. Das ist die Gestalt, die sie im stationären Zustand bzw. im Gleichgewicht hat (GLEICHGEWICHTSZUSTAND) und woran jeder sie sofort als Quietsch-Ente erkennt.

ELASTIZITÄT beschreibt den Grenzfall vollständig REVERSIBLER Deformationen. Irreversible Deformationen bezeichnet man als PLASTISCHE DEFORMATIONEN. Beispielsweise handelt es sich bei der Deformation, welche einen Stapel Spielkarten von der in Abb. 44c gezeigten Ausgangslage in die in Abb. 44d gezeigte Endlage überführt, um eine plastische Deformation. In diesem Kapitel werden ausschließlich elastische Deformationen untersucht.

Bei einer Feder muss man, wie bereits diskutiert, Energie zuführen, wenn man die Gleichgewichts-Gestalt ändern möchte (Kap. 12.4.1). Die Federkonstante k im Hookeschen Gesetz (Kap. 12.4.1) ist zum Beispiel eine phänomenologische Beschreibung des Elastizität von Federn. Durch die Energiezufuhr ändert sich die innere Energie der Feder. Bei der Feder nimmt man gewöhnlich ihre Auslenkung x aus der Ruhelage als Zustandsvariable. Die reversibel gespeicherte bzw. elastische Energie einer Feder ist das Paradebeispiel für eine potentielle Energieform, weil es sich durch ein so einfaches Modell beschreiben lässt.

Wenn man die energetischen Prozesse für die Gestaltänderung einer realen Quietsch-Ente aufstellen will, dann kann das eine ziemlich komplizierte Sache werden. Das liegt zum einen daran, dass man sie auf viele unterschiedliche Arten deformieren kann und deshalb von vorn herein nicht klar ist, was eine geeignete Zustandsvariable wäre. Zweitens ist ihre Deformationen zwar überwiegend elastisch aber doch auch ein klein wenig plastisch, denn man wird die Energie, die man für die Entendeformation aufgewandt hat, nicht mehr vollständig zurückbekommen. Zum dritten besteht eine Komplikation darin, dass die Ente nicht unbedingt aus einem homogenen Material besteht. Es soll hier daher gar nicht erst versucht werden, so etwas Schwieriges wie eine reale Quietsch-Ente zu modellieren, denn dieses Buch ist eine Einführung in die Physik. Stattdessen sollen

hier nur einige wichtige Standarddeformationen nebst ihren beschreibenden Zustandsvariablen vorgestellt werden: Dehnung (bzw. Stauchung), Scherung und Torsion, sowie die reine Volumendeformation. Jede davon könnte als vollkommen elastischer (reversibler) oder vollkommen plastischer (nicht-reversibler) Grenzfall realisiert sein oder irgendwo dazwischen liegen. In diesem Kapitel und im nächsten Kapitel werden nur Modelle für vollkommen elastischen Deformationen von Medien besprochen, die im Gleichgewichtszustand homogen und isotrop sind.

18.4.1. Dehnung

Am nächsten kommt die DEHNUNG eines homogener Stabes entlang der Stabachse ran an die Federdeformation. Zur Beschreibung kann man die x -Achse eines Koordinatensystems parallel zu Dehnungsrichtung legen. Mit x sei die Länge des Stabes im Gleichgewicht bezeichnet. Bei der Dehnung muss Energie zugeführt werden, die dann als potentielle Energie im Stab reversibel gespeichert ist. Analog zur Feder beschreibt man sie, indem man die Längenänderung Δx oder besser noch die relative Längenänderung $\varepsilon = \Delta x/x$ als Zustandsvariable wählt. Wenn man annimmt, dass sich die potentielle Energie in einem homogenen Stab auf das ganze Gleichgewichtsvolumen V_0 gleichmäßig verteilt und diesen Faktor herauszieht, dann wählt man

$$E_{pot}(\varepsilon) = \frac{1}{2}V_0E\varepsilon^2$$

als Hookesches Energiemodell. Dies definiert die materialspezifische Konstante E , die man als ELASTIZITÄTSMODUL bezeichnet. Der Elastizitätsmodul von Eisen liegt beispielsweise in der Größenordnung von $E \approx 10^{11} \text{ J/m}^3$.

Eine Dehnung geht immer mit einer Kontraktion der Querdimension einher. Für isotrope Materialien ist jede Richtung orthogonal zur Dehnungsachse gleichwertig, und man kann entlang irgendeiner davon die y -Achse legen. Die relative Längenänderung und die relative Querkontraktion $\Delta y/y$ erweisen sich als proportional zueinander:

$$\frac{\Delta y}{y} = -\mu \frac{\Delta x}{x} \quad (18.2)$$

Der materialspezifische positive Faktor μ heißt POISSONZAHL.

18.4.2. Reine Formänderungen

Für allgemeine Deformationen gibt es zwei wichtige Grenzfälle, nämlich den einer REINEN VOLUMENÄNDERUNG und den einer REINEN FORMÄNDERUNG. Reine Formänderungen sind Gestaltsänderungen, bei denen das Volumen konstant bleibt. Je nachdem, wie man eine Deformation an einer Quietsch-Ente experimentell realisiert, wird man näher an dem einen oder anderen Grenzfall liegen und das entsprechende Energiemodell zur Beschreibung heranziehen: Wenn ein Taucher die Quietsch-Ente zum Meeresboden mitnimmt, ist die reine Volumenänderung das adäquatere Modell. Wenn man ihr den Hals verdreht, handelt es sich in guter Näherung um eine Torsion und damit eher um eine reine Formänderung.

Scherung. Als Scher-Deformation bzw. SCHERUNG bezeichnet man eine Deformation welche die in Abb. 44c gezeigte Ausgangslage (Scherwinkel $\alpha = 0$) in die in Abb. 44d gezeigte Endlage (Scherwinkel $\alpha \approx 25^\circ$) überführt. Es ist offensichtlich, dass es sich um den Typus einer reinen Formänderung ohne Volumenänderung handelt. Erfolgt die Scherung elastisch, kann man den Scherwinkel α als Zustandsvariable wählen und die elastische Energie durch die Standardpotenzreihe um $\alpha = 0$ darstellen. Da es nicht auf die Scherrichtung ankommt, können in der Reihenentwicklung nur gerade Potenzen von α auftreten. Meist genügt das erste Glied

$$E_{pot}(\alpha) = \frac{1}{2}V_0G\alpha^2 \quad (18.3)$$

dieser Reihenentwicklung, um die experimentellen Beobachtungen ausreichend genau zu beschreiben. Der materialspezifische Faktor G heißt SCHERMODUL.

Torsion. Ein Draht habe eine Länge h und einen Radius r . Wenn Ober- und Unterseite um den Winkel φ gegeneinander verdreht sind, bezeichnet man diese Art der Deformation als TORSION. Auch die Torsion stellt eine reine Formänderung dar. Anders als bei der Scherung bleibt aber nicht nur das Volumen unverändert, sondern auch die Oberfläche $\mathcal{A}_G = 2\pi(r^2 + rh)$ des zylindrischen Körpers. Da es nicht auf die Drehrichtung ankommt, treten in der Entwicklung der inneren Energie in der Standardpotenzreihe nur gerade Potenzen von φ auf. In erster Näherung ist dann die potentielle Energie durch

$$E_{pot}(\varphi) = \frac{1}{2}V_0k_T\varphi^2 \quad (18.4)$$

mit einer Torsionskonstanten k_T darstellbar.

Ideale Fluide. Wenn der Schermodul G und die Torsionskonstante k_T eines homogenen Mediums verschwinden, ist jede Formänderung ohne Energiekosten (abgesehen von der Grenzflächenenergie, Abschnitt. 18.7.1) möglich, d.h. das Medium ist bei konstantem Volumen für jegliche Form im Gleichgewicht. Solche Medien werden als IDEALE FLUIDE bezeichnet.

18.5. Elastische Volumendeformationen

Auch in diesem Kapitel geht es um Elastizität. Wegen der Bedeutung der elastischen Volumendeformationen ist ihnen ein gesondertes Kapitel gewidmet.

18.5.1. Druck

Da ideale Fluide ihre Form energieneutral ändern können, sind hier Volumendeformationen die einzig energetisch relevanten elastischen Gestaltsdeformationen. Für Festkörper liegen hingegen nur dann reine Volumendeformationen vor, wenn die Deformationen

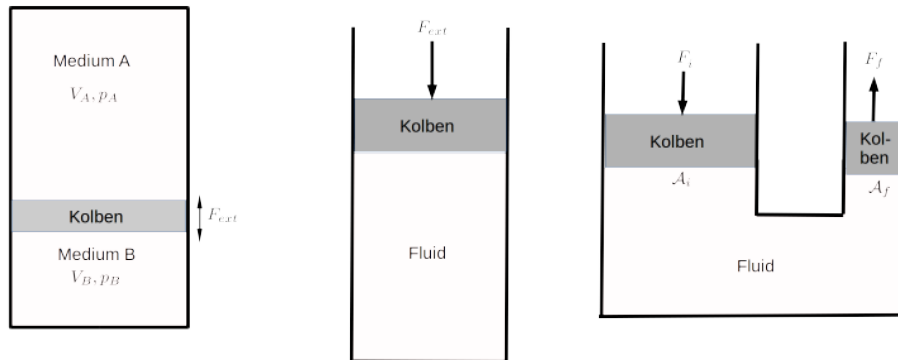


Abbildung 45: (a) Druckausgleich. (b) Elastische Energie und Arbeit (c) Weg-Kraft-Transformation

winkel- und verhältnistreu sind. Daher wird der Einfachheit halber im Folgenden ein ideales Fluid betrachtet und zwar als ein mechanisches System, dessen potentielle Energie allein vom Volumen V abhängt, d.h. $E_{pot} = E_{pot}(V)$. Das Differential ist

$$dE_{pot} = \frac{dE_{pot}}{dV} dV = -\mathbf{p} dV \quad (18.5)$$

Die hierdurch definierte neue Größe

$$\mathbf{p} = \mathbf{p}(V) = -\frac{dE_{pot}(V)}{dV} \quad (18.6)$$

wird als DRUCK bezeichnet. Der Druck ist selbst wieder eine Funktion der Zustandsvariablen, also hier allein des Volumens V . Da Energie und Volumen Skalare sind, ist auch der Druck eine skalare Größe. Die SI-Einheit des Drucks ist PASCAL:

$$[\mathbf{p}] = 1 \text{ Pascal} = 1 \text{ Pa} = 1 \text{ J/m}^3. \quad (18.7)$$

Der Luftdruck liegt beispielsweise ungefähr bei $1 \text{ bar} = 10^5 \text{ Pa} = 1000 \text{ hPa}$ (1000 Hektopascal). Ein Vergleich mit Gl. 15.3 zeigt, dass man den Druck als eine verallgemeinerte Kraft auffassen kann.

18.5.2. Druckausgleich

Zwei Fluide A und B mögen ein abgeschlossenes mechanisches Gesamtsystem G mit konstantem Volumen

$$V = V_A + V_B$$

bilden. Die Volumina V_A und V_B seien durch einen Kolben voneinander getrennt (Abb. 45a). Die Energie E_{pot} des abgeschlossenen Gesamtsystems ist die Summe der Energien von A und B, d.h.

$$E_{pot} = E_{pot,A} + E_{pot,B}.$$

Wenn man den Kolben verschiebt, so wird eines der beiden Volumina vergrößert und das andere um den gleichen Betrag verkleinert. Wegen $dV_A = -dV_B$ folgt für die Änderung der mechanischen Energie des Gesamtsystems

$$dE_{pot} = dE_{pot,A} + dE_{pot,B} = -p_A dV_A - p_B dV_B = (p_B - p_A) dV_A.$$

Ist $p_A > p_B$ und vergrößert sich das Volumen V_A , dann gibt das System G entweder über den Kolben Energie nach außen ab, leistet also Arbeit, oder es kann die frei werdende Energie dissipieren. Ist das System G abgeschlossen, so ist nur eine irreversible Dissipation der mechanischen Energie möglich. Dann darf der Vorgang der Verschiebung des Kolbens nur so lange ablaufen, bis die mechanische Energie ein Minimum erreicht hat. Das ist das mechanische Gleichgewicht (s. Kap. 17). Im Minimum ist $dE_{pot} = 0$ für $dV_A \neq 0$ und das bedeutet, dass der Druck im mechanischen Gleichgewicht ausgeglichen ist (DRUCKAUSGLEICH):

$$p_A = p_B. \quad (18.8)$$

Überträgt man diesen Gedankengang auf ein Medium als Ganzes, dann ist die Konsequenz, dass der Druck in einem homogenen Medium überall gleich ist, wenn das Medium sich im mechanischen Gleichgewicht befindet.

18.5.3. Weg-Kraft-Transformation

Ein Fluid sei in einem Zylinder eingeschlossen und im mechanischen Gleichgewicht. Abgesehen von einem Kolben mit einer Querschnittsfläche \mathcal{A} seien alle anderen Wände starr. Wenn man den Kolben um eine infinitesimale Strecke dx in den Zylinder hineindrückt, erhöht sich die elastische Energie des Fluids um $dE_{pot} = -p dV = p \mathcal{A} dx$. Dazu muss von außen Energie zugeflossen sein. Im reversiblen (dissipationsfreien) Grenzfall ist die am System geleistete Arbeit

$$dW = F_{ext} dx = dE_{pot} \quad (18.9)$$

gleich der Erhöhung der elastischen Energie. Die Zufuhr der Energie von außen kann durch eine Kraft F_{ext} orthogonal zur Kolbenfläche beschrieben werden, welche über die Strecke dx entlang der x -Richtung wirksam gewesen ist. Da externe Kraft, Flächennormale des Kolbens und Bewegungsrichtung parallel zueinander sind, kann das Problem als eindimensional betrachtet werden. Die vom Kolben auf das Fluid ausgeübte Kraft

$$F_{ext} = p \mathcal{A}, \quad (18.10)$$

ist der Fläche proportional und wirkt orthogonal zu ihr bzw. parallel zur Flächennormale (Abb. 45b).

Man kann nun eine zweiten mit einem Kolben versehene Öffnung an den Zylinder anschließen und das Fluid als Arbeitsmedium betrachten (Abb. 45c). Wenn man den Prozess so führt, dass über diesen Kolben gerade so viel Arbeit $dW_f = -F_f dx_f$ vom Fluid nach außen abgegeben wird, wie es an Arbeit $dW_i = F_i dx_i$ über den ersten Kolben aufgenommen hat, dann hat sich der Zustand des Fluids nicht verändert. Das Fluid hat einen KREISPROZESS ausgeführt: die zuerst zugeführte Arbeit dW_i hat seine elastische Energie erhöht und im zweiten Schritt wurde der Ausgangszustand durch Abgabe der Arbeit dW_f wieder hergestellt. Sowohl die potentielle Energie als auch das Volumen des Arbeitsmediums sind nach Abschluss des Zyklus unverändert geblieben. Mit der Volumenänderung $dV = \mathcal{A}_f dx_f + \mathcal{A}_i dx_i$ folgt

$$F_f/F_i = -dx_i/dx_f = \mathcal{A}_f/\mathcal{A}_i. \quad (18.11)$$

Die am Fluid angreifenden äußeren Kräfte transformieren sich im gleichen Verhältnis wie die Kolbenflächen. Ist die eingangsseitige Querschnittsfläche des Kolbens kleiner als die ausgangsseitige, dann ist die ausgangsseitige Kraft größer als die eingangsseitige. In ähnlicher Weise wie dieser HYDRAULISCHE TRANSFORMATOR funktionieren auch andere primitive mechanische Maschinen wie Hebel, schiefe Ebene oder Flaschenzüge als Kraft-Weg-Transformatoren. Die gerade angestellte Betrachtung ist selbstverständlich stark idealisiert, d.h. Dissipation oder ein anderer Verbleib der extern zugeführten Energie im Fluid blieben unberücksichtigt.

Um die Dissipation klein zu halten, muss der Prozess u.a. quasistatisch durchgeführt werden, d.h. man muss den Transformationsprozess sehr langsam ablaufen lässt, so dass die relativ zur Umgebung auftretenden Geschwindigkeiten klein bleiben.

18.5.4. Kompressibilität

Ein Medium sei bei einem Druck \mathbf{p} im mechanischen Gleichgewicht mit seiner Umgebung. Eine infinitesimale Änderung des Umgebungsdrucks um $d\mathbf{p}$ ruft eine relative Volumenänderung des Mediums um dV/V hervor. Im infinitesimalen Grenzfall ist ergibt sich meist ein linearer Zusammenhang:

$$\frac{dV}{V} = -\kappa d\mathbf{p}.$$

Die materialspezifische positive Größe

$$\kappa = -\frac{1}{V} \frac{dV}{d\mathbf{p}}. \quad (18.12)$$

heißt KOMPRESSIBILITÄT. Wenn zwei Materialien der gleichen Änderung des Umgebungsdrucks ausgesetzt sind, so ist dasjenige kompressibler, bei dem die größere relative Volumenänderung eintritt. Typische Größenordnungen für die Kompressibilität sind in Tab. 6 angegeben.

Tabelle 6: Typische Größenordnungen für die Kompressibilität

	Kompressibilität
Gase (unter Normalbedingungen)	10^{-5} Pa^{-1}
Flüssigkeiten	10^{-10} Pa^{-1}
Festkörper	10^{-11} Pa^{-1}

Gase ändern ihr Volumen um ca. 10 ppm pro Pascal. Die Bezeichnung „ppm“ bedeutet „parts per million“, also „ein Millionstel“ und es ist $1 \text{ ppm} = 10^{-3} \text{ ‰} = 10^{-4} \text{ ‰}$, wobei die Zeichen ‰ und ‰ für Promille bzw. Prozent stehen. Die Kompressibilität kondensierter Materie ist wesentlich kleiner als die der Gase. Bei gleicher Druckänderung ist ihre relative Volumenänderung um 5-6 Größenordnungen kleiner. Für kondensierte Materie kann man daher oft vom stark vereinfachten Modell einer INKOMPRESSIBLEN FLÜSSIGKEIT oder eines INKOMPRESSIBLEN FESTKÖRPERS ausgehen, bei dem die bei Energiezufuhr zwar stets vorhandene, aber extrem kleine Volumenänderung unberücksichtigt bleibt. Die Bezeichnung „inkompressibel“ ist leider etwas missverständlich. Inkompressible Materie bezeichnet keineswegs ein Modell, bei dem der elastische Beitrag zur inneren Energie eines Körpers bzw. Fluids vernachlässigt wird, sondern nur die damit verbundene Änderung des Volumens.

Die Werte der Kompressibilität hängen vor allem für Gase davon ab, unter welchen Bedingungen die Messungen durchgeführt werden (s. Kap. 30.1 und 31). Der Kehrwert

$$K = \frac{1}{\kappa} = -V \frac{dp}{dV} \quad (18.13)$$

heißt KOMPRESSIOMODUL.

Wenn nur mechanische Energien von Belang sind, gilt wegen Gl. 18.6

$$K = V_0 \frac{d^2 E_{pot}}{dV^2}, \quad (18.14)$$

wobei V_0 das Gleichgewichtsvolumen ist. Die potentielle Energie $E_{pot}(V) = E_{pot}(V_0 + \Delta V)$ hängt nur von der Abweichung ΔV vom Gleichgewichtsvolumen ab. Durch zweimalige Integration erhält man aus Gl. 18.14 dann

$$E_{pot}(\Delta V) = \frac{1}{2} \frac{K}{V_0} \Delta V^2 = \frac{1}{2} V_0 K \left(\frac{\Delta V}{V_0} \right)^2. \quad (18.15)$$

Dieses Gesetz hat die gleiche Form wie die Elastizitätsgesetze, die in Kap. 18.4 vorgestellt wurden.

18.6. Enthalpie

Die Enthalpie ist ein wichtiger energetischer Begriff der Mechanik. Sie ist insbesondere für die technische Strömungsmechanik von Bedeutung (s. „*Romanos Physikkurs 2*“).

Vergleicht man einen Massenpunkt ($V = 0$) der Masse m mit einem ausgedehnten Körper, der bei gleicher Masse m ein Volumen V einnimmt, so enthalten beide genau die gleiche innere Energie $\mathcal{U} = mc^2$. Wenn sie im mechanischen Gleichgewicht mit der Umgebung gebildet werden sollen, muss jedoch eine unterschiedliche Verdrängungsenergie aufgebracht werden. Gleichgewicht bedeutet hier, dass beide Objekte den gleichen Druck wie ihre Umgebung haben, und daher ist die Verdrängungsenergie gerade gleich dem Produkt aus diesem Gleichgewichtsdruck \mathbf{p} und dem Gleichgewichtsvolumen V . Für einen Massenpunkt ist die Verdrängungsenergie somit null. Wenn man den Massenpunkt aber zu einem ausgedehnten Körper „aufbläst“, ist sie $\mathbf{p}V$. Um irgendein Medium bzw. einen Körper der Masse m im Gleichgewicht mit der Umgebung zu erzeugen, muss man also die ihm zustehende innere Energie zuzüglich seiner Verdrängungsenergie $\mathbf{p}V$ bereitstellen. Um dem Rechnung zu tragen, führt man die ENTHALPIE

$$\mathcal{H} = \mathcal{U} + \mathbf{p}V \quad (18.16)$$

als neue energetische Größe ein. Die Enthalpie ist gleichbedeutend mit einer BILDUNGS-ENERGIE für einen Körper, einen Stoff oder ein Medium im Druckgleichgewicht mit der Umgebung.

Wenn sich der Umgebungsdruck infinitesimal von \mathbf{p} auf $\mathbf{p} + d\mathbf{p}$ ändert und sich dabei weder Volumen noch Masse des Körpers ändern, so muss man von außen eine infinitesimale Energie $dW = Vd\mathbf{p}$ zuführen. Das kann man folgendermaßen einsehen: Man lässt den Körper beim Druck \mathbf{p} auf einen Massenpunkt schrumpfen, ändert den Druck auf $\mathbf{p} + d\mathbf{p}$ und bläst ihn anschließend wieder auf das Volumen V auf. Die Bilanz für die Verdrängungsenergie ist dann $(\mathbf{p} + d\mathbf{p})V - \mathbf{p}V = Vd\mathbf{p}$. Wenn sich das Volumen bei konstantem Druck von V auf $V + dV$ ändert, ergibt die analoge Überlegung, dass eine infinitesimale Energie $dW = \mathbf{p}dV$ von außen zugeführt werden muss.

Wird einem Medium eine Energie dW von außen zugeführt, so kann diese vollständig in eine Änderung der inneren Energie umgesetzt werden. Oft wird sie aber nur teilweise in innere Energie umgesetzt, weil die Änderung der inneren Energie oft auch mit einer Volumenänderung des Mediums verbunden ist oder weil dabei auch eine Druckänderung im Gleichgewicht mit der Umgebung stattfindet. Dann muss ein Teil der zugeführten Energie auch für diese Änderungen aufkommen. Daher ist die zugeführte Energie für ein Medium im Gleichgewicht mit seiner Umgebung gerade gleich der Änderung seiner Enthalpie:

$$d\mathcal{H} = d\mathcal{U} + d(\mathbf{p}V) = d\mathcal{U} + \mathbf{p}dV + Vd\mathbf{p} = dW . \quad (18.17)$$

Wenn die Vorgänge dissipationsfrei ablaufen und nur ein potentieller Energiebeitrag E_{pot} sich ändert, dann gilt entsprechend, dass die zugeführte Arbeit dW gleich der Enthalpieänderung ist:

$$d\mathcal{H} = dE_{pot} + \mathbf{p}dV + Vd\mathbf{p} = dW . \quad (18.18)$$

Auch hier ist zu beachten, dass die beiden Beiträge $\mathbf{p}dV + Vd\mathbf{p}$ eine Konsequenz der Verdrängungsenergie sind. Sie sind ein Effekt des Gleichgewichts mit der Umgebung und haben nichts mit der inneren Energie des Mediums selbst zu tun. Das ist analog zum Effekt des Auftriebs (Kap. 18.2). Wenn die potentielle Energie nur aus der elastischen

Energie des Mediums besteht und daher $dE_{pot} = -pdV$ ist, so ändert infinitesimalen Energiezufuhr dW sich die Enthalpie durch eine infinitesimale Zufuhr δW von Arbeit um

$$\delta W = d\mathcal{H} = -pdV + \mathbf{p}dV + Vd\mathbf{p} = Vd\mathbf{p},$$

denn einerseits wird die innere Energie des Mediums um $-pdV$ vergrößert, andererseits muss eine Arbeit von $+\mathbf{p}dV$ zur Verdrängung der Umgebung geleistet werden, so dass sich die zugeführte Arbeit in einer Druckerhöhung um $d\mathbf{p}$ bei näherungsweise konstantem Volumen V auswirkt.

18.7. Grenzflächenphänomene

18.7.1. Grenzflächenenergie

Das Besondere an der Torsion eines zylindrischen Körpers ist, dass nicht nur das Volumen unverändert bleibt, sondern auch seine Oberfläche bzw. genauer die Fläche, die ihn von seiner Umgebung abgrenzt: seine Grenzfläche \mathcal{A}_G . Fluide erweisen sich dann als Materialien, für welche sich die innere Energie bei Torsion nicht ändert. Das Modell des IDEALEN FLUIDS ist durch $k_T = 0$ definiert. So lange man weder Volumen noch die Fläche \mathcal{A}_G verändert, die es von seiner Umgebung abgrenzt, kann man im Idealfall Flüssigkeits- oder Gasschichten relativ zueinander verdrehen, ohne dass man dazu Energie aufwenden muss.

Wenn man bei einer Deformation jedoch die Fläche \mathcal{A}_G ändert, so tritt ein dazu proportionaler Beitrag zur inneren Energie auf. Wenn a und b die aneinander angrenzenden Medien bezeichnen und man \mathcal{A}_G als Zustandsvariable heranzieht, dann ist die potentielle Energie durch

$$E_{pot}(\mathcal{A}_G) = \varepsilon_{ab}\mathcal{A}_G \quad (18.19)$$

gegeben. Die Proportionalitätskonstante ε_{ab} wird als SPEZIFISCHE GRENZFLÄCHENERGIE bezeichnet und hat die physikalische Dimension $[\varepsilon_{ab}] = [J/m^2]$. Sie ist nicht für ein Material spezifisch wie das z.B. für den Elastizitätsmodul E oder den Torsionsmodul G der Fall ist, sondern hängt von beiden aneinander angrenzenden Materialien a und b ab. Ist ε_{ab} positiv, kostet die Bildung von Grenzfläche Energie und die beiden aneinander angrenzenden Stoffe versuchen, ihre gemeinsame Grenzfläche zu minimieren. Ist ε_{ab} negativ, versuchen die Stoffe möglichst viel gemeinsame Oberfläche zu bilden und aneinander zu haften, d.h. es tritt ADHÄSION auf.

fluid-fluid. Nur wenn ε_{fg} positiv ist, ist es möglich, dass sich zwischen zwei Fluiden f und g eine Grenzfläche ausbilden kann. Im anderen Fall kommt es zu einer vollständigen Auflösung des einen Fluids in das andere: Beispielsweise mischen sich Flüssigkeiten oder Gase oder sie bilden miteinander eine Lösung. Wenn es sich um eine flüssige und gasförmige Phase des gleichen Stoffs handelt, dann wird die flüssige Phase beispielsweise verdampfen.

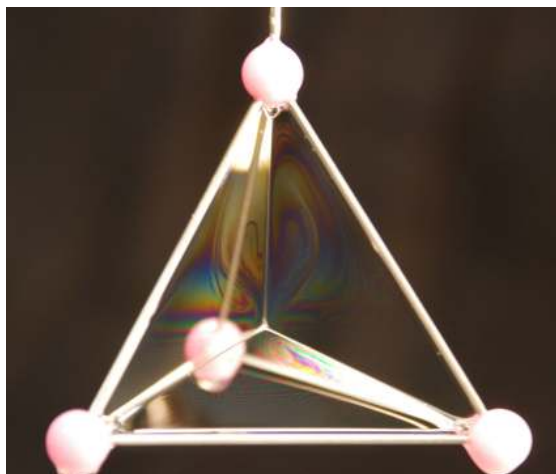


Abbildung 46: Demonstration der Ausbildung von Minimalflächen von Flüssigkeitslamellen. Die Randbedingungen sind durch einen Drahtrahmen vorgegeben.

Wenn ϵ_{fg} positiv ist und keine weiteren Energien eine Rolle spielen, nehmen Flüssigkeiten Konfigurationen mit minimaler Grenzfläche ein. Die typische Ausbildung von Minimalflächen bei gegebenen Randbedingungen (Anhaften der Flüssigkeit an einem Drahtrahmen) lassen sich durch Flüssigkeitslamellen leicht demonstrieren, deren fluid-fluid-Grenzflächenenergie wie z.B. bei Seifenwasser reduziert ist (Abb. 46). In Fluidumgebung streben Flüssigkeiten eine Kugelgestalt an, weil die Kugel bei vorgegebenem Volumen die geringste Oberfläche aller Körperformen hat. Kommen Reibungs- und Gravitationskräfte ins Spiel, weicht die Form eines Flüssigkeitstropfens von der Kugelgestalt mehr oder weniger ab. Die Kugelgestalt von Flüssigkeitstropfen lässt sich besonders schön durch nicht mischbare Flüssigkeiten gleicher Dichte demonstrieren.

fluid-fest. Die Grenzflächenenergie ϵ_{fs} zwischen einem Fluid f und einem Festkörper s (solid) kann jetzt nicht nur positiv, sondern auch negativ sein, weil sich ein Fluid gewöhnlich nicht in einem Festkörper auflöst. Ist sie positiv, dann treten für das Fluid ähnliche Phänomene wie für Fluid-Fluid-Grenzflächen auf – zumal gewöhnlich nur ein Teil der Grenzfläche mit dem festen Körper gebildet wird und ein anderer Teil mit einem anderen Fluid. Fällt ein Spritzer Quecksilber auf den Boden, dann hat es teilweise eine Grenzfläche mit dem Boden und teilweise mit der Luft. Da das Quecksilber aber trotzdem überall von Grenzflächen mit positiver Grenzflächenenergie umgeben ist, bildet es Kügelchen aus.

Meistens ist die Grenzflächenenergie jedoch negativ. Wenn das der Fall ist und die Oberfläche eines festen Körpers in ein Fluid eingetaucht war, dann haftet anschließend eine dünne Fluidschicht darauf, die sich nicht so einfach entfernen lässt. Eine Möglichkeit, das dennoch zu erreichen, besteht darin, den festen Körper zu erhitzen und so das Fluid abdampfen zu lassen. Das macht man beispielsweise, wenn man in einer Apparatur ein Ultrahochvakuum erreichen möchte.

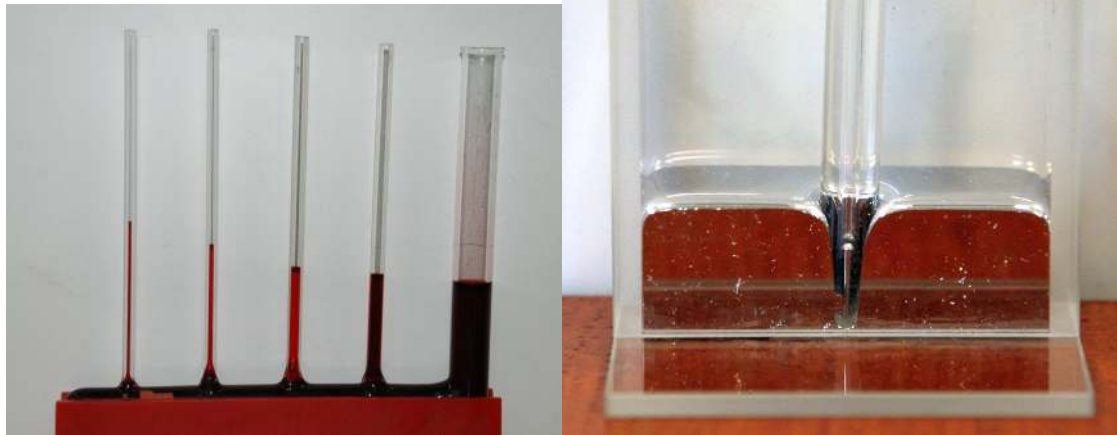


Abbildung 47: (a) Für negatives ε_{fs} nimmt die Steighöhe mit abnehmendem Kapillardurchmesser zu. Die Steighöhe ist umgekehrt proportional zum Radius r (0.8 mm, 1.3 mm, 3.0 mm, 4.2 mm und 14.5 mm). (b) Demonstration der Kapillardepression für positives ε_{fs} an der Quecksilber-Glas-Grenzfläche.

Befindet sich eine dünne Flüssigkeitsschicht zwischen zwei Glasplatten, dann lassen sich die Platten nur schwer durch orthogonal zur Plattenfläche wirkende Kräfte trennen. Es ist aber möglich, wenn sie tangential abschert, denn der Schermodul von Fluiden ist null. Wenn eine Flüssigkeit einen Teil der Grenzschicht mit einem Gas bildet (Grenzflächenenergie ε_{fg}) und einen Teil mit einem Festkörper (Grenzflächenenergie ε_{fs}), dann kommt es zur vollständigen Benetzung des Festkörpers, wenn $\varepsilon_{fs} + \varepsilon_{fg} < 0$.

Die potentiellen Energien verschiedener Grenzflächen können miteinander konkurrieren, aber auch mit anderen potentiellen Energiebeiträgen: In Kapillaren vom Radius r steigt die Flüssigkeitssäule so lange an bis bei einer Höhe h die Kosten für die aufzuwendende gravitative potentielle Energie (proportional zum Volumenelement $dV = \mathcal{A} dh = \pi r^2 dh$ auf der Höhe h gerade gleich dem Gewinn an Energie durch Bildung von Grenzfläche (proportional zum zylindrischen Flächenelement $d\mathcal{A}_G = 2\pi r dh$) ist. Im energetischen Gleichgewicht gilt:

$$\rho g h dV + \varepsilon_{fs} d\mathcal{A}_G = 0$$

Daraus ergibt sich die Steighöhe

$$h = -\frac{2\varepsilon_{fs}}{\rho g r} \quad (18.20)$$

Die Grenzflächenenergie ε_{fs} lässt sich auf diese Weise bestimmen (Abb. 47a). Ist ε_{fs} positiv wie für die Grenzfläche Quecksilber-Glas, wird h negativ und es kommt zur KAPILLARDEPRESSION (Abb. 47b).

18.7.2. Haftung

Das Phänomen der Haftung ist eines der äußerst seltenen Phänomene, wo der Begriff der Kraft tatsächlich von Bedeutung ist. Es seien zwei feste Platten A und B betrachtet, deren aneinander angrenzenden Ebenen die Flächengröße \mathcal{A} haben. Da ein fester Körper im Vergleich zu einem Fluid extrem starr ist, kann sich seine Oberfläche jedoch nicht so ohne weiteres der Oberfläche eines zweiten festen Körpers anpassen. Ferner kann man nur von Ebenen sprechen, wie sie uns nach Maßgabe der technischen Möglichkeiten erscheinen, und das sind keine planen Ebenen im mathematischen Sinn. Damit soll gesagt sein, dass zwischen zwei festen Körpern in der Realität i.A. keine perfekte Grenzfläche vorliegt. Auch wenn man große Anstrengungen unternimmt und die Oberflächen der beiden Feststoffen extrem plan und auf extrem geringe Rauheit hin poliert, so erweisen sich die durch solche mechanischen Technologien erzielbaren Oberflächen auf der mikroskopischen Skala immer noch als sehr rau und lassen keinen Kontakt über den gesamten Bereich \mathcal{A} der gegenüberliegenden Flächen zu. Im Regelfall ist die Grenzfläche \mathcal{A}_G des tatsächlichen fest-fest-Kontakts zwischen den festen Körpern fast vernachlässigbar klein: $\mathcal{A}_G \ll \mathcal{A}$. Obwohl die spezifische Grenzflächenenergie ε_{AB} zwischen den festen Körpern A und B betragsmäßig groß und negativ ist, führt dieser Umstand dazu, dass die Adhäsion zwischen ihnen nur gering ist. Über den größten Teil der Fläche, nämlich $\mathcal{A} - \mathcal{A}_G \approx \mathcal{A}$, sind die beiden Oberflächen in Kontakt mit der dazwischen befindlichen Luft. Da die Kohäsion von Gas jedoch extrem klein ist und keine Scherkräfte auftreten, trägt die Luft nicht zur Haftung der beiden Körper aneinander bei. Es gibt verschiedene Strategien, wie sich die Haftung zweier fester Körpern aneinander verbessern lässt:

1. Man bringt eine Flüssigkeit zwischen die beiden Ebenen, die eine negative spezifische Grenzflächenenergie zu beiden festen Körpern hat. Da die Kohäsion von Flüssigkeiten relativ zu Gasen extrem groß ist, ist es schwierig, die beiden Ebenen nun in der zu ihnen orthogonalen Richtung auseinander zu ziehen. Hingegen ist es leicht, A und B durch Abscheren in tangentialer Richtung zu trennen, weil Flüssigkeiten keine Scherfestigkeit haben.
2. Man bringt eine Flüssigkeit zwischen beide Ebenen, welche die Ebenen der Fläche \mathcal{A} vollständig benetzt, und lässt diese erstarren. Dann ist $\mathcal{A}_G \approx \mathcal{A}$. Das ist die grundlegende Idee beim Kleben, Löten und Schweißen. Wenn man die Ebenen voneinander trennen will, muss nun auch in tangentialer Richtung eine Kraft aufgewandt werden, deren Betrag eine kritische Schwelle

$$F_{crit} = \mu_K \mathcal{A}_G \approx \mu_K \mathcal{A} \quad (18.21)$$

übersteigt. Der Wert des Koeffizienten μ_K hängt einerseits von der Wechselwirkung des Klebstoffs mit den Oberflächen ab (Adhäsion) und zum anderen vom inneren Zusammenhalt des Klebstoffs (Kohäsion). Manchmal ist die KOHÄSION, also der Zusammenhalt eines homogenen Festkörpers A, geringer als die ADHÄSION zwischen den aneinander haftenden festen Körper A und B und die Kohäsion des Klebstoffs. Dann kommt es im Medium A zu einem BRUCH bevor der Betrag F_{crit} der Mindestkraft für die Trennung von A und B erreicht werden kann. Wenn

eine unerwünschte Klebung auftreten kann, wie z.B. beim Braten vorkommt, kann man sich durch eine Teflonbeschichtung der Pfanne dagegen vorsehen, weil für die meisten Materialien der Betrag der spezifischen Grenzflächenenergie mit Teflon sehr klein ist.

- Die wirksame Grenzfläche \mathcal{A}_G kann man dadurch vergrößern, dass man die Ebenen der Körper gegeneinander presst, also eine Kraftkomponente F_\perp orthogonal zu den in Kontakt befindlichen Ebenen wirken lässt. Empirisch findet man, dass in einem weiten Wertebereich gilt, dass die wirksame Grenzfläche \mathcal{A}_G einerseits proportional zur Fläche \mathcal{A} anwächst und andererseits zur pro Flächeneinheit wirkenden Normalkraft (also proportional zu F_\perp/\mathcal{A}) anwächst. Insgesamt gilt also $\mathcal{A}_G \propto \mathcal{A}(F_\perp/\mathcal{A}) \propto F_\perp$. Infolgedessen wächst die kritische Kraftschwelle der Presshaftung

$$F_{crit} = \mu_H F_\perp \quad (18.22)$$

allein proportional mit F_\perp . So lange der Betrag $F = |\vec{F}|$ einer an den Platten angreifenden Kraft kleiner als der kritische Wert ist, also $F < F_{crit}$, passiert gar nichts. Für $F \geq F_{crit}$ reißt die Adhäsion ab und es setzt i.A. eine beschleunigte Bewegung ein.¹⁸

Der Wert des Haftkoeffizienten μ_H ist ein empirischer Parameter, der von den beteiligten Materialien, der Einwirkungsdauer und der Oberflächenbeschaffenheit abhängen kann. Wenn man die Kraft F_\perp wieder verringert, dann geht in der Regel F_{crit} aufgrund elastischer Effekte auch wieder zurück. Das ist aber nicht immer so. Wenn man beispielsweise extrem gut polierte Oberflächen durch Zusammendrücken in sehr guten Kontakt bringt, dann bildet sich die so erzielte Kontaktfläche unter Umständen nicht mehr zurück, auch wenn man die Kontaktflächen wieder entlastet.

18.7.3. Dissipation an Grenzflächen

Der physikalische Terminus technicus „Reibung“ hat nicht unbedingt mit dem Begriff „reiben“ zu tun, auch wenn beide Worte im Deutschen verführerisch ähnlich klingen und der technische Begriff der Reibung vermutlich etymologisch aus dem Begriff des Reibens hervorgegangen ist.¹⁹ Als Fachbegriff der TRIBOLOGIE bezieht er sich auf jede Art und Weise, wie man zwei Stoffe in engen Kontakt bringen kann. Dieser enge Kontakt kann zwar auch durch Reiben zustande kommen, muss aber nicht. Reibung (im Sinne des hier verwendeten physikalischen Fachbegriffs) kann auch durch gegenseitiges Anpressen zweier Stoffe realisiert werden.

Zu den tribologischen Phänomenen gehören also die gerade besprochene Haftung und die Grenzflächenenergie sowie die Reibungselektrizität, die in Kap. 35.1 besprochen wird.

¹⁸Das ist nichts Ungewöhnliches, wie folgende Analogie zeigt: Stellen Sie einen Körper der Masse m auf den Tisch, hängen Sie eine Feder an und versuchen Sie, ihn damit nach oben zu ziehen. Auch hier kann man zunächst an der zunehmenden Auslenkung der Feder erkennen, dass eine zunehmende Kraft $F = kx$ angreift, aber es passiert erst einmal nichts. Erst wenn die Kraft F den Wert $F_{crit} = mg$ überschreitet, setzt eine beschleunigte Bewegung ein.

¹⁹Im Englischen ist das z.B. nicht so: Reiben heißt „to rub“ bzw. „to grind“ und Reibung „friction“.

Diese Phänomene demonstrieren, dass zwischen zwei Grenzflächen eine Wechselwirkung besteht. Bei einer Relativbewegung der Grenzflächen führt sie zu einer Energiedissipation. Die Energiedissipation ist unabhängig von der Richtung der Relativbewegung und wächst mit zunehmendem Betrag $|v|$ der Relativgeschwindigkeit an (s. Kap. 12.5). In diesem Kapitel sollen drei typische Phänomene der Dissipation kurz andiskutiert werden, welche bei relativ zueinander bewegten Grenzflächen beobachtet werden.

Gleitreibung. Als einfachen Modellfall diskutieren wir eine feste Platte der Fläche \mathcal{A} , die auf einer planen Oberfläche eines festen Körpers mit der Relativgeschwindigkeit v gleitet. Der dissipative Energieverlust $I_v = \dot{E}_v$ pro Zeiteinheit bzw. die Verlustleistung $P_v = \dot{E}_v$ hängt von der tatsächlichen Kontaktfläche \mathcal{A}_G zwischen den festen Körpern ab und diese ist wiederum proportional zur Kraft F_\perp , die orthogonal zur Platte wirkt (s. Kap. 18.7.2). Experimentell findet man, dass die dissipative Verlustleistung richtungsunabhängig und damit proportional zu $|v|$ anwächst und erhält man

$$P_v = -\mu_G F_\perp |v| \quad (18.23)$$

Das Minuszeichen weist auf die Abnahme der kinetischen Energie²⁰ durch GLEITREIBUNG hin. Der GLEITREIBUNGSKOEFFIZIENT μ_G hängt nicht nur von den Materialien ab, aus denen die Oberflächen bestehen, sondern auch von der Oberflächenbeschaffenheit, z.B. von der Rauheit der beiden Oberflächen.

Viskosität. Über einer Ebene befinde sich eine Fluidschicht der Dicke d . Darauf gleite eine Platte der Fläche \mathcal{A} mit einem Geschwindigkeitsbetrag v relativ zur ruhenden Ebene. Man beobachtet dabei, dass die Geschwindigkeit durch Dissipation abnimmt. Wenn man sie konstant halten will, muss man der Platte von außen gerade so viel Energie pro Zeiteinheit zuführen, wie durch die dissipative Verlustleistung verloren geht. Dazu kann man z.B. der Platte eine entsprechende externe Leistung P zuführen und so die Verlustleistung für eine gegebene Geschwindigkeit empirisch ermitteln. Es ergibt sich

$$P_v = -\eta \frac{\mathcal{A}}{d} v^2 \quad (18.24)$$

Die Proportionalitätskonstante η ist eine materialspezifische Größe des Fluids und heißt VISKOSITÄT.²¹

Ein wichtiger Modellfall ist die Relativbewegung zwischen einer Kugel mit dem Radius r und einem ausgedehnten Fluid der Viskosität η . In einem ausgedehnten Fluid erreicht die Relativgeschwindigkeit zwischen Fluid und Kugel in einiger Entfernung $d \propto r$ von der Bewegungsrichtung des Kugelzentrums den Maximalwert v . Die effektiv wirksame Fläche der Kugel skaliert mit $\mathcal{A} \propto r^2$ und daher $P_v \propto -\eta(r^2/r)v^2$. Aus einer Rechnung, die sehr

²⁰bezogen auf das Bezugssystem des Endzustands, bei dem beide Flächen wieder relativ zueinander ruhen.

²¹Dies soll vorläufig als Definition der Viskosität eines Fluids genügen. Das Phänomen der inneren Reibung eines Fluids wird in Kap. ?? genauer besprochen werden.

aufwändig ist und deshalb hier nicht präsentiert werden soll (Herleitung z.B. in [11]), erhält man für die Proportionalitätskonstante den Wert 6π und damit das STOKESSCHE REIBUNGSGESETZ

$$P_v = -6\pi r \eta v^2 \quad (18.25)$$

für kugelförmige Körper.

Kinematische Dissipation. Auch wenn sich ein Objekt in Fluiden mit vernachlässigbar kleiner Viskosität bewegt, tritt mit zunehmender Geschwindigkeit eine immer deutlicher werdende Dissipation auf. Empirisch zeigt sich oft, dass sie proportional zur Querschnittsfläche \mathcal{A}_\emptyset ist, die das Objekt orthogonal zur Richtung der Relativbewegung darbietet. Das empirische Gesetz kann man in der Form

$$P_v = \frac{1}{2} c_W \mathcal{A}_\emptyset \rho v^2 |v| \quad (18.26)$$

aufschreiben, wobei ρ die Dichte des Fluids und $\frac{1}{2}c_W$ eine dimensionslose Proportionalitätskonstante ist. Der c_W -Wert hängt vom Profil bzw. der Form des Objekts ab und muss i.A. empirisch ermittelt werden.

19. Funktionen mehrerer Variabler (M3)

19.1. Partielle Ableitung

Partielle Ableitungen und Potentiale werden hier am Beispiel einer Funktion $f(x, y)$ der beiden allgemeinen Variablen x und y erläutert. Das Beispiel kann leicht auf den Fall mehrerer Variabler ausgedehnt werden.

Wenn man $y = y_0$ konstant hält, dann kann man f auch als eine Funktion von nur einer Variablen auffassen, nämlich x , und so wie üblich differenzieren. Das Ergebnis ist die *partielle Ableitung* nach x :

$$\left(\frac{\partial f}{\partial x}\right)_{y_0} = \frac{df(x, y_0)}{dx}$$

Sie ist durch das Symbol „ ∂ “ gekennzeichnet.

Die Änderung df der Funktion $f(x, y)$ an der Stelle (x_0, y_0) ergibt sich (unter bestimmten Voraussetzungen) durch das TOTALE DIFFERENTIAL

$$df = \left(\frac{\partial f}{\partial x}\right)_{y_0} dx + \left(\frac{\partial f}{\partial y}\right)_{x_0} dy, \quad (19.1)$$

das sich aus den beiden partiellen Differentialen $\left(\frac{\partial f}{\partial x}\right)_{y_0} dx$ und $\left(\frac{\partial f}{\partial y}\right)_{x_0} dy$ zusammensetzt. Wenn das überall möglich ist, so stellt

$$df = \left(\frac{\partial f}{\partial x}\right)_y dx + \left(\frac{\partial f}{\partial y}\right)_x dy. \quad (19.2)$$

das totale Differential dar. In dieser Notation wird die jeweils bei der Differentiation konstant gehaltene Variable als Index bei den runden Klammern angegeben.²²

19.2. Potential

Falls für eine Funktion f das totale Differential df existiert, wird es auch als POTENTIAL bezeichnet. Das Potential

$$f = \int df + C \quad (19.3)$$

ist bis auf eine Konstante C bestimmt und gewissermaßen die Erweiterung des Begriffs der Stammfunktion auf den Fall der Funktionen mehrerer Variabler. Das Potential

$$f = f(x, y) = \int_{(x_0, y_0)}^{(x, y)} df \quad (19.4)$$

hat dabei für jeden beliebigen Weg in einer kompakten Definitionsmenge seiner Variablen den gleichen Wert. Es ist also wegunabhängig, und durch seine Startvariablen x_0 und y_0 bestimmt.

20. Systeme mit mehreren Freiheitsgraden

In den vorangegangenen Kapiteln haben wir verschiedene Energiebeiträge sowie geeignete Parameter kennengelernt, durch die man sie beschreiben kann. Ein Beispiel ist die durch Gl. 18.19 beschriebene Grenzflächenenergie und der Parameter Grenzfläche \mathcal{A}_G , der sie als Zustandsgröße beschreibt. Die Grenzfläche ist aber nur einer der vielen möglichen unabhängigen Parameter, welche die Energie \mathcal{U} eines thermodynamischen Systems bestimmen. Bis hier hin haben wir so getan als ob man jede einzelne Zustandsgröße isoliert betrachten kann. Das ist möglich, wenn man alle anderen Zustandsgrößen konstant hält. Dann ist die innere Energie so als ob sie die Funktion nur einer einzigen Variablen ist. Im allgemeinen ist die innere Energie jedoch eine Funktion mehrerer Variabler. Beispielsweise könnte es eine adäquate Beschreibung eines physikalischen System erforderlich machen, ein Energiemodell aufzustellen, bei dem die innere Energie eines Mediums von seiner Grenzfläche \mathcal{A}_G , seinem Volumen V und seinem Scherwinkel α abhängt, also $\mathcal{U} = \mathcal{U}(\mathcal{A}_G, V, \alpha)$ eine Funktion dieser drei Zustandsvariablen sein. Das soll hier nun allgemeiner betrachtet werden.

In einem kontinuierlichen Medium hätte man streng genommen unendlich viele Parameter zu erfassen, wenn man seine Energie vollständig beschreiben wollte. In der Praxis ist das nicht möglich. In unseren Modellen werden nur einige wenige Zustandsvariable

²²Wenn klar ist, welche Variable bei der partiellen Ableitung konstant gehalten wird, kann man diese Klammern auch weglassen.

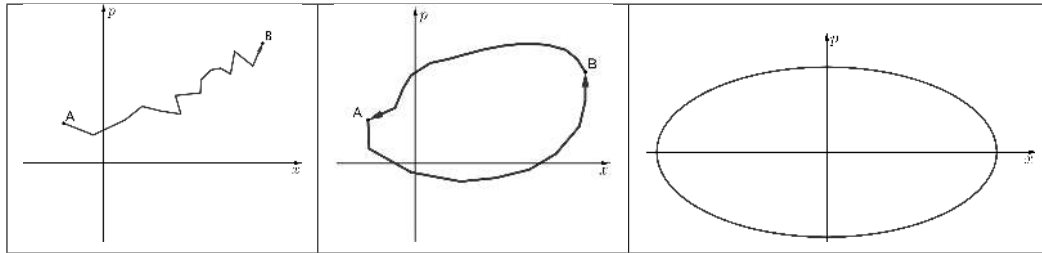


Abbildung 48: (a) Zustandsdiagramm (Phasenraumdiagramm bzw. x - p -Diagramm) eines Prozesses, der in A beginnt und in B endet. (b) Zustandsdiagramm für einen von A über B nach A zurückführenden Kreisprozess. Am Ende haben die Zustandsvariablen des Systems wieder den gleichen Wert, den sie vorher hatten. (c) Zustandsdiagramm des idealen harmonischen Oszillators für eine gegebene Gesamtenergie E_0 .

berücksichtigt, und das heißt wiederum, dass fast alle Zustandsvariable bis auf eben diese wenigen Ausnahmen ignoriert werden (Im Kontinuumsmodell bleiben also immer noch unendlich viele Parameter unberücksichtigt!). Insbesondere wird ein Parameter einfach ignoriert, wenn seine Auswirkungen so klein sind, dass sie für alle durchführbaren Messungen unterhalb der Schranke der Messunsicherheit liegt. Für energetische Modelle zieht man also i.A. eine Anzahl von f voneinander unabhängigen Zustandsparametern in Betracht, die man für ein physikalisches Problem als relevant erkannt hat und die man auch tatsächlich experimentell kontrollieren und/oder messen kann. Diese unabhängige Parameter z_1, \dots, z_f legen den ZUSTAND des Modellsystems eindeutig fest. Die innere Energie

$$\mathcal{U} = \mathcal{U}(z_1, \dots, z_f)$$

ist dann eine ZUSTANDSFUNKTION dieser f ZUSTANDSVARIABLEN bzw. ZUSTANDSKOORDINATEN. Unabhängigkeit einer Zustandsvariablen von den übrigen $f - 1$ anderen Zustandsvariablen bedeutet, dass man die $f - 1$ anderen Variablen festhalten kann und dass die Energie dann alleine eine Funktion der einen Variablen ist. Die Änderung jeder Zustandsvariablen soll also völlig frei und unabhängig von den Änderungen einer anderen Zustandsvariablen sein. Jede unabhängige Zustandsvariable stellt einen FREIHEITSGRAD des Modellsystems dar. Ein Satz unabhängiger Zustandsvariabler bildet den ZUSTANDSRAUM, d.i. der Raum aller möglicher Zustände.

Die Dimension des Zustandsraums ist gleich der Anzahl f der Freiheitsgrade des Modells der inneren Energie eines Systems.

20.1. Zustandsdiagramme

Die im Laufe eines physikalischen Prozesses eingenommenen Zustände eines Systems kann man sich in einer Art von Koordinatensystem graphisch veranschaulichen, bei dem jeder

Zustandsvariablen eine Koordinatenachse zugeordnet ist (Abb. 48a). Das ist das ZUSTANDSDIAGRAMM. Ist speziell eine Variable die Positionskoordinate x und die andere die Impulskoordinate p , so bezeichnet man das Zustandsdiagramm auch als PHASENRAUMDIAGRAMM. Wenn ein im Punkt A beginnender Prozess über mehrere Stationen im Zustandsdiagramm führt und dann schließlich wieder im Punkt A endet, so bezeichnet man ihn als KREISPROZESS (Abb. 48b). Nach Durchlaufen eines Kreisprozesses hat ein physikalisches System den gleichen Zustand und damit auch die gleiche innere Energie wie zu Beginn des Prozesses. Somit hat das System in einem Kreisprozesses genausoviel Energie von der Umgebung aufgenommen, wie es wieder an die Umgebung abgegeben hat. Ein Spezialfall davon ist der Kreisprozess eines abgeschlossenen Systems wie er beispielsweise bei der Bewegung eines idealen (dissipationsfreien) Oszillator vorlag. Die Energie eines abgeschlossenen Systems ist konstant ($dU = 0$). Wegen

$$E_0 = \frac{1}{2m}p^2 + \frac{k}{2}x^2$$

ist das Phasenraumdiagramm bzw. Zustandsdiagramm dieses Kreisprozesses eine Ellipse (Abb. 48c).

20.2. Extensive und intensive Größen

Jedes makroskopische physikalische System setzt sich aus vielen Teilsystemen zusammen bzw. kann aus vielen Teilsystemen zusammengesetzt gedacht werden. Selbst wenn man so weit in Materie hineinzoomt, dass sie hinreichend homogen erscheint, ist ein Stück dieser homogenen Materie i.A. immer noch ein Makrosystem und gedanklich in Teilsysteme zerlegbar; im Rahmen der Kontinuumsmodells der Materie darf diese Zerlegung unendlich weit fortgesetzt werden, ohne dass das Teilsystem den Status eines Makrosystems verliert, denn das kontinuumsphysikalische System ist hierin das extreme Gegenmodell des Massenpunkts: Ein Punkt hat keine Teile und ein Kontinuum deren überabzählbar unendlich viele.

Sei G ein System, das sich aus den beiden Teilsystemen A und B zusammensetzt. Eine physikalische Größe X habe für Teilsystem A den Wert X_A , für B den Wert X_B und für das Gesamtsystem den Wert X_G . Wenn sich der Wert für das Gesamtsystem gemäß

$$X_G = X_A + X_B \tag{20.1}$$

additiv aus den Werten der Teilsysteme zusammensetzt, dann ist die Größe X eine EXTENSIVE GRÖSSE. Ist X eine Zustandsvariable, die Gl. 20.1 erfüllt, so spricht man von einer EXTENSIVEN ZUSTANDSGRÖSSE bzw. extensiven Zustandsvariablen.

Wenn man Größen bildet, bei denen man zwei extensive Zustandsvariable ins Verhältnis setzt, also z.B. durcheinander dividiert, so werden diese als INTENSIVE GRÖSSEN bezeichnet. Sind diese Zustandsgrößen, so spricht man von INTENSIVEN ZUSTANDSGRÖSSEN.

Besteht ein physikalisches System aus einem einzigen Stoff, d.h. ist er chemisch homogen, so sind seine extensiven Größen proportional zur Stoffmenge (Kap. 18.2). Man kann

stoffspezifische bzw. materialspezifische Größen dadurch definieren, dass man extensive Größen durch die Stoffmenge dividiert. Diese Größen sind intensive Größen. Ein Beispiel für eine so konstruierte, für ein Material charakteristische Größe, ist ihr SPEZIFISCHES VOLUMEN $vV/m = 1/\rho$, welche gleich der reziproken Dichte ist. Das spezifische Volumen ist das Verhältnis aus dem Volumen V , welches das Medium einnimmt zur Masse m des Mediums, welches hier das Maß für die Stoffmenge ist.

Beispiele: Energie E und Impuls p wurden in der Relativitätstheorie als additive Erhaltungsgrößen konstituiert und sind daher von Natur aus extensive Größen. Die Geschwindigkeit ist beispielsweise keine extensive Größe, denn wenn ein Teilchen A die Geschwindigkeit v_A hat und ein Teilchen B die Geschwindigkeit v_B , dann ist die Geschwindigkeit des aus beiden Teilchen bestehenden Gesamtsystems i.A. nicht die Summe dieser beiden Geschwindigkeiten. Nimmt Teilsystem A das Volumen V_A ein und B das Volumen V_B , dann ist das Gesamtvolumen $V_G = V_A + V_B$. Das Volumen ist also eine extensive Größe, aber i.A. keine Erhaltungsgröße, denn das Gesamtvolumen eines abgeschlossenen Systems kann sich bei einem physikalischen Prozess ändern. Wenn man den elastischen Beitrag $-pdV$ einer Änderung $dU = -pdV$ der inneren Energie betrachtet, dann sind sowohl dU als auch dV extensiv. Der Druck p ist eine intensive Größe.

20.3. Fundamentalform der inneren Energie

Die Zustandsvariablen z_1, \dots, z_f der inneren Energie legen den Zustand eines physikalischen Systems *eindeutig* fest. Das bedeutet, dass die inneren Energie $\mathcal{U}(z_1, \dots, z_f)$ stets einen eindeutigen Wert hat. Er hängt nicht vom Weg bzw. Prozess ab, über den man zum Zustand gelangt ist. Die innere Energie \mathcal{U} ist somit ein Potential (s. Kap. 19) und hat ein vollständiges Differential

$$d\mathcal{U} = \frac{\partial \mathcal{U}}{\partial z_1} dz_1 + \dots + \frac{\partial \mathcal{U}}{\partial z_f} dz_f.$$

Auch die partiellen Ableitungen

$$Z_j = Z_j(z_1, \dots, z_f) = \frac{\partial \mathcal{U}(z_1, \dots, z_f)}{\partial z_j} \quad (20.2)$$

stellen in der Regel Zustandsvariable dar: Häufig kann man nämlich nach einer der ursprünglichen Zustandsvariablen auflösen, beispielsweise nach der Variablen z_i . Diese Variable ist dann eine Funktion $z_i = z_i(z_1, \dots, z_{i-1}, Z_j, z_{i+1}, \dots, z_f)$. Wenn man sie in die innere Energie einsetzt, dann hängt die innere Energie wieder von f Zustandsvariablen ab, wobei allerdings z_i nun durch Z_j ersetzt worden ist. Das kann sich zu einem verwirrenden Dschungel an neuen Zustandsvariablen auswachsen, denn man kann ausgehend von dieser neuen Formulierung der inneren Energie wieder neue Zustandsvariable kreieren.

Um nun ein einheitliches Beschreibungssystem zu schaffen, führt man die Fundamentalform der inneren Energie ein. Damit bezeichnet man eine standardisierte Formulierung

der inneren Energie, bei der man für die innere Energie ausschließlich extensive Zustandsvariable verwendet. Drückt man die innere Energie $\mathcal{U} = \mathcal{U}(z_1, \dots, z_f)$ in der Standardform aus, so sind die Zustandsvariablen z_1, \dots, z_f aller f Freiheitsgrade des Systems so ausgewählt, dass sie extensive Zustandsvariable sind. Die sich dann aus den partiellen Ableitungen (Gl. 20.2) sich ergebenden Zustandsvariablen sind dann allesamt keine extensiven Zustandsvariablen. Man bezeichnet sie als INTENSIVE ZUSTANDSGRÖSSEN bzw. intensive Zustandsvariablen. Durch die Fundamentalform ist jeder extensiven Zustandsgröße damit eine intensive Zustandsgröße zugeordnet.

Warum tut man das? Betrachten wir das Volumen V eines Mediums als extensive Zustandsvariable und die innere Energie eines Systems mit f Freiheitsgraden in der Form $\mathcal{U} = \mathcal{U}(V, z_2, \dots, z_f)$. Wenn mehr Variablen als nur das Volumen im Spiel sind, dann ist die Definition des Drucks, wie sie in Abschnitt 18.5.1 formuliert wurde, nicht mehr eindeutig. Um die Eindeutigkeit wieder herzustellen, vereinbart man, den Druck \mathfrak{p} ausgehend von der Fundamentalform der inneren Energie zu definieren:

$$\mathfrak{p} = \mathfrak{p}(V, z_2, \dots, z_f) = - \left(\frac{\partial \mathcal{U}}{\partial V} \right)_{z_2, \dots, z_f} \quad (20.3)$$

Der Druck ist also die partielle Ableitung nach dem Volumen einer Funktion von extensiven Zustandsvariablen, wobei diese bei der Bildung der Ableitung konstant gehalten werden.

Selbstverständlich könnte man auch einen anderen „Druck“ ausgehend von einer Formulierung der inneren Energie definieren, in der einige der Zustandsvariablen nicht extensiv sind. Nur ist dieser „Druck“ dann eben nicht der Druck. Die Fundamentalform der inneren Energie ist eine Konvention, mit der ein Gerüst aus klar definierten Begriffen für energetische Modellsysteme etabliert wird, die mehr als eine Zustandsvariable enthalten. Sie bringt Ordnung in den mit der Anzahl der Freiheitsgrade immer üppiger wuchernden Dschungel.

20.4. Konjugierte Variable

Was man mit dem Wort „konjugiert“ bezeichnet, wurde bereits in Kap. 2.2 erläutert. Wenn man von konjugierten Variablen spricht, so bezeichnet man damit ein Paar von Variablen, die in einer symmetrischen wechselseitige Beziehung zueinander stehen. Die Legendre-Transformation ermöglicht es, eine solche Beziehung herzustellen. Darauf soll hier aber nicht im Detail eingegangen werden. Für Interessierte sei hier auf den Anhang verwiesen (Kap. E). Hier soll nur die grundsätzliche Definition konjugierter Variabler skizzenhaft vorgestellt werden:

Die Legendre-Transformation stellt eine symmetrische Beziehung zwischen zwei Funktionen (bzw. Potentialen) her, so dass die (partielle) Ableitung nach der Variablen der einen Funktion gerade die Variable der anderen Größe ist. Wenn eine solche Beziehung besteht, dann bilden sowohl die beiden Funktionen als auch die beiden Variablen ein konjugiertes Paar.

Beispiel 1: Im Spezialfall können die beiden Funktionen zusammenfallen. Das ist z.B. bei der inneren Energie

$$\mathcal{U}(p, x) = \mathcal{U}_0 + \frac{p^2}{2m} + \frac{1}{2}kx^2$$

eines harmonischen Oszillators der Fall. Man bildet die partielle Ableitung nach der Impulsvariablen, erhält man die Geschwindigkeit

$$v = \frac{\partial \mathcal{U}}{\partial p} = p/m = v(p).$$

Substituiert man die Geschwindigkeit in die innere Energie, so erhält man die innere Energie in einer Formulierung durch die Geschwindigkeitskoordinate:

$$\mathcal{U}(v, x) = \mathcal{U}_0 + \frac{1}{2}mv^2 + \frac{1}{2}kx^2.$$

Die partielle Ableitung ergibt

$$p = \frac{\partial \mathcal{U}}{\partial v} = mv = p(v).$$

Impuls und Geschwindigkeit stellen hier ein konjugiertes Paar von Zustandsvariablen dar, wobei der Impuls die extensive und die Geschwindigkeit die intensive Zustandsvariable ist.

Beispiel 2: Das physikalische System bestehe nur aus einem einzigen Teilchen, das im gewählten Inertialsystem den Impuls p hat. Innere Energie \mathcal{U} und Energie E haben in diesem Fall die gleiche Bedeutung. Die Masse im Ausdruck für die Energie $E(p) = \sqrt{m^2c^4 + p^2c^2}$ sei eine Konstante und die Energie damit allein eine Funktion des Impulses. Das ist hier die Zustandsvariable. Das Differential der Funktion ist

$$dE = Z(p)dp.$$

Da nur eine einzige Zustandsvariable p vorliegt, ist die partielle Ableitung gleich der gewöhnlichen Ableitung

$$Z = \frac{dE}{dp} = \frac{pc}{\sqrt{m^2c^4 + c^2p^2}}.$$

Die Größe Z ist wiederum eine Zustandsvariable. Vergleicht mit Gl. 11.16, so erkennt man, dass Z gleich der Geschwindigkeit v ist. Indem man den Impuls durch die Geschwindigkeit ausdrückt und in den Ausdruck für die Energie einsetzt, kann man durch bilden der Ableitung zeigen, dass auch hier Impuls und Geschwindigkeit ein konjugiertes Variablenpaar bilden. Das Differential der Energie kann in der Form

$$dE = vdp$$

mit der Geschwindigkeit

$$v(p) = \frac{dE}{dp}$$

geschrieben werden, wobei p die extensive und v die intensive konjugierte Variable ist.

Alle Energieformen der Mechanik gehören entweder zu den kinetischen oder zu den potentiellen Energieformen und werden i.A. durch intensive oder extensive Zustandsvariable beschrieben. Für Formen der potentiellen Energie sind das üblicherweise Lage- oder Positionsvariable, wie z.B. die Ortskoordinate x als intensive Zustandsvariable oder wie z.B. das Volumen V als extensive Zustandsvariable. Die Zustandsvariablen aller Formen kinetischer Energien sind üblicherweise primär zeitliche Ableitungen solcher Koordinaten, wie z.B. die Geschwindigkeitskoordinate $v = \dot{x}$. Kinetische Energieformen sind, wie der Name schon sagt, Träger der Kinetik. Die besondere Bedeutung der kinetischen Energie liegt darin, dass ohne sie keine Energie von einem Ort zu einem anderen übertragen werden kann und auch keine Energie von einer potentiellen Energieform auf eine andere. Zeitliche Änderungen von Zuständen können nur unter Beteiligung einer kinetischen Energie ablaufen. Anstelle einer Geschwindigkeit kann man auch den dazu konjugierten Impuls heranziehen, um eine kinetische Energie aufzustellen. Beides hat Vor- und Nachteile. Theoretiker ziehen extensive Zustandsgrößen wie die Impulse vor, weil für extensive Zustandsgrößen klar ist, wie sich die Zustandsgröße des Gesamtsystems aus den Werten der Teilsysteme ergibt. Für den Experimentator spielen intensiven Zustandsgröße eine größere Rolle, weil beispielsweise Geschwindigkeiten als zeitliche Änderung einer mechanischen Positionsvariable unmittelbar beobachtbar sind.

Geht man in der Kontinuumsmechanik von der Fundamentalform für die innere Energie aus, dann sind die extensive und die intensive Zustandsvariable stets konjugierte Variable zueinander.

Beispiel: Die innere Energie sei in der Fundamentalform durch $\mathcal{U}(V)$ mit dem Volumen als extensiver Zustandsvariabler gegeben. Dieses Modell hat nur einen Freiheitsgrad. Die Ableitung liefert den Druck

$$\mathbf{p} = -\frac{d\mathcal{U}}{dV}$$

als intensive Zustandsvariable.

Für das Differential der Enthalpie wurde in Kap. 18.6 gezeigt, dass

$$d\mathcal{H} = Vd\mathbf{p}.$$

Wenn man die Enthalpie speziell so formuliert, dass der Druck \mathbf{p} die Zustandsvariable ist, so gilt für die Ableitung

$$\frac{d\mathcal{H}}{d\mathbf{p}} = V.$$

Damit erweisen sich die innere Energie $\mathcal{U}(V)$ und die Enthalpie $\mathcal{H}(\mathbf{p})$ als wechselseitig konjugierte Funktionen, deren Variable konjugierte Zustandsvariable sind. Man kann zeigen, dass die beiden Funktionen Legendre-Transformierte zueinander sind (s. Kap. E).

20.5. Warum und wie werden neue Zustandsvariable eingeführt?

Ein System soll zunächst durch $f = n$ Freiheitsgrade zufriedenstellend beschrieben sein und somit durch eine innere Energie $\mathcal{U}(z_1, \dots, z_n)$. Wenn man nun entdeckt, dass neben

diesen bekannten und durch die extensiven Zustandsvariablen z_1, \dots, z_n beschriebenen Energiebeiträge eine weitere, damit noch nicht erfasste Energieform im System schlummert, so ergänzt man das Energiemodell um einen weiteren Freiheitsgrad z_{n+1} . Dem System werden infolgedessen nun $f = n + 1$ Freiheitsgrade zugeschrieben, d.h. seine innere Energie

$$\mathcal{U} = \mathcal{U}(z_1, \dots, z_n, z_{n+1})$$

ist nun eine Funktion von $n + 1$ Zustandsvariablen z_1, \dots, z_{n+1} . Aus Konsistenzgründen geht man von der Fundamentalform aus und legt die neue Zustandsvariable $z_{n+1} = z_f$ als extensive Zustandsvariable fest. Der Beitrag des neuen Freiheitsgrads zum Differential $d\mathcal{U}$ ist

$$Z_f dz_f \tag{20.4}$$

mit der zu z_f konjugierten intensiven Zustandsvariablen

$$Z_f = \frac{\partial \mathcal{U}(z_1, \dots, z_n, z_f)}{\partial z_f}.$$

Existenz und grundlegende Eigenschaften eines neuen energetischen Freiheitsgrades können meist aus dem Verhalten der bekannten n Energiebeiträge erschlossen werden. Nachfolgend werden drei Beispiele für die Konstruktion neuer (nicht-mechanischer) Energiebeiträge vorgestellt, nämlich der thermische Energiebeitrag, der chemische Energiebeitrag und der elektrische Energiebeitrag. Als jeweils extensive Zustandsvariable führt man die Entropie S , die Stoffmenge n bzw. die elektrische Ladung q ein.

Teil IV.

Phänomenologische Thermodynamik

Die PHÄNOMENOLOGISCHE THERMODYNAMIK befasst sich mit makroskopischen Systemen. Sie werden grundsätzlich in ihrem Schwerpunktsystem (Kap. 11.5) beschrieben. Die Energie eines makroskopischen Systems der Masse m ist somit gleich der relativistisch invarianten inneren Energie $U = mc^2$. Daher wird im Folgenden das Wort „innere“ oft weglassen und einfacher von der Energie U eines thermodynamischen Systems gesprochen. Im Rahmen der phänomenologischen Thermodynamik wird Materie als ein kontinuierliches Medium aufgefasst und das Kontinuumsmodell (Kap. 18.1) zugrunde gelegt.

21. Die Hauptsätze der Thermodynamik

In diesem Abschnitt wird eine neue Energieform eingeführt, welche gleichberechtigt neben die mechanischen Energieformen tritt und das mechanische Energiemodell zum thermodynamischen Energiemodell ergänzt. Das Vorgehen ist das Musterbeispiel für die Einführung neuer Energieformen und wird sich später noch an vielen Beispielen in ähnlicher Weise wiederholen (chemische Energie, elektrische Energie, usw.). Ausgangspunkt sei ein mechanisches Modellsystem, für welches zunächst n unabhängige phänomenologische Parameter bzw. phänomenologische Zustandsvariable zur Charakterisierung des mechanischen Energiezustands des Systems identifiziert werden konnten. Diese phänomenologischen bzw. „sichtbaren“ Parameter repräsentieren $f = n$ unabhängige energetische Freiheitsgrade des mechanischen Systems. In Kap. 14.3 wurde z.B. für reale makroskopische schwingende Systeme ein Modellsystem entworfen, das nur aus $f = 2$ Freiheitsgraden mit der Koordinate x und der Geschwindigkeit \dot{x} als Zustandsvariablen. Es erscheint unglaublich und fast schon arrogant, wenn Physiker behaupten, sie könnten ein reales naturgegebenes Objekt mit all seiner inneren Komplexität auf nur zwei Zustandsvariable abbilden. Und dennoch sind sie damit erfolgreich. Aber wir haben auch bereits gesehen, dass das nicht perfekt funktioniert, denn es treten dissipative „Energieverluste“ auf, die man im Laufe der Zeit irgendwann nicht mehr ignorieren kann.

Hält man am Modell fest, dann kann man sie als offene Systeme betrachten, welche die dissipierte Energie nach außen abgeben. Man kann aber auch einen anderen Standpunkt einnehmen und danach trachten, das offene System zu einem abgeschlossenen System zu vervollständigen, indem man die dissipierte Energie ins Modell einschließt. Dazu muss man der Fundamentalförm der inneren Energie mindestens einen neuen Freiheitsgrad hinzufügen. Der neue Energiebeitrag soll als THERMISCHE ENERGIE bezeichnet werden. Genau genommen ist die thermische Energie der Sammelbegriff für eine gigantisch große

Zahl an winzig kleinen Energiebeiträgen, für die man keine „sichtbaren“ Zustandsvariablen angeben kann. Sie ist die Zusammenfassung sehr vieler mikroskopischer Energiebeiträge, deren Änderungen so winzig klein sind, dass man sie auf der makroskopischen Ebene nicht durch Zustandsvariable erfassen kann. Dennoch wirken sie sich im Laufe der Zeit zunehmend makroskopisch aus. Obwohl ihre Energiebeiträge im einzelnen vernachlässigbar klein sind, ergeben sie dennoch den makroskopisch an der Zeitentwicklung der n Freiheitsgrade beobachtbaren Effekt der Dissipation, weil extrem viele (im Kontinuumsmodell unendlich viele) dieser individuell nicht fassbaren Freiheitsgrade an der Dissipation beteiligt sind.

Im einfachsten Fall modelliert man alle diese nicht erfassbaren Beiträge durch einen einzigen Energiebeitrag, den thermischen Energiebeitrag, und beschreibt ihn in der Fundamentalform der inneren Energie durch eine extensive Zustandsvariable S , der man den Namen ENTROPIE gibt. Fundamentalform der inneren Energie \mathcal{U} eines thermodynamischen Systems besteht also aus der neuen extensiven Zustandsvariablen S und den n anderen extensiven Zustandsvariablen z_1, z_2, \dots, z_n , also aus insgesamt $f = n + 1$ Freiheitsgraden:

$$\mathcal{U} = \mathcal{U}(S, z_1, z_2, \dots, z_n). \quad (21.1)$$

Die zur Entropie S konjugierten Zustandsgröße

$$T = T(S, z_1, z_2, \dots) = \left(\frac{\partial \mathcal{U}}{\partial S} \right)_{z_1, z_2, \dots} \quad (21.2)$$

bezeichnet man als TEMPERATUR. Eine Änderung der inneren Energie wird also durch

$$d\mathcal{U} = TdS + \frac{\partial \mathcal{U}}{\partial z_1} dz_1 + \frac{\partial \mathcal{U}}{\partial z_2} dz_2 + \dots + \frac{\partial \mathcal{U}}{\partial z_n} dz_n \quad (21.3)$$

beschrieben mit dem thermischen Beitrag TdS zum Energiedifferential.

Mit der Zuschreibung einer neuen Energieform und ihrer zueinander konjugierten Zustandsvariablen allein ist es aber nicht getan, denn das wesentliche Charakteristikum dissipativer Prozesse ist, IRREVERSIBEL zu sein. Wenn man ein abgeschlossenes System ($d\mathcal{U} = 0$) mit der Energiebilanz

$$TdS + \frac{\partial \mathcal{U}}{\partial z_1} dz_1 + \frac{\partial \mathcal{U}}{\partial z_2} dz_2 + \dots + \frac{\partial \mathcal{U}}{\partial z_n} dz_n = 0 \quad (21.4)$$

betrachtet, so unterscheidet sich die thermische Energie dadurch von allen anderen Energieformen, dass sie die besondere Eigenschaft hat, eine äußerst selbstsüchtige Energieform zu sein: sie nimmt von den anderen Energieformen auf, gibt aber nichts zurück: Sie ist sozusagen der Dagobert Duck aller Energieformen.

Zentrale Aufgabe der Thermodynamik ist die Modellierung der an makroskopischen Systemen beobachteten Irreversibilität und die Ausarbeitung ihrer Konsequenzen.

In einem abgeschlossenen System makroskopischer Dimension geht es also einerseits darum, dass man im Laufe der Zeit eine Abnahme der durch mechanische Zustandsvariable beschriebenen Energiebeiträge beobachtet wird und dass die thermische Energie das energetische Gegenkonto dazu liefert. Bei einem solchen Vorgang nimmt die thermische Energie zu. Der umgekehrte Vorgang, dass nämlich die totale mechanische Energie zunimmt und infolgedessen auf eine entsprechende Abnahme der thermischen Energie geschlossen werden müsste, wird erfahrungsgemäß nicht beobachtet. Das macht die Irreversibilität aus und diese Erfahrungstatsache für makroskopische Systeme muss nun noch modellmäßig formuliert werden. Das geht verblüffend einfach. Da über die Temperatur T bis hier hin nichts festgelegt ist, setzt man die Annahme, dass T positiv ist. Das irreversible Anwachsen des thermischen Beitrags TdS ist folglich modelliert, wenn man als weitere Annahme setzt, dass die Entropie in abgeschlossenen Systemen nicht abnehmen darf, d.h. dass in allen Prozessen, welche man in der Natur für abgeschlossene Systeme beobachtet, $dS \geq 0$ gilt. Das grundlegende Modell der Thermodynamik beruht somit auf einer axiomatische Basis, die durch die folgenden drei HAUPTSÄTZE DER THERMODYNAMIK formuliert ist:

Hauptsätze der Thermodynamik

1. Existenz der thermischen Energie

Um die Energie eines abgeschlossenen Systems vollständig zu beschreiben, muss neben den mechanischen Energiebeiträgen eine thermische Energie postuliert werden, welche die daraus dissipierte Energie kontiert.

2. Verbot der Entropieabnahme

Die extensive Zustandsvariable der thermischen Energie heißt Entropie. In abgeschlossenen Systemen ist eine Abnahme der Entropie verboten.

3. Positivität der Temperatur

Die intensive Zustandsvariable der thermischen Energie heißt Temperatur. Sie ist als positive Größe definiert.

Die Deklaration der Positivität der Temperatur ist eine reine Konvention. Man hätte die Temperatur auch als negative Größe definieren können, aber dann hätte man für den zweiten Hauptsatz eine Entropiezunahme verbieten müssen. Wenn die Positivität der Temperatur konventionsgemäß gesetzt ist, dann ist das Verbot der Entropieabnahme das zentrale Axiom der Thermodynamik ist, denn es modelliert die Irreversibilität, die man an abgeschlossenen makroskopischen Systemen beobachten kann. Wenn eine der f nichtthermischen Energieformen Energie an den thermischen Energiebeitrag TdS abgibt und Gl. 21.4 erfüllt, dann ist das wegen $dS \geq 0$ ein zulässiger Prozess. Da eine Abnahme der Entropie nach dem zweiten Hauptsatz verboten ist und die Temperatur definitionsgemäß positiv ist, kann der umgekehrte Prozess aber nicht ablaufen: Keine der f nichtthermischen Energieformen kann Energie von der thermischen Energieform zurückgewinnen!

Der Prozess der Dissipation ist im Rahmen des durch die drei Hauptsätze axiomatisch eingeführten Modells irreversibel.

Wenn in einem abgeschlossenen System G physikalische Prozesse stattfinden, bei denen die Entropie konstant bleibt (isentropen Prozesse), so sind sie prinzipiell reversibel. Sie können genauso gut in der einen wie in der umgekehrten Richtung ablaufen. Energieübertragungen von einer mechanischen Energieform in eine andere sind prinzipiell in der einen wie in der anderen Richtung möglich. Wird jedoch mechanische Energie so abgegeben, dass die Entropie dabei anwächst, so ist der Rückweg verbaut, weil die Entropie nicht abnehmen darf.

Die Erfahrung zeigt, dass während des Ablaufs eines makroskopischen Prozesses stets ein wenig von den kinetischen Energiebeiträgen irreversibel dissipiert wird. Im Modell der Thermodynamik wird dieses Phänomen auf ein Anwachsen der Entropie zurückgeführt. Der Abfluss von kinetischer Energie an die thermische Energie wirkt so wie eine Art Mehrwertsteuer auf kinetische Energie. Aus leidvoller Erfahrung mit Finanzämtern ist ja bekannt, dass es zum Wesen der Steuer gehört, dass man Geld an das Finanzamt zahlt, aber nicht vom Finanzamt bekommt. Der Vorgang ist irreversibel.

Wer auch immer sich bewegt, wirtschaftet, weder rastet noch ruht, der zahlt Steuer. Wer sich nicht bewegt, nicht. Steuern zahlt man nur so lange, wie man, jung und dynamisch, auch tatsächlich wirtschaftlich tätig sein kann. Analog laufen physikalische Prozesse nur so lange, wie sich mechanische Energieformen ändern können bzw. so lange, wie die Prozessdynamik nicht zum Stillstand gekommen ist. Wenn schließlich aber irgendwelche Restriktionen, denen die mechanischen Zustandsvariablen unterliegen, dafür sorgen, dass keine weitere Energie aus den mechanischen Energieformen mehr dissipiert werden kann, dann erreicht das thermodynamische System seinen stationären Zustand: Es wird keine weitere zeitliche Änderung von Zustandsvariablen mehr beobachtet. Dieser zeitlich stationäre Zustand wird als THERMODYNAMISCHES GLEICHGEWICHT bezeichnet. Thermodynamische Prozesse können nur so lange ablaufen, bis die Entropie im Rahmen der Restriktionen, denen die mechanischen Zustandsvariablen unterliegen, das Maximum angenommen hat. Es ist also gleichbedeutend zu sagen, dass ein physikalisches System das thermodynamische Gleichgewicht erreicht hat, oder zu sagen, dass es die nach Maßgabe der Restriktionen maximale Entropie erreicht hat.

Wenn man im Sportstudio einen Ball in der Hand hält, hat dieser eine gewisse potentielle Energie. Wenn man ihn loslässt, wird diese teilweise in kinetische Energie umgewandelt und teilweise dissipiert. Da man das Sportstudio getrost als abgeschlossenes System betrachten kann, modelliert die Thermodynamik die Energieerhaltung dadurch, dass der Verlust an kinetischer Energie durch eine Zunahme der thermischen Energieform bilanziert wird. Die Irreversibilität des Vorgangs wird durch ein Anwachsen der Entropie modelliert. Es laufen so lange thermodynamische Prozesse ab, bei denen kinetische Energie reversibel in potentielle Energie und zurück umgewandelt wird und zugleich stets ein wenig davon irreversibel in thermische Energie überführt wird, bis der Ball nach einer gewissen Zeit schließlich auf dem Boden zur Ruhe gekommen ist und alle makroskopische und damit sichtbare Bewegung aufhört. Damit wird am Ende schließlich ein stationärer Zustand erreicht. Das ist der thermodynamische Gleichgewichtszustand unter der gege-

benen Restriktion, welche der Boden erzwingt. Die Entropie im Sportstudio ist dann maximal.

Es sei nun angenommen, dass das thermodynamische Gleichgewicht unter den vorgegebenen Restriktionen erreicht wurde und man keine Änderung von mechanischen Zustandsvariablen mehr feststellt. Wenn man nun zu irgendeinem Zeitpunkt eine der Restriktionen für die mechanischen Zustandsvariablen aufhebt, dann werden wieder dynamische Prozesse aufleben und so lange weiterlaufen, bis die Entropie im Rahmen der verbleibenden Restriktionen wieder ihr Maximum erreicht hat. Das durch $dS = 0$ gekennzeichnete Entropiemaximum ist das Kriterium dafür, dass das neue thermodynamische Gleichgewicht unter der Maßgabe der verbleibenden Restriktionen erreicht worden ist.

Wenn der Boden des Sportstudios eine Falltür zu einem Kellerraum hat (so etwas haben zugegebenermaßen die wenigsten Sportstudios), dann stellt das Öffnen der Falltür die Aufhebung der bisherigen Restriktion dar. Nach dem Öffnen wird sich der Ball wieder bewegen und der Vorgang der Umwandlung von potentieller in kinetische Energie (von der wieder ein bisschen Energie abgezuckt bzw. dissipiert wird) wieder losgehen. Er wird so lange unter Anwachsen der Entropie ablaufen, bis die Entropie des Sportstudios ein neues Maximum erreicht hat. Im neuen thermodynamischen Gleichgewicht wird der Ball auf dem Kellerboden des Studios liegen und dort ruhen, weil eine weitere Energieumwandlung durch den als Restriktion wirkenden Kellerboden unterbunden ist. Die ursprüngliche potentielle Energie, die abgegeben wurde, ist auch weiterhin im abgeschlossenen Sportstudio, nur eben auf viele „unsichtbare“ Energiebeiträge verteilt.

22. Physiker und ihre Verbotstafeln

Typische Verbote der Physik sind oft so formuliert wie der zweite Hauptsatz der Thermodynamik. Ein anderes Beispiel für Verbote sind die Erhaltungssätze wie z.B. der Energie- oder Impulserhaltungssatz, auch wenn sie positiv formuliert sind. Man kann sie nämlich auch negativ formulieren: alle Vorgänge, bei denen die Energie oder der Impuls eines abgeschlossenen Systems nicht erhalten sind, sind verboten. Sie werden später noch andere Erhaltungssätze kennenlernen, wie z.B. die Ladungserhaltung oder Leptonenzahlerhaltung, die im Grunde genommen gewisse physikalische Prozesse durch Verbote ausschließen, die man erfahrungsgemäß nicht beobachtet.

Woher wissen die Physiker eigentlich so genau, dass alle anderen Vorgänge verboten sind? Woher weiß man, dass das von Wolfgang Pauli formulierte Pauliverbot absolute Gültigkeit hat? Nun, die Physiker wissen das eigentlich nicht. Es sind Hypothesen. Aber es sind solche, von denen bislang noch nie beobachtet wurde, dass es auch nur einen einzigen Vorgang gibt, der ihnen widerspricht. Ihre Glaubwürdigkeit wächst mit dem Gewicht der Physikgeschichte und mit der Tätigkeit der vielen Physiker auf der Welt, welche diese Hypothesen gewissermaßen täglich auf den Prüfstand stellen.

Und doch ist es in der Physikgeschichte vorgekommen, dass eine Beobachtung auftrat, die einer dieser Hypothesen widersprach, obwohl ganze Physikergenerationen über Jahrzehnte an die absolute Gültigkeit des Verbots geglaubt haben, von der Wahrheit der

Verbotshypothese überzeugt waren. Ein Beispiel hierfür ist die Paritätserhaltung. Man ging bis Mitte des 20. Jahrhundert (!) davon aus, dass physikalische Gesetze symmetrisch gegenüber Raumspiegelungen sind. Das bedeutet, dass, wenn man sich einen physikalischen Vorgang in einem Spiegel betrachtet, man ihn in der gleichen Weise in der Natur auffinden kann, wie er im Spiegel aussieht. Die Physikerin Chien-Shiung Wu konnte jedoch experimentell zeigen, dass das für Vorgänge der schwachen Wechselwirkung nicht richtig ist.

Wenn Physiker auch nur eine einzige Ausnahme von einem ihrer als Verbote formulierten Hypothesen feststellen, dann geben sie die Hypothese schlicht und einfach auf oder schränken den Gültigkeitsbereich ein. Verbote muss man ohnehin immer mit dem Beipackzettel lesen: Das Verbot der Entropieabnahme gilt *nur für abgeschlossene Systeme!* Auf dem neuen Beipackzettel für die Paritätserhaltung steht heute: Alle Wechselwirkungen erfüllen die Paritätserhaltung – mit Ausnahme der schwachen Wechselwirkung. Physiker glauben an ihre Hypothesen, aber nicht so, wie ein religiöser Mensch felsenfest seinem Glauben anhängt, weil er einer überlieferten Schrift absolutes Vertrauen schenkt. Physiker sind sich bewusst, dass sie Suchende sind. Sie lassen die Falsifizierbarkeit ihrer Hypothesen zu. Ihr „Glaube“ ist ein vorläufiger, der durch bessere Einsichten umgestoßen werden kann. Genau das macht die Physik zu einer Wissenschaft und eine Religion nicht.

Aber wie geht man mit Vorgängen um, welche durch Verbote der Physik nicht ausgeschlossen sind? Hier schlägt das positive Rechtsverständnis der Physiker zu: Was nicht verboten ist, ist erlaubt! Wenn man für bestimmte Prozesse kein Verbot ausspricht, dann bedeutet das, dass man nicht vollständig ausschließen kann, dass sie auftreten könnten. Ein Beispiel ist die in der makroskopischen Welt beobachtete Nicht-Reversibilität von Prozessen wie z.B. beim gedämpften Oszillator. Die Erfahrung ist, dass man die Dämpfung mit technischen Verbesserungen und Tricks verkleinern kann. Sie hängt gewissermaßen von der Ingenieurskunst ab. Daher kann man keine untere Schranke von vorne herein angeben, von der sicher ist, dass sie nicht unterschritten werden könnte. Somit muss alles bis hin zum Grenzfall des ungedämpften Oszillators zu den Möglichkeiten gezählt werden, welche in der Natur auftreten könnten. Genauso verhält es sich mit den reversiblen Prozessen der Thermodynamik: Sie sind der Grenzfall der erlaubten und damit denkbaren Prozesse, auch wenn sie im Regelfall nicht auftreten. Sie stehen für theoretische Schlüsse zur Verfügung, denn verboten sind sie nicht.

Interessanterweise realisiert die Natur tatsächlich fast immer das, was nicht verboten ist. Als die Physiker Tsung-Dao Lee und Chen Ning Yang erkannten, dass die Paritätserhaltung für die schwache Wechselwirkung nicht zwingend war, da hieß das nur, dass so ein Prozess denkmöglich ist, aber bedeutete eigentlich noch nicht, dass er von der Natur tatsächlich realisiert ist. Das bewies erst das Experiment der Physikerin Wu. Den Nobelpreis gab es jedoch für die beiden Theoretiker, weil sie uns von der Denkschranke der Paritätserhaltung befreit hatten.

Mit der Verbotstafel der Entropieabnahme, welche durch den 2. Hauptsatz der Thermodynamik formuliert ist, hat es eine besondere Bewandtnis: Wie Sie später noch lernen werden, gibt es so ein Verbot streng genommen eigentlich nicht. Es ist eine ad-hoc-Hypothese, um die überwältigende Zahl an Erfahrungsberichten zur Dissipation in der

Alltags- und Küchenphysik zu modellieren. Es ist eine ad-hoc-Hypothese für die Erfahrung an makroskopischen – je größer desto besser – Systemen. Im Folgenden sollen die Konsequenzen des durch die Hauptsätze der Thermodynamik definierten physikalischen Denkmodells ausgearbeitet werden. Sie beschreiben in der Tat in hervorragender Weise die Erfahrungen mit makroskopischen Systemen.

23. Arbeit und Wärme

Mit der Einführung der Entropie müssen viele Begriffe der Mechanik nachgeschärft werden. Beispielsweise ist nun der Druck (Kap. 20.3) nun durch

$$p = p(S, V, z_2 \dots z_n) = - \left(\frac{\partial \mathcal{U}}{\partial V} \right)_{S, z_2 \dots z_n} \quad (23.1)$$

definiert (mit $z_1 = V$), d.h. als partielle Ableitung der inneren Energie nach dem Volumen für konstant gehaltene Entropie. Dass der Druck i.A. nicht allein eine Funktion des Volumens ist, sondern auch der Entropie und anderer Zustandsvariablen ist, liegt an materialspezifischen Kopplungen zwischen den Zustandsvariablen von Materie. Sie i.A. sehr komplexe Ursachen. Daher muss die Abhängigkeit des Drucks von den als relevant erkannten Zustandsvariablen trotz der großen Fortschritte der Computational Physics in den letzten Jahrzehnten meist erst einmal experimentell ermittelt werden.

Für offene thermodynamische Systeme, also Systeme, welche Energie mit der Umgebung austauschen, muss insbesondere auch der Begriff der Arbeit W aus der Mechanik (Kap. 16) nachgeschärft werden: Die Abgabe bzw. Aufnahme von Energie durch ein offenes thermodynamisches System stellt nur dann reine Arbeit W dar, wenn bei dem Prozess die Entropie konstant bleibt!

Im Gegensatz zu abgeschlossenen Systemen ist die Abnahme der Entropie für offene thermodynamische Systeme nicht verboten. Wenn ein offenes System Energie nach außen abgibt oder von außen aufnimmt und sich dabei nur seine Zustandsvariable Entropie und sonst keine weitere Zustandsvariable ändert, dann wird ausschließlich WÄRME zugeführt bzw. abgegeben. Die Wärme wird im Folgenden mit dem Symbol Q bezeichnet. Wie schon gesagt, spricht der zweite Hauptsatz über offene thermodynamische Systeme keine Einschränkung aus (er bezieht sich nur auf abgeschlossene Systeme!). Daher kann die Entropie eines Systems nicht nur zunehmen, indem Wärme von außen zufließt, sondern sie kann auch abnehmen, indem Wärme nach außen abfließt!

Die innere Energie eines offenen Systems nimmt zu, indem ihm Arbeit oder Wärme zugeführt wird. Sie nimmt ab, wenn das System Arbeit leistet oder wenn es Wärme abgibt. Durch die Begriffe Wärme und Arbeit wird eine durch das vollständige Differential beschriebene differentielle Änderung

$$d\mathcal{U} = TdS + Z_1 dz_1 + \dots + Z_f dz_n = \delta Q + \delta W \quad (23.2)$$

der inneren Energie aufgeteilt in eine Änderung des partiellen Differentials

$$TdS = \delta Q, \quad (23.3)$$

welches durch Zufuhr von Wärme verursacht wird, und eine Änderung des partiellen Differentials

$$Z_1 dz_1 + \dots + Z_f dz_n = \delta W, \quad (23.4)$$

welches auf die Zufuhr von Arbeit zurückzuführen ist. Wenn sich die Entropie eines thermodynamischen Systems verringern soll, so geht das nur im Zusammenspiel mit einem zweiten System, dessen Entropie mindestens um den gleichen Betrag zunimmt. Die Entropieänderung dS des ersten Systems ist dann negativ. Da die Temperatur immer positiv ist, muss das erste System notwendigerweise dabei auch Energie abgeben abgeben. Das ist die als Wärme bezeichnete Energieabgabe.

Nun ist die Temperatur keine Konstante, sondern eine Funktion der Entropie. Wenn ein System seine Entropie von einem Anfangswert S_α zu einem Endwert²³ S_ω ändert und alle anderen Zustandsvariablen unverändert bleiben, dann ergibt sich die zugeführte Wärme dadurch, dass alle differentiellen Beiträge aufintegriert werden:

$$Q = \int_{S_\alpha}^{S_\omega} T(S) dS.$$

Prozesse, bei denen ein Austausch von Wärme unterbunden ist, werden als ADIABATISCHE PROZESSE bezeichnet. Wenn sich dabei nur eine der extensiven Zustandsvariablen ändert, die hier mit z bezeichnet sein soll, so ist ihre konjugierte Zustandsvariable Z allein eine Funktion von z . Dann erhält man die zugeführte Arbeit durch das Integral

$$W = \int_{z_\alpha}^{z_\omega} Z(z) dz. \quad (23.5)$$

Die nachfolgenden drei Ausdrücke sind konkrete Beispiele für Gl. 23.5, d.h. Beispiele dafür, in welcher Weise Arbeit bei einer adiabatischen Zustandsänderung aufgenommen

²³Der Buchstabe ω ist der letzte Buchstabe im mit $\alpha, \beta \dots$ bzw. „alfa,beta“ beginnenden griechischen Alfa bet.

werden kann:

$$W = \int_{x_\alpha}^{x_\omega} F dx$$

$$W = \int_{V_\alpha}^{V_\omega} \mathbf{p} dV$$

$$W = \int_{p_\alpha}^{p_\omega} v dp$$

Hier bezeichnet F eine konservative Kraft (z.B. die Hookesche Federkraft $F = kx$) und dx eine Wegänderung, \mathbf{p} den Druck in einem homogenen Medium und dV eine Volumenänderung, v eine Geschwindigkeit und dp eine Impulsänderung.

24. Das Standardmodell der Thermodynamik

Um den Einstieg in die Thermodynamik für Physikanfänger nicht allzu schwierig zu machen, wird die Thermodynamik meist anhand energetischer Modelle erläutert, die nur aus zwei Beiträgen zur Energie \mathcal{U} eines makroskopischen Systems bestehen, nämlich aus dem für die Dissipation unverzichtbaren THERMISCHEN ENERGIEBEITRAG und irgendeinem der mechanischen Energiebeiträge.

Da Prozesse des Energieaustauschs zwischen verschiedenen Energieformen bzw. zwischen verschiedenen Systemen ohne die als Mittler fungierende kinetische Energie nicht zustande kämen, mag es zunächst verwundern, warum nicht auch noch mindestens ein Beitrag zur kinetischen Energie berücksichtigt werden muss. Man kann jedoch zur weiteren Vereinfachung die Thermodynamik im QUASISTATISCHEN GRENZFALL betrachten, bei dem man annimmt, dass alle Prozesse so langsam ablaufen, dass vorübergehend auftretende Beiträge zur kinetischen Energie im Vergleich zu den Umsätzen anderer Energieformen vernachlässigbar klein bleiben.

Im STANDARDMODELLS DER THERMODYNAMIK wählt man als mechanischen Energiebeitrag die volumenelastische Energie (s. Abschnitt 18.5). Da das für Beschreibung des Materialverhaltens von Fluiden ein wichtiger Energiebeitrag ist, ist dieses energetische Standardmodell insbesondere von Bedeutung für die Materialphysik und die Chemie. Hat man die Thermodynamik anhand dieses stark vereinfachten Standardmodells verstanden, so sind die dabei gewonnenen Erkenntnisse leicht auf komplexere Fälle übertragbar. Das Standardmodell geht von den beiden extensiven Zustandsvariablen Entropie S und Volumen V aus, bzw. von der Fundamentalform

$$\mathcal{U} = \mathcal{U}(S, V). \quad (24.1)$$

mit dem totalen Differential

$$d\mathcal{U} = \left(\frac{\partial \mathcal{U}}{\partial S} \right)_V dS + \left(\frac{\partial \mathcal{U}}{\partial V} \right)_S dV = \delta Q + \delta W \quad (24.2)$$

mit den partiellen Differentialen

$$\delta Q = \left(\frac{\partial \mathcal{U}}{\partial S} \right)_V dS$$

und

$$\delta W = \left(\frac{\partial \mathcal{U}}{\partial V} \right)_S dV. \quad (24.3)$$

Die zu den extensiven Zustandsvariablen konjugierten Zustandsvariablen sind die TEMPERATUR und der DRUCK:

$$T = T(S, V) = \left(\frac{\partial \mathcal{U}}{\partial S} \right)_V \quad (24.4)$$

$$\mathfrak{p} = \mathfrak{p}(S, V) = - \left(\frac{\partial \mathcal{U}}{\partial V} \right)_S. \quad (24.5)$$

Definitionsgemäß sind diese beiden intensiven Zustandsvariablen positive Größen.

Wenn Materie stabil sein soll, so muss die innere Energie bei Abnahme des Volumens anwachsen. Andernfalls würde Materie unter dissipativer Abgabe von Energie an die Umgebung ein immer kleineres Volumen einnehmen und schließlich „verschwinden“. Daher ist die partielle Ableitung $\left(\frac{\partial \mathcal{U}}{\partial V} \right)_S$ negativ. Das ist der Grund für das in Gl. 24.5 explizit auftretende negative Vorzeichen. Mit diesen intensiven Zustandsvariablen geschrieben, lautet die Fundamentalform der inneren Energie im Standardmodell:

$$d\mathcal{U} = TdS - pdV = \delta Q + \delta Q \quad (24.6)$$

mit

$$\begin{aligned} \delta Q &= TdS \\ \delta W &= -pdV \end{aligned}$$

Zustandsgleichungen Gleichungen, die wie 24.5 und 24.4 Zusammenhänge zwischen intensiven und extensiven Zustandsvariablen herstellen, heißen ZUSTANDSGLEICHUNGEN. Zustandsgleichungen werden meist dazu herangezogen, einen Satz von f unabhängigen Zustandskoordinaten bzw. Zustandsvariablen auf einen anderen unabhängigen Satz von f Zustandskoordinaten transformieren.

Im Standardmodell stellen sie Zusammenhänge zwischen jeweils drei Zustandsvariablen her. In Gl. 24.4 und Gl. 24.5 wird beispielsweise eine (intensive) Zustandskoordinate als Funktion von jeweils zwei unabhängigen (extensiven) Zustandskoordinaten ausgedrückt. In der Regel kann man Gl. 24.4 nach der Entropie auflösen und Gl. 24.5 nach dem Volumen. Damit erhält man die Zustandsgleichungen in der Form

$$S = S(T, V) \quad (24.7)$$

$$V = V(S, \mathfrak{p}) \quad (24.8)$$

Setzt man Gl. 24.7 in Gl. 24.5 ein, so kann man die Entropie gegen die Temperatur tauschen und erhält eine Zustandsgleichung in der Form

$$T = T(V, \mathbf{p}), \quad (24.9)$$

welche einen Zusammenhang zwischen Temperatur, Volumen und Druck beschreibt. Bekannte Beispiele für Zustandsgleichungen dieser Form sind die Van-der-Waals Zustandsgleichung (Gl. 30.2) und die Zustandsgleichung des idealen Gases.

Man kann aber genauso gut das Volumen gegen den Druck tauschen. Indem man nämlich Gl. 24.8 in die Fundamentalform der inneren Energie einsetzt, erhält man die innere Energie in der Form

$$\mathcal{U} = \mathcal{U}(S, V(S, \mathbf{p})) = \mathcal{U}(S, \mathbf{p}).$$

Sie ist dann als Funktion der unabhängigen Zustandsvariablen S und \mathbf{p} gegeben. Solche Koordinatentransformationen ändern selbstverständlich nichts an der Zahl der Freiheitsgrade bzw. der Anzahl der für eine Systembeschreibung benötigten unabhängigen Zustandsvariablen.

25. Thermisches Gleichgewicht

In einem abgeschlossenes System laufen so lange Prozesse ab, bis das thermodynamische Gleichgewicht erreicht ist, d.h. bis die Entropie unter den gegebenen Bedingungen maximal ist. Ein bereits besprochener möglicher Weg zum thermischen Gleichgewicht besteht darin, dass im System vorhandene kinetische Energie dissipativ in thermische Energie umgewandelt wird. So lange keine Einschränkungen vorhanden sind, kann die kinetische Energie aus im System vorhandener Energie requiriert werden, so dass letztends über den Zwischenschritt der kinetischen Energie deren potentielle Energie irreversibel in thermische Energie umgewandelt wird, was gleichbedeutend ist, dass die Entropie wächst, ihr aber im thermodynamischen Modell verboten ist, abzunehmen. Das beschreibt die bereits in der Mechanik thematisierte Erfahrungstatsache der Dissipation mechanischer Energie.

Hier soll nun ein zweiter Weg zum thermischen Gleichgewicht vorgestellt werden. Wie wir gleich sehen werden, kann er auftreten, wenn das System aus Teilsystemen besteht, welche unterschiedliche Temperaturen aufweisen und wenn zwischen ihnen ein Energieaustausch in Form von Wärme möglich ist. Das Besondere dabei ist, dass mechanische Energiebeiträge bei diesem Prozess überhaupt keine Rolle spielen. Die einfachst mögliche Modellsituation, die man hier studieren kann, besteht aus einem abgeschlossenes Gesamtsystem G , das nur aus zwei homogenen Teilsystemen A und B mit den Temperaturen T_A bzw. T_B besteht. Zwischen ihnen sei nur ein Energieaustausch in Form von Wärme möglich. Von A werde eine infinitesimale Wärme δQ_A abgegeben (δQ_A negativ) oder aufgenommen (δQ_A positiv) und entsprechendes gelte für das Teilsystem B . Im Rahmen des thermodynamischen Modells bleibt die innere Energie \mathcal{U} eines abgeschlossenen Systems G aber konstant und daher gilt

$$d\mathcal{U} = \delta Q_A + \delta Q_B = T_A dS_A + T_B dS_B = 0 \quad (25.1)$$

Für die hier interessierenden irreversiblen Prozesse muss die Entropie S von G zudem anwachsen, d.h. die Änderung der extensiven Größe S muss positiv sein:

$$dS = dS_A + dS_B > 0.$$

Aus beidem folgt

$$\left(1 - \frac{T_A}{T_B}\right) dS_A > 0.$$

Ist $T_A > T_B$, so kann A nur Wärme abgeben und für $T_A < T_B$ ist nur der umgekehrte Vorgang möglich. Das Modell macht folglich die klare Ansage, dass ein System A in Wärmekontakt mit einem System B sich so lange unter Wärmeabgabe irreversibel fortentwickeln kann, wie seine Temperatur höher ist als die von B. Sobald sich die Temperaturen jedoch angeglichen haben, ist kein weiterer irreversibler Austausch von Wärme mehr möglich. Ist $T_A = T_B$, so folgt aus Gl. 25.1, dass $dS = dS_A + dS_B = 0$. Die Entropie ist somit maximal und die beiden Systeme sind folglich im THERMISCHEN GLEICHGEWICHT. Dieser Zustand ist zeitlich stationär, d.h. er ändert sich nicht mehr.

Was in der Natur prinzipiell möglich ist, geschieht üblicherweise auch. Es ist daher nur eine Frage der Zeit, bis sich das thermische Gleichgewicht bei Wärmekontakt eingestellt hat. Die Modellanalyse besagt schlussendlich, dass sich für zwei Systeme, die in Wärmekontakt sind, irgendwann ein Temperaturgleichgewicht einstellen wird. Zwar gibt uns die Modellaussage keine Auskunft darüber, wie schnell oder wie langsam der Prozess abläuft, aber sie reicht aus, um als theoretische Grundlage der Temperaturmessung zu dienen. Zusammengefasst lauten die wichtigsten Aussagen, die aus einer Analyse des thermodynamischen Modells hervorgehen:

Können zwischen Systemen nur Prozesse stattfinden, bei denen Wärme ausgetauscht werden kann, dann

- ist die Abgabe von Wärme bzw. Entropie von einem System mit niedrigerer Temperatur an ein System mit höherer Temperatur nicht möglich.
- streben die Systeme irreversibel einem thermischen Gleichgewichtszustand zu, bei dem ihre Temperaturen gleich sind.

Wenn Wärme in einem rein dissipativen (irreversiblen) Prozess von einem Körper zu einem anderen fließt, bezeichnet man den Körper, von dem die Wärme abfließt, als den HEISSEN Körper und den anderen als den KALTEN Körper.

Ein Gesamtsystems, das aus homogenen Teilsystemen besteht, die unterschiedliche Temperaturen haben, hat keine Temperatur. Man kann einfach keine angeben. Sie ist nicht definiert. Für die Entropie ist die Sachlage anders. Da sie eine extensive Größe ist, ist die Entropie des Gesamtsystems gleich der Summe der Entropien der Teilsysteme. Auch wenn noch kein thermisches Gleichgewicht herrscht, ist die Entropie eines aus zwei Teilen bestehenden Gesamtsystems per definitionem

$$S_G = S_A + S_B.$$

Im Nichtgleichgewicht sind die Entropien von G stets kleiner als die Entropie im Gleichgewichtszustand. Die hier mit einem Querstrich gekennzeichnete Gleichgewichtsentropie

$$\bar{S}_G = \max(S_G)$$

ist die maximale Entropie, welche durch reinen Wärmeaustausch erreicht werden kann. Dem Gesamtsystem G lässt sich erst im thermischen Gleichgewicht eine Temperatur zuordnen: Da nun überall die gleiche Temperatur herrscht, d.h. hier $T_A = T_B$, und damit auch das Gesamtsystem homogen geworden ist, kann man die Temperatur von G gleich der GLEICHGEWICHTSTEMPERATUR setzen, d.h.

$$T_G := T_A = T_B.$$

26. Thermometrie

Messen beruht auf dem Vergleich mit einem Maßstab, auf dem der Messwert abgelesen werden kann. Maßstäbe für Temperaturen bezeichnet man als THERMOMETER. Wenn ein Thermometer mit einem Objekt so in Kontakt gebracht wird, dass zwischen beiden Wärme ausgetauscht werden kann und wenn das Thermometer nach einer Weile schließlich einen stationären Wert anzeigt, dann sind die von ihm angezeigte Temperatur und die Temperatur des Objekts gleich. Die Geschichte der Temperaturmessung bzw. der THERMOMETRIE ist lang. Die älteste uns bekannte Beschreibung eines Thermometers mit einer Temperaturskala stammt von Galenos von Pergamon (2. Jh.). Seine bedeutenden medizinischen Werke wurden im christlichen Europa vernichtet, im arabischen Raum wiederentdeckt und kommen erst ein Jahrtausend später im Abendland der Renaissance zur Entfaltung. Die Thermometrie der Neuzeit beginnt mit der Wiederentdeckung des Thermometers durch Galileo Galilei.

26.1. Praktische Thermometrie.

Grundlage des Temperaturvergleichs der praktischen Thermometrie ist das sich in thermischem Kontakt stets nach einer gewissen Zeit einstellende thermische Gleichgewicht zwischen Messobjekt und Thermometer. In der Praxis verwendet man leicht messbare temperaturabhängige Größen als Temperaturindikatoren, beispielsweise das Volumen eines Fluids wie Quecksilber oder Alkohol, den elektrischen Widerstand von Platin-Messwiderständen oder die Thermospannung von Thermoelementen. Auf der Basis dieser Phänomene kann man THERMOSKOPE realisieren, d.h. Anzeigergeräte für die Temperatur. Mit einem Thermoskop kann man erst einmal nur qualitativ feststellen, ob die Temperatur gleich geblieben, zu- oder abgenommen hat. Man kann jedoch die Änderung der als Temperaturindikator verwendeten Messgröße selbst heranziehen, um eine Temperaturänderung DEM GRAD NACH zu charakterisieren, d.h. um zu charakterisieren, ob sich eine



Abbildung 49: Réaumur-Skala (links) und Celsius-Skala (rechts) an einem Ausdehnungsthermometer in Wien. Die beiden Temperaturskalen haben unterschiedliche Einteilung aber den gleichen durch einen Temperaturfixpunkt definierten Nullpunkt.

Temperatur mehr oder weniger stark geändert hat. Besonders günstig ist es, wenn man für solche Grad-Thermometer einen Temperaturindikator heranzieht, der eine monotone Funktion der Temperatur ist.

Die ersten Grad-Thermometer beruhten auf der Volumenänderung von Fluiden mit der Temperatur. Löst man Gl. 24.9 beispielsweise nach dem Volumen auf, so ist

$$V = V(T, p). \quad (26.1)$$

Sorgt man dafür, dass der Druck auf ein Fluid konstant bleibt, ist das Volumen folglich allein eine Funktion der Temperatur, d.h. im Standardmodell ist $V = V(T)$. Mit zunehmender Temperatur dehnt sich das Volumen im Regelfall aus. Als verlässlicher Temperaturindikator für den Alltag hat sich beispielsweise die Volumenausdehnung des Quecksilbers bei Atmosphärendruck bewährt. Für den alltags- und küchenrelevanten Temperaturbereich ist der Zusammenhang zwischen der relativen Volumenausdehnung des Quecksilbers und der Temperaturänderung in exzellenter Näherung linear und kann durch

$$\frac{\Delta V}{V} \approx \gamma_{th} \Delta T \quad (26.2)$$

mit einem in guter Näherung konstanten Faktor γ_{th} beschrieben werden. Gl. 26.2 ist auch für viele andere Materialien über gewisse Temperaturbereiche eine taugliche Näherung, wobei der thermische Volumenausdehnungskoeffizient γ_{th} eine materialspezifische Größe darstellt.

Historisch wurden verschiedene solche Gradmaß-Thermometer vorgeschlagen und praktisch meist als Ausdehnungsthermometer realisiert. Im Wien der Jahrhundertwende waren beispielsweise Ausdehnungsthermometer in Gebrauch auf denen der Grad der Tem-

peraturänderung auf der Celsius-Skala und der Réaumur-Skala abgelesen werden konnten (Abb. 49).

Thermoskope werden aber erst dadurch zu THERMOMETERN im eigentlichen Sinne, dass sie durch Kalibration in den Eichbüros an die Temperatureinheit Kelvin angeschlossen werden. Erst dann charakterisieren sie eine Temperaturänderung nicht nur dem Grade nach, sondern zeigen die Temperatur T selbst an.

26.2. Definition der Temperatureinheit

Betrachtet wird ein System C, das einen Kreisprozess ausführt (C steht hier für „cyclic“ bzw. Carnot). Zunächst soll es im Temperaturgleichgewicht mit einem System A bei einer Temperatur T_A eine Wärme Q_A aufnehmen, dann seine Temperatur durch Zustandsänderungen auf eine andere Temperatur T_B bringen und im Temperaturgleichgewicht mit einem System B bei der Temperatur T_B eine Wärme Q_B abgeben. Zum Schluss soll der Kreisprozess durch allfällige weitere Zustandsänderungen vollendet werden. Volumen V und Entropie S haben am Ende des Prozesses also die gleichen Werte wie am Anfang. Dass das Volumen am Ende gleich ist, kann durch Abmessen kontrolliert werden, und bei konstantem Volumen ist die Entropie genau dann gleich, wenn die Temperatur $T = T(S, V)$ gleich ist. Um das zu kontrollieren, genügt ein (unkalibriertes) Thermoskop. Durch Aufnahme von Wärme erhöht sich die Entropie um $\Delta S_A = Q_A/T_A$.²⁴ Darüber hinaus kann sich die Entropie durch irreversible Vorgänge um ΔS_i erhöhen. In einem Kreisprozess müssen beide Entropieerhöhungen durch die Entropieerniedrigung ΔS_B rückgängig gemacht werden, welche durch Abgabe von Wärme $Q_B = T_B \Delta S_B$ bei der Temperatur T_B auftritt. Die Entropiebilanz für einen Kreisprozess lautet daher:

$$\Delta S_A + \Delta S_B + \Delta S_i = 0.$$

Zumindest für die bei konstanter Temperatur ablaufende Entropieänderungen treten wegen $T = T(S, V) = \text{const.}$ notwendigerweise auch Volumenänderungen auf. Infolgedessen kann C auch Arbeit leisten, d.h. eine Energiemenge W in Form von Arbeit nach außen abgeben (W negativ). Da die innere Energie am Ende eines Kreisprozesses die gleiche ist wie am Anfang, ist die Energiebilanz null, d.h.

$$Q_A + Q_B + W = T_A \Delta S_A + T_B \Delta S_B + W = 0$$

Aus den beiden Gleichungen folgt

$$(T_A - T_B) \Delta S_A - T_B \Delta S_i + W = 0.$$

²⁴Zur Klarstellung: Die Entropie „fließt“ nicht von einem System zu einem anderen. Was von einem System zum anderen fließt, ist die Wärme. Wenn aber Wärme von einem System A zu einem System B fließt, hat das zur Folge, dass sich die Zustandsgröße Entropie in A verringert und in B erhöht. Deshalb werden wir in Kap. 36 konsequenterweise auch von einem Wärmestrom und nicht von einem Entropiestrom sprechen.

Arbeit kann also nur abgegeben werden, so lange $(T_A - T_B)\Delta S_A - T_B\Delta S_i$ positiv ist, und daher ist das erst einmal nur möglich, wenn $T_A > T_B$. Ferner kann um so mehr Arbeit abgegeben werden, desto kleiner ΔS_i ist. Die durch den Kreisprozess geleistete Arbeit ist dann maximal, wenn dabei gar keine Entropie irreversibel produziert wird, wenn also der Kreisprozess im reversiblen Grenzfall ausgeführt wird. Solche ideale reversible Kreisprozesse werden als CARNOTSCHES KREISPROZESSE bezeichnet. Das bemerkenswerte an Carnotschen Kreisprozessen ist, dass das Verhältnis

$$\frac{W}{Q_A} = \frac{T_B}{T_A} - 1 \quad (26.3)$$

der abgegebenen Arbeit zur zugeführten Wärme nur vom Verhältnis der Temperaturen abhängt, bei denen die Wärme aufgenommen bzw. abgegeben wurde.

Wenn das System A eine Wärmemenge Q_A abgibt, so ändert sich sein Zustand. Es sei nun angenommen, dass sein Volumen konstant gehalten wird. Dann kommt es nur zu einer Entropieabnahme. Sie würde eine Temperaturänderung bewirken. Damit das nicht geschieht, muss die Entropieabnahme des System A fortlaufend durch einen Prozess kompensiert werden, der eine gleich große Entropiezunahme bewirkt. Das kann man dadurch erreichen, dass man dem System A von außen Arbeit zuführt und sie darin dissipieren lässt. Das kann man beispielsweise dadurch realisieren, dass man als System A einen geschlossener Behälter mit Wasser nimmt, in dem ein von außen angetriebener Rührer läuft und der in Wärmekontakt mit dem System C steht. Wenn die Temperatur von A konstant gehalten wird, ist die von außen zugeführte Arbeit W_A (positiv) gleich der abgegebenen Wärme, also $W_A = Q_A$. Infolgedessen gilt nun

$$T_A = \frac{W_A}{W_A - W} T_B.$$

Wenn man also die Temperatur T_B kennt, dann kann man die Temperatur T_A durch rein mechanische Messgrößen bestimmen, nämlich durch Messung der Arbeit W_A und W . Das ist Grundidee zur Definition der thermodynamischen Temperaturskala in der Einheit Kelvin, denn man muss nur noch ein physikalisches Phänomen finden, welches seine Temperatur T_B bei Zufuhr von Wärme nicht ändert (was man wiederum durch ein unkalibriertes Thermoskop kontrollieren kann). Durch dieses Phänomen erhält die Kelvin-Skala ihren Temperaturfixpunkt.

Es gibt mehrere Phänomene, welche dafür geeignet sind. Man hat sich dafür entschieden, den Temperaturfixpunkt der Kelvin-Skala durch ein Bad zu realisieren, bei dem ein Gemisch von drei Phasen eines Stoffes, einer gasförmigen, einer flüssigen und einer festen Phase, im thermischen Gleichgewicht stehen. Auch mit einem unkalibrierten Thermoskop lässt sich empirisch nachweisen, dass sich die Gleichgewichtstemperatur nicht ändert, wenn man einem Gemisch aus drei koexistierenden Phasen Wärme zuführt. Durch die Wärmezufuhr ändern sich nämlich lediglich die Mengenanteile der beteiligten Phasen. Die charakteristische Gleichgewichtstemperatur T_{III} für die Koexistenz dreier Phasen wird als TRIPELPUNKT bezeichnet. Da der Tripelpunkt materialspezifisch ist, muss man ihn für die Definition der Kelvin-Skala ein genau spezifiziertes Referenzmedium festlegen.

Dies ist das Wiener Standard-Ozean-Wassers, dessen Tripelpunkt mit $T_{III} = 273.16 \text{ K}$ definiert wurde. Der ungewöhnliche Zahlenwert hat historische Gründe.

Wenn man also einen Carnotschen Kreisprozess mit einem thermischen Bezugssystem B realisiert, dessen Bezugstemperatur durch den Tripelpunkt $T_B = T_{III} = 273.16 \text{ K}$ des Wiener Referenzmediums festgelegt und damit unabhängig von der zugeführten Wärmemenge Q_B fixiert ist (genau das meint man mit einem Temperaturfixpunkt), dann ergibt sich die Temperatur T_A aus der Messung von W_A und W und somit letztlich aus der Messung der Änderung mechanischer (und damit beobachtbarer) Zustandsvariabler. Damit ist das SI-Maß **KELVIN**

$$[T] = 1 \text{ Kelvin} = 1 \text{ K}$$

zur Messung der Temperatur T prinzipiell realisiert.

Das bis hier geschilderte Messverfahren ist die exakte theoretische Prozedur für die Messung der Temperatur in der Einheit Kelvin. Es lässt sich jedoch nur mit der Expertise und Messtechnik nationaler Eichbüros realisieren und ist daher nicht alltagstauglich. Man kann jedoch dem Wert des Temperaturindikators eines Thermoskops, das mit dem System A in thermischem Kontakt ist den Wert T_A in Kelvin zuordnen und so aus jedem Thermoskop ein in Kelvin geeichtes Thermometer machen. Beispielsweise kann ein Messwiderstand als Thermoskop herangezogen werden, wobei sein Widerstandswert als Temperaturindikator dient. Wenn man die Widerstandswerte mit dem geschilderten Verfahren auf Temperaturen kalibriert, wird dieses Thermoskop zu einem Thermometer.

Bei der obigen vereinfachten Schilderung, wie die Einheit der Temperatur definiert wird, mag der Eindruck entstehen, dass nur Temperaturen T_A ein Messwert zugeordnet werden kann, die größer als die Temperatur T_B sind. Gl. 26.3 zeigt jedoch, dass im Fall $T_A < T_B$ die einzige Änderung darin besteht, dass das System C beim Carnot-Prozess Arbeit W aufnehmen muss anstatt abzugeben, weil W für $T_A < T_B$ positiv wird.

Celsiuskala. Neben dieser thermodynamischen Temperaturskala ist im Alltag noch eine Temperaturskala in Gebrauch, deren Nullpunkt der Temperatur $T = 273.15 \text{ K}$ entspricht. Für Temperaturangaben in dieser ebenfalls SI-konformen Celsius-Skala wird als Einheit die Bezeichnung Grad CELSIUS ($^{\circ}\text{C}$) in diesem Buch das Größensymbol ϑ_C verwendet. Man berechnet den Grad-Wert der Celsiuskala also aus der Temperatur T gemäß

$$\vartheta_C = (T - 273.15 \text{ K})^{\circ}\text{C}/\text{K}.$$

Der Unterschied zwischen der Temperatur, bei der Wasser bei einem vereinbarten Normaldruck p_0 gefriert bzw. siedet, beträgt beispielsweise $\Delta T \approx 100.00 \text{ K}$ bzw. $\Delta\vartheta_C = 100.00^{\circ}\text{C}$, d.h. die Zahlenwerte für die Temperatur und die entsprechende Angabe in Celsius-Graden sind gleich. Beide Angaben sind äquivalent. Der **NORMALDRUCK** beträgt $p_0 \approx 1.0 \text{ bar} = 1.0 \cdot 10^5 \text{ Pa}$.

27. Revision von Begriffen aus der Mechanik

27.1. Thermodynamisches Druckgleichgewicht

Das Volumen eines physikalischen Systems unterliegt keiner prinzipiellen Einschränkung.²⁵ Gase können in der Tat prinzipiell ihr Volumen beliebig vergrößern (s. Kap. 28.3).

Es sei nun ein aus zwei Teilsystemen A und B bestehendes abgeschlossenes System G im thermischen Gleichgewicht betrachtet. Dann sind die Temperaturen alle gleich. Dissipative Prozesse aufgrund von Wärmeaustausch sind damit ausgeschlossen. Welchen Druck hat dieses heterogene Gesamtsystem G? Das ist zunächst einmal undefiniert, denn der Druck ist eine intensive Zustandsvariable.

Wenn man das aus den beiden Teilvolumina V_A und V_B hervorgehende Gesamtvolumen $V = V_A + V_B$ der Bedingung unterwirft, dass es konstant sein soll, dann hat das die Wirkung eines Erhaltungssatzes und auch ähnliche Folgen, die nun herausgearbeitet werden sollen. Für die Volumenänderungen gilt $dV_A = -dV_B$ und somit gilt für den Energieaustausch innerhalb des abgeschlossenen Systems

$$T_G dS_G + (p_B - p_A) dV_A = 0$$

So lange $p_A > p_B$ kann durch eine Vergrößerung des Volumens von System A und eine entsprechende Verringerung des Volumens von B die elastische Energie $(p_B - p_A) dV_A$ irreversibel dissipiert werden. Das ist ein zulässiger Prozess, bei dem die Entropie S_G des Gesamtsystems anwächst. Der Vorgang hört auf, sobald es mit $p_A = p_B$ zum Druckausgleich kommt. Die im System vorhandene potentielle Energie hat dann ihr Minimum erreicht, denn es gibt dann keine Prozesse mehr, durch die sie weiter verringert werden kann (vgl. mit Kap. 17). Thermodynamische Systeme streben einem Maximum der Entropie zu. Das Maximum der Entropie wird für einen Zustand erreicht, bei dem alle Teilsysteme, aus denen sie sich zusammensetzen, die gleichen Werte für die intensiven Zustandsvariablen annehmen, die sich ändern dürfen. Der sich damit einstellende stationäre Zustand ist hier das THERMODYNAMISCHE GLEICHGEWICHT DES DRUCKS. Im Standardmodell ist das thermodynamische Gleichgewichtszustand daher erreicht, wenn überall sowohl gleiche Temperatur als auch der derselbe Gleichgewichtsdruck

$$p_G = p_A = p_B$$

herrschen. Im ursprünglichen Nichtgleichgewichtssystem ist der Druck des Gesamtsystems undefiniert, während das Volumen einfach die Summe der Teilvolumina ist. Erst im thermodynamischen Gleichgewicht hat der Druck des Gesamtsystems einen Wert, nämlich p_G .

27.2. Enthalpie im Standardmodell

Die Ergänzung der inneren Energie \mathcal{U} um die thermische Energie zieht eine entsprechende Ergänzung für die Enthalpie $\mathcal{H} = \mathcal{U} + pV$ (s. Kap. 18.6) nach sich. Führt man einem

²⁵Für die elektrische Ladung eines physikalischen Systems ist das beispielsweise anders (Kap. 35.3).

Objekt, das sich im Druckgleichgewicht mit seiner Umgebung befindet, von außen eine infinitesimale Energie in Form von Arbeit δW oder Wärme δQ zu, so ändert sich seine Enthalpie um

$$d\mathcal{H} = d\mathcal{U} + d(\mathbf{p}V) = \delta Q + \delta W, \quad (27.1)$$

weil sowohl seine eigene (innere) Energie infinitesimal um $d\mathcal{U}$ anwächst als auch seine Wechselwirkungsenergie mit der Umgebung, nämlich seine Verdrängungsenergie um $d(\mathbf{p}V) = \mathbf{p}dV + Vd\mathbf{p}$. Die Änderung der Enthalpie ist nur dann gleich jener der inneren Energie, wenn sich weder Druck noch Volumen ändern.

Bei der Aufstellung der Thermodynamik geht man zunächst von der inneren Energie aus, weil für sie ein Erhaltungssatz gilt. In der Enthalpie ist stets in gewissen Maße ein wenig Energie seiner Umgebung mit verpackt. Sie ist daher keine Erhaltungsgröße und auch keine sinnvolle Größe für isolierte Systeme, aber wichtig für jedes reale System das im Gleichgewicht mit einer Umgebung ist. Änderungen der inneren Energie sind reine „Innenpolitik“, d.h. eine rein interne Angelegenheit des Systems, und daher braucht man auch nichts über die Umgebung zu sagen, wenn man sie angibt. Die Enthalpie inkludiert die „Außenpolitik“ und daher müssen bei der Angabe von Enthalpien die Umgebungsbedingungen stets mitspezifiziert werden. In Tabellen werden die Enthalpien für die Bildung chemischer Stoffe meist unter STANDARDBEDINGUNGEN angegeben. In der Physik ist das meist $\vartheta_C = 0^\circ\text{C}$ und $\mathbf{p} \approx 1 \text{ bar}$, in der Chemie meist $\vartheta_C = 25^\circ\text{C}$ und $\mathbf{p} = 1 \text{ bar}$. Ausgehend von solchen Standardenthalpien kann man die Enthalpien für andere Umgebungsbedingungen ausrechnen, indem man die geänderte Verdrängungsenergie in Rechnung stellt. Das ist kein Problem, denn sowohl \mathbf{p} als auch V sind Zustandsvariable des Systems und damit ist auch ihr Produkt, die Verdrängungsenergie $\mathbf{p}V$, eine Zustandsgröße. Es ist daher egal auf welchem Wege der Endzustand bzw. der Gleichgewichtszustand erreicht wird. Wegen der Wegunabhängigkeit sind sowohl die Verdrängungsenergie als auch die Enthalpie zwar Potentiale im Sinne der Analysis (s. Kap. 19.2), aber – wie schon erwähnt – keine Erhaltungsgrößen des Systems: diese Potentiale können sich auch dadurch ändern, dass sich die Umgebung verändert, beispielsweise, wenn sich der Druck der Umgebung ändert (das steckt in dem Term $Vd\mathbf{p}$ für die Änderung der Verdrängungsenergie). Damit ist gerechtfertigt, die infinitesimale Änderung auf der rechten Seite von Gl. 27.1 als Differential zu schreiben:

$$d\mathcal{H} = d\mathcal{U} + \mathbf{p}dV + Vd\mathbf{p} = TdS + Vd\mathbf{p}. \quad (27.2)$$

Ist die Enthalpie als Funktion der Zustandsvariablen Entropie und Druck formuliert, also $\mathcal{H} = \mathcal{H}(S, \mathbf{p})$, so hat sie das vollständige Differential

$$d\mathcal{H} = TdS + Vd\mathbf{p} = \left(\frac{\partial\mathcal{H}}{\partial S}\right)_{\mathbf{p}} dS + \left(\frac{\partial\mathcal{H}}{\partial\mathbf{p}}\right)_S d\mathbf{p}. \quad (27.3)$$

und somit gilt

$$T(S, \mathbf{p}) = \left(\frac{\partial \mathcal{H}}{\partial S} \right)_{\mathbf{p}}$$

$$V(S, \mathbf{p}) = \left(\frac{\partial \mathcal{H}}{\partial \mathbf{p}} \right)_S$$

28. Thermische Materialeigenschaften

Materialeigenschaften müssen i.A. empirisch erforscht werden. Hier sollen als Beispiel thermische Ausdehnung γ_{th} (Kap. 28.2), die spezifischen Wärmekapazitäten c_p und c_v sowie der Joulesche Expansionskoeffizient (Kap. 28.3) im Rahmen des thermodynamischen Standardmodells diskutiert werden. Diese Größen sind für die Aufstellung der allgemeinen Zustandsgleichung wichtig (Kap. 30).

28.1. Nomenklatur thermodynamischer Prozesse

In diesem Abschnitt werden *offene* thermodynamische Systeme im Standardmodell betrachtet, die Wärme und/oder Arbeit mit der Umgebung austauschen. Jeden solchen Prozess kann man sich als Abfolge von infinitesimalen Prozessschritten vorstellen. Bei jedem wird eine infinitesimale Wärmemenge δQ und/oder eine Arbeitsmenge δW unter bestimmten *Prozessbedingungen* ausgetauscht, die man in vier Typen einteilen kann:

Prozesstyp	Bedingung	Prozesstyp	Bedingung
isentrop	$dS = 0$	isotherm	$dT = 0$
isochor	$dV = 0$	isobar	$d\mathbf{p} = 0$

Im Standardmodell ist ein isentroper Prozess zugleich ein adiabatischer Prozess, d.h. ein Prozess, bei dem kein Wärmeaustausch stattfindet, und ein isochorer Prozess zugleich ein diabatischer Prozess, d.h. ein Prozess, bei dem kein Arbeitsaustausch stattfindet.

28.2. Thermische Ausdehnung

Der isobare Koeffizient der thermischen Volumenausdehnung ist durch

$$\gamma_{th}(T, V) = \frac{1}{V} \left(\frac{\partial V}{\partial T} \right)_{\mathbf{p}} \quad (28.1)$$

definiert. Er ist i.A. eine Funktion der Temperatur und des Volumens. Normalerweise ist γ_{th} positiv, d.h. die Ausdehnung nimmt mit der Temperatur zu. Typische Beispiele für die Größenordnung der thermischen Ausdehnung sind in Tab. 7 aufgeführt. In einem Temperaturbereich um 4 °C herum tritt für Wasser jedoch eine Anomalie auf: Hier wird γ_{th} negativ.

Tabelle 7: Größenordnung des isobaren thermischen Ausdehnungskoeffizienten γ_{th} für feste, flüssige und gasförmige Körper unter NORMBEDINGUNGEN, d.h. bei atmosphärischem Luftdruck ($p \approx 1 \text{ bar} \approx 1 \cdot 10^5 \text{ Pa}$) und 0°C .

Aggregatzustand	Ausdehnungskoeffizient γ_{th} unter Normbedingungen
fest	10^{-5} K^{-1}
flüssig	$1 \times 10^{-3} \text{ K}^{-1}$
gasförmig	$3 \times 10^{-3} \text{ K}^{-1}$

Für eingeschränkte Temperaturintervalle ist die lineare Näherung

$$\Delta V \approx V \gamma_{th} \Delta T$$

mit einem konstanten Ausdehnungskoeffizienten γ_{th} eine ausreichende Näherung für den Zusammenhang zwischen Volumen- und Temperaturänderung, was z.B. für das Quecksilberthermometer genutzt wird (s. Gl. 26.2). Für Gase, findet man allerdings empirisch

$$\gamma_{th}(T, V) \approx (1 - V_{min}/V) \frac{1}{T} \quad (28.2)$$

mit einem materialspezifischen Parameter V_{min} . Für sehr große Volumina, d.h. $V \gg V_{min}$ nimmt der thermische Ausdehnungskoeffizient zunächst proportional zu $1/T$ ab. In diesem Grenzfall nimmt das relative Volumen $\Delta V/V$ proportional zur relativen Temperatur $\Delta T/T$ ab. Wenn das Volumen jedoch in die Größenordnung von V_{min} kommt, sinkt der Ausdehnungskoeffizient γ_{th} rascher. Für $V = V_{min}$ wird er null, d.h. V_{min} stellt das kleinstmögliche Volumen dar, auf das sich ein Fluid bei tiefen Temperaturen zusammenziehen kann.

28.3. Expansionskoeffizient und Wärmekapazitäten

Mit Hilfe der Zustandsgleichung 24.4 kann man die Entropie durch die Zustandsvariable Temperatur ersetzen, d.h. die Energie

$$\mathcal{U} = \mathcal{U}(S(T, V), V) = \mathcal{U}(T, V) \quad (28.3)$$

liegt dann als Funktion der unabhängigen Variablen Temperatur und Volumen vor. Bildet man das totale Differential

$$d\mathcal{U} = \left(\frac{\partial \mathcal{U}(T, V)}{\partial T} \right)_V dT + \left(\frac{\partial \mathcal{U}(T, V)}{\partial V} \right)_T dV, \quad (28.4)$$

so bezeichnet man den Koeffizienten

$$\gamma = \left(\frac{\partial \mathcal{U}(T, V)}{\partial V} \right)_T$$

als JOULESCHEN EXPANSIONSKOEFFIZIENTEN und den Koeffizienten

$$\Gamma_V = \left(\frac{\partial \mathcal{U}(T, V)}{\partial T} \right)_V \quad (28.5)$$

als ISOCHORE WÄRMEKAPAZITÄT.

Bei isobarer Prozessführung ändert sich das Volumen des Mediums. Im Gleichgewicht mit seiner Umgebung muss die Verdrängungsarbeit berücksichtigt werden. Daher geht man für die Definition der isobaren Wärmekapazität von der Enthalpie $\mathcal{H}(T, p)$ als Funktion von Temperatur und Druck aus. Bildet man das totale Differential von $\mathcal{H}(T, p)$,

$$d\mathcal{H} = \left(\frac{\partial \mathcal{H}}{\partial T} \right)_p dT + \left(\frac{\partial \mathcal{H}}{\partial p} \right)_T dp, \quad (28.6)$$

so stellt der Koeffizient

$$\Gamma_p = \left(\frac{\partial \mathcal{H}(T, V)}{\partial T} \right)_p \quad (28.7)$$

für konstanten Druck ($dp = 0$) die ISOBARE WÄRMEKAPAZITÄT dar.

Wärmekapazitäten sind extensive temperaturabhängige Größen. Für homogene Stoffe sind sie wie alle extensiven Größen proportional zur Stoffmenge (s. Kap. 20.2). Dividiert man sie daher durch die Masse m , hier in ihrer Funktion als vorläufiges Maß der Stoffmenge, erhält man stoffspezifische Größen, nämlich die SPEZIFISCHE ISOCHORE WÄRMEKAPAZITÄT

$$c_V = \frac{\Gamma_V}{m}. \quad (28.8)$$

bzw. die SPEZIFISCHE ISOBARE WÄRMEKAPAZITÄT

$$c_p = \frac{\Gamma_p}{m}. \quad (28.9)$$

Isochore Wärmekapazität. Wird einem System Energie bei konstantem Volumen zugeführt, so kann das im Rahmen des Standardmodells nur in Form von Wärme geschehen. Nach Gl. 28.4 resultiert die Zufuhr einer infinitesimalen Wärme δQ_V in einer Erhöhung der inneren Energie um

$$d\mathcal{U} = \Gamma_V dT = \delta Q_V.$$

Der Index V soll darauf hinweisen, dass die Wärme δQ_V bei festgehaltenem Volumen zugeführt wird. Daher kann man die isochore Wärmekapazität

$$\Gamma_V = \frac{\delta Q_V}{dT} \quad (28.10)$$

dadurch messen, dass man die Wärme sowie die resultierende Temperaturänderung bestimmt und das Verhältnis bildet. Ein Wärmespeicher hat dann eine hohe Wärmekapazität, wenn er sehr viel Wärme aufnehmen oder abgeben kann (δQ_V), ohne dass sich dadurch seine Temperatur merklich ändert (dT).

Joulescher Expansionskoeffizient. Um den Jouleschen Expansionskoeffizienten für Gase zu messen, bestimmt man die bei einer freien Expansion auftretende Temperaturänderung eines isolierten Systems. Dieses kann z.B. durch zwei Behälter realisiert werden, die durch ein Ventil miteinander verbunden sind und so isoliert sind, dass weder Arbeit noch Wärme mit der Umgebung austauscht werden kann. Ein Behälter enthält das Gas, während der andere evakuiert ist. Nun öffnet man das Ventil und lässt das Gas frei expandieren. Da das System isoliert ist, gilt

$$d\mathcal{U} = mc_V dT + \Upsilon dV = 0. \quad (28.11)$$

Bei bekannter Wärmekapazität c_V des Gases folgt der Joulesche Expansionskoeffizient durch Messung der Volumenänderung ΔV und der beim Versuch beobachteten Temperaturänderung ΔT :

$$\Upsilon \approx \frac{mc_V \Delta T}{\Delta V}. \quad (28.12)$$

Empirisch findet man in guter Näherung

$$\Upsilon \approx a\rho^2 \quad (28.13)$$

mit einer empirischen Materialkonstanten a und der Dichte $\rho = m/V$. Für stark verdünnte Gase ($\rho \rightarrow 0$) geht der Joulesche Expansionskoeffizient gegen null.

Isobare Wärmekapazität. Die isobare Wärmekapazität Γ_p wird dadurch gemessen, dass man eine Wärme δQ_p bei konstantem Druck ($d\mathbf{p} = 0$) zuführt und die resultierende Temperaturerhöhung dT misst. Mit Gl. 27.1 und Gl. 28.6 folgt

$$d\mathcal{H} = \Gamma_p dT = \delta Q_p, \quad (28.14)$$

und somit ergibt sich die isobare Wärmekapazität

$$\Gamma_p = \frac{\delta Q_p}{dT}, \quad (28.15)$$

einfach aus dem Verhältnis der beiden Messgrößen. Da die zugeführte Wärme einerseits die innere Energie erhöhen muss und andererseits die Verdrängungsarbeit gegen den konstanten Druck leisten muss, ist $\Gamma_p \geq \Gamma_V$.

Differenz der spezifische Wärmekapazitäten Für Gase findet man experimentell über einen weiten Temperaturbereich

$$c_p - c_V \approx \text{konstant} > 0 \quad (28.16)$$

und für kondensierte Materie in der Regel

$$c_p - c_V \approx 0.$$

29. Umwandlung von Arbeit in Wärme

Wir blicken hier noch einmal auf das Ausgangsproblem der Dissipation zurück, welches in Kap. 21 durch die Einführung des thermischen Energiebeitrags und durch den zweiten Hauptsatzes für die Entropie gelöst wurde. Wie stellt sich das Phänomen nun im Lichte der thermodynamischen Begriffe dar?

Das Feuerbohren ist eine schon seit Urzeiten bekannte Methode, um durch Reibung ein Feuer zu entzünden. Jeder, der mal einen stumpfen Bohrer mit einer Bohrmaschine betrieben hat, weiß, dass er glühend heiß werden kann. Graf Rumford, ein Spion, Physiker und Sozialreformer (Armenhäuser, Rumford-Suppe) im 18. Jahrhundert, war der erste, der aus der Wärmeabgabe beim Kanonenbohren zu dem Schluss kam, dass Wärme eine Form der Energie sein muss. Wenn man stumpfe Bohrer verwendete, dann wirkte nämlich ein Kanonenbohrer wie eine Maschine, die fortlaufend mechanische Energie aufnahm, also Arbeit, und Wärme abgab. Das physikalische Phänomen ist hier die Dissipation von kinetischer Energie, welche auftritt, wenn Teile eines physikalischen Systems eine Relativbewegung gegeneinander aufweisen.

Mitte des 19. Jahrhunderts untersuchte James Prescott Joule dieses Phänomen quantitativ: Er betrieb dazu einen Rührer in einem Behälter mit Wasser, der von außen mechanisch angetrieben wurde. Die Abnahme der Energie des äußeren Antriebssystems kann an den mechanischen Parametern des äußeren Systems abgelesen und somit gemessen werden. Die kinetische Energie, die dem Wasser durch Rühren zugeführt wird, wird erfahrungsgemäß dissipiert und erhöht somit die Entropie des Wasserbehälters und somit dessen thermische Energie. Wenn Arbeit zugeführt und dabei zugleich vollständig dissipiert wird, dann ist dieser Vorgang äquivalent zu einer Zufuhr von Wärme. Die zugeführte und im Behälter dissipierte Arbeit kann der Wasserbehälter u.a. als Wärme abgeben. Wenn genauso viel Wärme abgeführt wird, wie an Arbeit zugeführt und dissipiert wurde, dann verbleibt das System (der Wasserbehälter) in einem stationären Zustand: Die Verringerung der Entropie durch Wärmeabgabe wird ausgeglichen, indem genauso viel Entropie durch Dissipation „nachproduziert“ wird.

Wenn man den Wasserbehälter thermisch isoliert, erhöht sich mit der Entropie auch die Temperatur, und zwar genauso, als ob ihm Wärme zugeführt würde. Im Jouleschen Experiment gibt der Rührer beispielsweise eine infinitesimale Arbeit δW in Form von kinetischer Energie an das Wasser ab. Diese wird dissipiert, generiert Entropie und erhöht damit auch die Temperatur. Wenn man den flüchtigen Zwischenschritt des Umsatzes in kinetische Energie und seine anschließende Dissipation auslässt, wird also Arbeit in thermische Energie umgesetzt. Wenn der Wasserbehälter offen ist und der Prozess somit bei konstantem Druck abläuft, gilt:

$$\delta W \xrightarrow{\text{Dissipation}} \delta Q = TdS = d\mathcal{H} = \Gamma_p dT .$$

Der Pfeil stellt hier die Transformation der Arbeit durch den irreversiblen Prozess der Dissipation dar, so dass das Resultat einer gleich großen Zufuhr von Wärme äquivalent wäre ($\delta Q = \delta W$). Joule fand heraus, dass sich die Temperatur einer Stoffmenge von 1 kg

Wasser um $\Delta T = 1\text{ K}$ erhöht, wenn man eine Rührarbeit von $\Delta W = 4.2\text{ kJ}$ verrichtet. Daraus folgt, dass 1 kg Wasser eine Wärmekapazität von $\Gamma_p = 4.2\text{ kJ/K}$ hat bzw. eine spezifische Wärmekapazität $c_p = 4.2\text{ kJ/kgK}$.

Der Titel dieses Kapitels „Umwandlung von Arbeit in Wärme“ stellt den Sachverhalt nicht ganz korrekt dar. Eigentlich ist es so, dass die zugeführte Arbeit genauso wirkt wie eine Zufuhr von Wärme, wenn sie so zugeführt wird, dass sie vollständig dissipiert wird. Das kann beispielsweise geschehen, wenn sie als kinetische Energie einer Relativbewegung von Teilen des Systems auftritt, wie bei dem betrachteten Beispiels des Rührvorgangs. Es kann aber auch dadurch geschehen, dass elektrische Energie in einem Tauchsieder dissipiert wird und dieser die Energie in Form von Wärme an das Wasser überträgt, in den er eintaucht.

30. Allgemeine Zustandsgleichung

Für konstanten Druck findet man mit Gl. 27.2 und 28.4

$$d\mathcal{H} = \Gamma_p dT = d\mathcal{U} + p dV = \Gamma_V dT + \Upsilon dV + p dV.$$

und aus Gl. 28.1 ergibt sich ferner, dass $dV = \gamma_{th} V dT$ und somit

$$\Gamma_p - \Gamma_V = (p + \Upsilon) \gamma_{th} V$$

bzw.

$$(p + \Upsilon) \frac{1}{\rho} = (c_p - c_V) \frac{1}{\gamma_{th}}, \quad (30.1)$$

wobei hier $\rho = m/V$ die Dichte ist. Alle auftretenden Materialkonstanten, d.h. c_p , c_V , Υ , γ_{th} und ρ , sind i.A. Funktionen der Temperatur T und des Volumens V , so dass Gl. 30.1 den Zusammenhang der drei Zustandsvariablen p , T und V beschreibt und somit eine ALLGEMEINE ZUSTANDSGLEICHUNG darstellt. Die Zustandsgleichungen für spezielle Stoffe sind Spezialfälle davon. Es sollen nun zwei wichtige Beispiele vorgestellt werden.

30.1. Van-der-Waals Zustandsgleichung für Fluide

Setzt man Gl. 28.2 und Gl. 28.13 in die allgemeine Zustandsgleichung (Gl. 30.1) ein, so erhält man die empirische Zustandsgleichung

$$(p + a\rho^2)(V - V_{min}) = m(c_p - c_V)T, \quad (30.2)$$

die als VAN-DER-WAALSISCHE ZUSTANDSGLEICHUNG bezeichnet wird. Sie ist sowohl für Gase wie Flüssigkeiten eine gute Modellgleichung, d.h. ganz allgemein eine gute Zustandsgleichung für Fluide. Für ausreichend große Volumina kann man V_{min} gegen V vernachlässigen. Das ist für Gase eine gute Näherung. Durch partielles Differenzieren

von Gl. 30.2 nach dem Druck erhält man (nach einiger Rechnerei) für die isotherme Kompressibilität von Gasen

$$\kappa_T(\mathbf{p}, \rho) = -\frac{1}{V} \left(\frac{\partial V}{\partial \mathbf{p}} \right)_T \approx \frac{1}{\mathbf{p} - a\rho^2}. \quad (30.3)$$

30.2. Die Zustandsgleichung verdünnter Gase

Im Grenzfall verdünnter Gase, d.h. für Gase mit verschwindender Dichte ($\rho \rightarrow 0$) und verschwindendem Eigenvolumen ($V_{min} \rightarrow 0$), ergeben sich aus den Gln. 28.13, 28.2 und Gl. 30.3 die Materialkonstanten

$$\begin{aligned} \Upsilon &= \left(\frac{\partial \mathcal{U}(T, V)}{\partial V} \right)_T = 0 \\ \gamma_{th} &= +\frac{1}{V} \left(\frac{\partial V}{\partial T} \right)_\mathbf{p} = \frac{1}{T} \\ \kappa_T &= -\frac{1}{V} \left(\frac{\partial V}{\partial \mathbf{p}} \right)_T = \frac{1}{\mathbf{p}} \end{aligned} \quad (30.4)$$

Die Van-der-Waalsche Zustandsgleichung, Gl. 30.2, vereinfacht sich damit zur Zustandsgleichung

$$\mathbf{p}V = m(c_p - c_V)T \quad (30.5)$$

verdünnter Gase.

Für $\Upsilon = 0$ folgt für $\mathcal{U}(T, V)$ aus Gl. 28.4 ferner, dass

$$d\mathcal{U} = mc_V dT,$$

d.h. dass die innere Energie vom Volumen überhaupt nicht abhängt und somit allein eine Funktion $\mathcal{U}(T, V) = \mathcal{U}(T)$ der Temperatur ist! Wenn man des weiteren den empirischen Befund berücksichtigt, dass die spezifischen Wärmen c_V und c_p für verdünnte Gase temperaturunabhängig werden, ergibt sich für die innere Energie verdünnter Gase der einfache Zusammenhang

$$\mathcal{U}(T) = mc_V T. \quad (30.6)$$

mit Gl. 30.5 für die Enthalpie

$$\mathcal{H} = \mathcal{U} + \mathbf{p}V = mc_p T.$$

31. Adiabatische Prozesse

Bei adiabatischen Prozessen tauscht ein System nur Arbeit mit seiner Umgebung aus, während der Austausch von Wärme durch geeignete Maßnahmen unterbunden ist. Wegen $\delta Q = 0$ gilt

$$d\mathcal{U} = TdS - pdV = \Gamma_V dT + \Upsilon dV = \delta W. \quad (31.1)$$

Im reversiblen bzw. isentropen Grenzfall mit $dS = 0$ ist mit der Zufuhr von Arbeit eine Volumenänderung verbunden, welche eine Temperaturänderung zur Folge hat. Aus Gl. 31.1 und Gl. 30.1 ergibt sich hierfür:

$$dT = -\frac{1}{\Gamma_V}(\mathfrak{p} + \Upsilon)dV = -\frac{1}{\gamma_{th}}(c_p/c_V - 1)\frac{dV}{V}. \quad (31.2)$$

Für Gase kann es durch adiabatische Komprimierung des Volumens zu einer beträchtlichen Temperaturerhöhung kommen. Dieser Effekt ist die physikalische Grundlage des pneumatischen Feuerzeugs und der Zündung des Dieselmotors. Umgekehrt kann man durch Expansion die adiabatische Kühlung eines Gases erzielen. Da die Wärmeabgabe an eine auf anderer Temperatur befindliche Umgebung nicht perfekt ausgeschaltet werden kann, verwendet man dazu eine schnell ablaufende Prozessführung (rasche Komprimierung bzw. Expansion). Sie entspricht näherungsweise einer adiabatische Prozessführung. Setzt man beispielsweise in Gl. 31.2 die Eigenschaften eines verdünnten Gases ein, so erhält man mit $\gamma_{th} = 1/T$:

$$\frac{dT}{T} = (1 - c_p/c_V)\frac{dV}{V},$$

Durch Integration beider Gleichungsseiten ergibt sich $\ln T = (1 - c_p/c_V)\ln V + C$ mit einer Integrationskonstanten C und daraus die ADIABATENGLICHUNG

$$TV^{c_p/c_V - 1} = \text{konstant} \quad (31.3)$$

Adiabatengleichungen beschreiben die Abfolge der Zustände verdünnter Gase bei einem adiabatischen Prozess. Wenn man sie für andere Zustandsvariable aufstellen möchte, verwendet man die Zustandsgleichung, im vorliegenden Fall also die Zustandsgleichung des verdünnten Gases, um eine Variablentransformation durchzuführen. Aus Gl. 30.5 folgt beispielsweise $T/V \propto \mathfrak{p}$. Daher kann man die Temperatur in Gl. 31.3 leicht durch den Druck substituieren:

$$\mathfrak{p}V^{c_p/c_V} = \text{konstant} \quad (31.4)$$

Differenziert man diese Gleichung partiell nach dem Druck (die Entropie ist bei der adiabatischen Zustandsgleichung konstant), so ergibt sich $V^{c_p/c_V} + \mathfrak{p}\frac{c_p}{c_V}V^{c_p/c_V}\frac{1}{V}\left(\frac{\partial V}{\partial \mathfrak{p}}\right)_S = 0$. Daraus folgt für die adiabatische Kompressibilität verdünnter Gase:

$$\kappa_S = -\frac{1}{V}\left(\frac{\partial V}{\partial \mathfrak{p}}\right)_S = (c_V/c_p)\frac{1}{\mathfrak{p}}.$$

Sie ist um einen Faktor (c_V/c_p) kleiner als die isotherme Kompressibilität (Gl. 30.4).

32. Wärmepumpen und Wärmekraftmaschinen

Von der Haushaltstechnik bis zur industriellen Produktion sind primär nur makroskopische Zustandsvariable bzw. deren Energiebeiträge von praktischer Bedeutung. Bei jedem Gerät oder Motor geht es darum, eine makroskopische Wirkung zu erzielen, d.h. dass etwas makroskopisch „Sichtbares“ bzw. Beschreibbares geschieht, und somit stehen für den Maschinenbauingenieur erst einmal die f „sichtbaren“ mechanischen Zustandsvariablen und eventuell noch die Dissipationskinetik im Vordergrund.

Wenn System A mit System B Wärme austauscht, so ist das nur möglich, wenn es weder die Energieerhaltung noch das Verbot der Entropieabnahme für das aus A und B bestehende Gesamtsystem G verletzt. Gibt A eine infinitesimale Wärme δQ ab (negativ), so muss B die Wärme $-\delta Q$ aufnehmen (positiv). Daher muss bei Wärmezufuhr an B mindestens ein Sockelbetrag von $T_B dS_B = -T_B dS_A$ in Form von thermischer Energie in B abgelegt werden. Dieser Anteil kommt gewissermaßen „entwertet“ in B an und kann nicht mehr in irgendeine andere Energieform von B überführt werden, weil sonst die Entropie des abgeschlossenen Gesamtsystems sinken würde. Das ist aber nach dem zweiten Hauptsatz verboten.

Wird die Wärme $\delta Q = T_A dS_A$ von A abgegeben, so ist nur die Differenz der an B zugeführten Energie $-T_A dS_A$ und der in die thermische Energieform von B abgelegten Energie $T_B dS_B \geq -T_B dS_A$ frei verfügbar, d.h. nur die Energie

$$(T_A - T_B)dS_B \geq -(T_A dS_A + T_B dS_B) \geq 0. \quad (32.1)$$

Die letzte Ungleichung ist lediglich die triviale Feststellung, dass B maximal so viel Energie aufnehmen kann, wie A abgibt.

Wenn A an B Wärme abgibt und somit Entropie, dann ist die von B aufgenommene Entropie dS_B positiv. Aus Gl. 32.1 folgt, dass dieser Vorgang nur dann erfolgen kann, wenn die Temperatur T_A größer oder gleich T_B ist.

Dividiert man Gl. 32.1 durch die Wärme $-\delta Q = -T_A dS_A$, die von A zur Verfügung gestellt wurde, so ist der Bruchteil der Energie, der ohne Einschränkung frei verfügbar ist, durch den Wirkungsgrad

$$\eta = 1 + \frac{T_B dS_B}{T_A dS_A}$$

in den Grenzen

$$0 \leq \eta \leq \eta_{max}$$

gegeben, wobei der maximal mögliche Wirkungsgrad der Umwandlung von Wärme aus A in nichtthermische Energieformen von B der CARNOTSCHEN WIRKUNGSGRAD

$$\eta_{max} = \frac{(T_A - T_B)dS_A}{T dS_A} = 1 - \frac{T_B}{T_A}. \quad (32.2)$$

ist. Durch Dissipation während des Entropietransports wird die Energie geschmälert, die auf die verschiedenen mechanischen Freiheitsgrade von B verteilt werden kann und daher

ist $\eta \leq \eta_{max}$. Darüber, wie diese Energie sich auf die Freiheitsgrade von B verteilt, machen die mechanischen und thermodynamischen Hauptsätze keine Aussage. Sie darf mit gleichem Recht innerhalb von B dissipiert werden oder irgendwie auf die f nichtthermischen Freiheitsgrade verteilt werden.

Welcher der möglichen Zustände von B am Ende erreicht wird unterliegt keinerlei Beschränkung. Der sich durch Aufnahme einer infinitesimalen Wärme δQ einstellende Endzustand von B hängt davon ab, welchen Weg man zulässt. Man kann beispielsweise die Änderung einiger der Zustandsvariablen durch geeignete Maßnahmen blockieren. Der Endzustand kann also durch eine entsprechende Wahl von Hemmnissen bzw. Zwangsbedingungen gesteuert werden.

WÄRMEPUMPEN und WÄRMEKRAFTMASCHINEN sind periodisch laufende Maschinen, die Wärme und Arbeit mit ihrer Umgebung austauschen. Ihr Arbeitsmedium – meist ein Gas – durchläuft dabei der Reihe nach verschiedene Zustände. Im Zustandsdiagramm des Mediums bzw. der Maschine C wird dabei ein bestimmter Pfad durchlaufen. Maschinen, die im zeitlich stationären Betrieb laufen, kehren nach jedem Zyklus zum gleichen Zustand zurück. Im Zustandsdiagramm legt das Arbeitsmedium und damit die Maschine folglich einen geschlossenen Weg zurück, wie er z.B. in Abb. 50 für den Stirling- und Ottomotor dargestellt ist. Das bedeutet insbesondere, dass die Summe aller der Maschine zugeflossenen Energien über einen solchen KREISPROZESS null ist, denn für die innere Energie des Arbeitsmediums muss für einen Zyklus gelten:

$$\Delta U = \oint dU = 0$$

Hier bedeutet das Symbol \oint eine Integration über den geschlossenen Pfad der Zustände des Kreisprozesses. Genauso muss die Maschine C alle ihr zugeführten Entropien wieder loswerden, aber zusätzlich auch noch jene Entropie ΔS_{gen} , welche die Maschine während des Zyklus durch irreversible Prozesse in ihrem Inneren generiert.

Wärmepumpe: Aufgabe der Wärmepumpe ist es, einem Wärmereservoir B bei einer niedrigeren Temperatur T_B pro Zyklus eine Wärme $\Delta Q_B = T_B \Delta S_B$ zu entnehmen und eine Wärme $\Delta Q_A = T_A \Delta S_B$ an ein Wärmereservoir A bei einer höheren Temperatur $T_A > T_B$ abzugeben. Bei begrenzter Wärmekapazität der Wärmereservoirs wird dadurch die niedrigere Temperatur weiter verringert und die höhere Temperatur weiter erhöht. Es ist offensichtlich, dass das kein erlaubter Prozess ist, denn es wäre die Umkehrung des bekanntermaßen irreversiblen Prozesses des Temperatúrausgleichs. Der Maschine fließt in jedem Zyklus die Entropie $\Delta S_B = \Delta Q_B / T_B$ zu. Diese muss sie zuzüglich der in der Maschine C durch irreversible Prozesse produzierten (positiv) Entropie ΔS_i wieder los werden. Daher muss gelten $\Delta S_A = -(\Delta S_B + \Delta S_i)$. Damit die Maschine ihre Entropiebilanz auf diese Weise in Ordnung bringen kann, muss Sie wegen $|\Delta Q_A| > |\Delta Q_B|$ mehr Wärme abgeben als ihr von B aus zufließt. Wegen der Energieerhaltung ist das so erst einmal nicht zu erfüllen. Möglich wird es jedoch, wenn der Maschine so viel Arbeit ΔW pro Zyklus von außen zufließt, dass sie den Fehlbetrag in der Energiebilanz

ausgleichen kann. Dann gilt insgesamt:

$$\begin{aligned}\Delta S_A + \Delta S_B + \Delta S_i &= 0 \\ \Delta Q_A + \Delta Q_B + \Delta W &= 0\end{aligned}$$

Wärmekraftmaschine: Aufgabe der Wärmekraftmaschine ist es, einem Wärmereservoir A bei einer Temperatur T_A pro Zyklus eine Wärme $\Delta Q_A = T_A \Delta S_A$ zu entziehen und möglichst viel davon als Arbeit ΔW nach außen abzugeben.²⁶ Die abgegebene Arbeit ist dabei die interessierende Größe. Sie lässt sich ökonomisch nutzen, beispielsweise, um eine Dampflokomotive in Bewegung zu setzen oder um Werkzeugmaschinen anzutreiben. Der Wirkungsgrad der Maschine ist der Quotient

$$\eta = -\frac{\Delta W}{\Delta Q_A} \quad (32.3)$$

Das Minuszeichen in Gl. 32.3 rührt daher, dass zugeführte Wärme positiv und abgegebene Arbeit negativ gezählt wird. Es ist offensichtlich, dass das so nicht geht, denn man muss ja über einen Zyklus die der Maschine zugeführte Entropie ΔS_A wieder aus ihr entfernen – zuzüglich der in ihr durch allfällige irreversible Prozesse erzeugten Entropie ΔS_i . Der „Trick“ mit dem man das Ziel dennoch erreichen kann, besteht darin, ein weiteres Wärmereservoir B zu beteiligen, dem man die zu entfernende Entropie $-\Delta S_B = \Delta S_A + \Delta S_i$ mittels einer Wärmemenge $\Delta Q_B = T_B \Delta S_B$ bei der Temperatur T_B zuschieben kann. Eine erste Konsequenz ist, dass die Maschine nur dann überhaupt Arbeit abliefern kann, wenn $T_A > T_B$ und die zweite, dass nur die Differenz zwischen aufgenommener und abgegebener Wärme als Arbeit $-\Delta W = \Delta Q_A + \Delta Q_B$ abgegeben werden kann und daher gilt:

$$\eta = 1 + \frac{\Delta Q_B}{\Delta Q_A} = 1 + \frac{T_B \Delta S_B}{T_A \Delta S_A} = 1 - \frac{T_B}{T_A} - \frac{T_B \Delta S_i}{T_A \Delta S_A} \leq 1 - \frac{T_B}{T_A} = \eta_{max} \quad (32.4)$$

Den größten Wirkungsgrad erhält man, was nicht verwunderlich ist, wenn der Kreisprozess *reversibel* abläuft, d.h. wenn die Maschine während eines Zyklus keine zusätzliche Entropie generiert, wenn also $\Delta S_i = 0$. Das ist der sogenannte Carnotsche Grenzfall einer Wärmekraftmaschine mit dem CARNOTSCHEN WIRKUNGS-GRAD $\eta_{max} = 1 - T_B/T_A$. Er hängt nur vom Temperaturverhältnis der beiden Wärmereservoirs ab, zwischen denen der Kreisprozess abläuft.

Wärmekraftmaschinen werden meist sehr idealisiert diskutiert. Mit dem Stirlingmotor und dem Ottomotor sollen hier zwei Beispiele für idealisierte Viertaktmotoren vorgestellt werden. Damit meint man, dass diese Wärmekraftmaschinen vier thermodynamische Prozess-Schritte aufweisen. Die Zustandsdiagramme für die Kreisprozesse der beiden Wärmekraftmaschinen sind in Abb. 50 dargestellt.

²⁶Zur Vereinfachung wird davon ausgegangen, dass die Wärmeaufnahme von A bei konstanter Temperatur T_A und die weiter unten betrachtete Wärmeabgabe an B bei konstanter Temperatur T_B erfolgen.

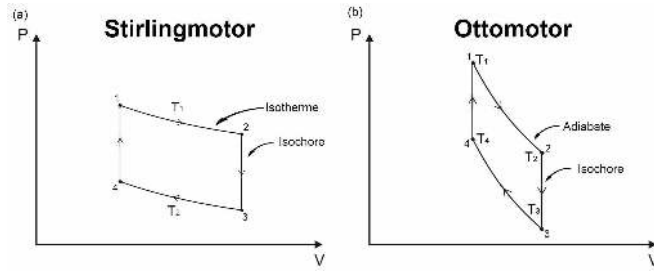


Abbildung 50: $p - V$ -Diagramme der Kreisprozesse von Wärmekraftmaschinen. (a) Stirlingmotor. (b) Ottomotor.

Stirlingmotor. Er besteht aus einem Arbeitsgas, einem Arbeitskolben und einem REGENERATOR, der zur Zwischenspeicherung von Wärme dient. Wie im $p - V$ -Diagramm gezeigt (Abb. 50a), besteht der Stirling-Prozess aus den folgenden vier Arbeitstakten:

1. Vom Zustand (T_1, V_1) ausgehend, lässt man das Arbeitsgas bei einer höheren Temperatur T_1 isotherm von V_1 nach V_2 expandieren. Die Wärmemenge ΔQ_1 wird dem auf T_1 befindlichen Wärmebad²⁷ entnommen und der Arbeitskolben leistet im Optimalfall eine Arbeit $\Delta W_1 = -\Delta Q_1$.
2. Das Arbeitsgas wird isochor auf T_2 abgekühlt. Dabei wird die Wärme $\Delta Q_2 = \Delta U_2$ abgegeben.
3. Man komprimiert das Arbeitsgas bei der niedrigeren Temperatur T_2 isotherm von V_2 nach V_1 . Dafür muss eine Arbeit $\Delta W_2 = -\Delta Q_3$ zugeführt werden.
4. Das Arbeitsgas wird isochor auf T_1 erhitzt. Dabei wird die Wärme $\Delta Q_2 = \Delta U_4$ entnommen und wieder der Ausgangszustand (T_1, V_1) erreicht. Damit ist der Zyklus des Kreisprozesses vollendet.

Der Betrieb des Stirlingmotors setzt bloß voraus, dass zwei Wärmereservoirs auf unterschiedlicher Temperatur vorliegen. Im Idealfall kann die in den Prozessschritten 2 und 4 abgegebene und wieder aufgenommene Wärme durch den Regenerator ideal zwischengespeichert werden und hebt sich bei der Bilanz für den Kreisprozess heraus. Während der beiden isothermen Prozessschritte wird dem Wärmebad mit höherer Temperatur eine Wärme ΔQ_1 entzogen und dem Wärmebad mit der niedrigeren Temperatur die Wärme ΔQ_2 abgegeben. Dadurch wird insgesamt die Arbeit $\Delta W = \Delta W_1 + \Delta W_2$ abgegeben.²⁸

Im theoretischen Grenzfall kann man mit dem Stirlingmotor prinzipiell den Carnotschen Wirkungsgrad erreichen. Wenn man einen Stirlingmotor im umgekehrten Zyklus laufen lässt, wird Arbeit verwendet um Wärme vom kalten zum warmen Bad zu pumpen, d.h. er arbeitet dann als Wärmepumpe.

²⁷ Unter einem Wärmebad versteht man ein System, das Energie nur als Wärme abgeben kann und das so groß ist, dass dabei seine Temperatur näherungsweise unverändert bleibt. So etwas gibt es in der Realität natürlich nicht. Der Begriff Wärmebad bezeichnet ein idealisiertes physikalisches Modell.

²⁸ ΔW_1 ist negativ, ΔW_2 positiv, ΔW also negativ, weil netto Arbeit vom Gas abgegeben wird.

Ottomotor. Im Gegensatz zum Stirlingmotor erfolgt die Wärmezufuhr nicht von außen, sondern geschieht durch einen Verbrennungsvorgang im Innenraum des Zylinders. Ferner wird das heiße Gas nicht abgekühlt und wieder neu verdichtet, sondern einfach ausgestoßen. Der Prozess ist in Abb. 50b skizziert.

1. Im ersten Takt wird das Arbeitsgas durch einen chemischen Verbrennungsvorgang isochor von T_4 auf T_1 erhitzt. Dabei wird chemische Energie in thermische Energie umgewandelt und infolgedessen die Entropie erhöht.
2. Man lässt das heiße Gas bei dem dann vorliegenden hohen Druck adiabatisch expandieren. Dabei wird Arbeit geleistet und die Temperatur fällt.
3. Das immer noch recht warme Gas wird über ein Ventil ausgestoßen und durch neues kaltes Gas-Luft-Gemisch bei T_2 isochor ersetzt.
4. Das Gemisch wird adiabatisch verdichtet, d.h. sein Volumen wird unter Aufwendung von Arbeit verkleinert. Dazu benötigt man wegen des bei niedrigerer Temperatur geringeren Drucks weniger Arbeit als im zweiten Schritt gewonnen wurde.

Durch den Verbrennungsvorgang im ersten Takt wird jene Entropie produziert, die durch den dritten Takt wieder ausgestoßen werden muss, damit keine Entropie im Motor verbleibt und ein Kreisprozess stattfinden kann. Da Ottomotoren recht schnell laufen sollen und es auch tun, sind der zweite und vierte Takt im Idealfall reversible adiabatische Vorgänge (s. Kap. 31), bei denen keine Entropieänderung stattfindet. Im dritten Takt verzichtet man bewusst auf den langsamen Abkühlungsprozess durch Wärmeabfuhr und tauscht warmes Gas einfach gegen kaltes Gas aus. Mit dem austretenden warmen Abgas wird eine ungenutzte Wärmemenge ΔQ_2 samt der darin enthaltenen Entropie ausgestoßen. Aus diesem Grund kann man den Carnotschen Wirkungsgrad prinzipiell nicht erreichen. Durch die Zeitersparnis im Vergleich zum Abkühlungsprozess und auch durch die schnellen adiabatischen Prozessschritte erzielt man mit dem Ottomotor jedoch eine größere Leistung als mit dem Stirlingmotor. Ein Maschinenbauingenieur wird versuchen, hier einen guten Kompromiss zwischen Leistung, Energieeffizienz und Schadstoffausstoß zu finden.

33. Gibbs-Energie

Ab diesem Kapitel soll über das Standardmodell der Thermodynamik hinausgegangen werden. Neben den Energiebeiträgen des Standardmodells möge ein weiterer Energiebeitrag für den betrachteten Stoff identifiziert werden, also ein neuer Energiebeitrag. Daher muss neben den Zustandsvariablen Entropie und Volumen eine weitere extensive Zustandsvariable berücksichtigt werden, die beispielsweise mit z bezeichnet sei und ihre intensive Zustandsvariable sei Z , so dass

$$d\mathcal{U} = TdS - pdV + ZdZ.$$

Zur inneren Energie $\mathcal{U} = \mathcal{U}(S, V, z)$ ist also ein entsprechender (allgemeiner) Energiebeitrag Zdz hinzugekommen. Das kann zum Beispiel ein chemischer (Kap. 34) oder elektrischer Energiebeitrag sein (Kap. 35). Da sich Systeme unter realen Bedingungen stets

in Kontakt mit der Umgebung befinden, ist die Enthalpie $\mathcal{H}(S, \mathbf{p}, z) = \mathcal{U} + pV$ für die Praxis, insbesondere die Laborpraxis, eine weit wichtigere Größe als die innere Energie. Sie hat das Differential

$$d\mathcal{H} = TdS + Vd\mathbf{p} + Zdz. \quad (33.1)$$

Dieses energetische Konzept der Enthalpie muss man zur GIBBS-ENERGIE bzw. FREIEN ENTHALPIE erweitern, wenn Vorgänge betrachtet werden, die sich nicht nur unter konstantem Umgebungsdruck, sondern auch im Temperaturgleichgewicht mit der Umgebung stehen. Die Gibbs-Energie \mathcal{G} ist das durch

$$\mathcal{G} = \mathcal{H} - TS \quad (33.2)$$

definierte thermodynamische Potential mit dem Differential

$$d\mathcal{G} = d\mathcal{H} - d(TS) = SdT + Vd\mathbf{p} + Zdz. \quad (33.3)$$

Wenn man die Gibbs-Energie als Funktion der Variablen T, \mathbf{p} und z formuliert, also $\mathcal{G} = \mathcal{G}(T, \mathbf{p}, z)$, so ist

$$d\mathcal{G} = \left(\frac{\partial \mathcal{G}}{\partial T} \right)_{\mathbf{p}, z} dT + \left(\frac{\partial \mathcal{G}}{\partial \mathbf{p}} \right)_{T, z} d\mathbf{p} + \left(\frac{\partial \mathcal{G}}{\partial z} \right)_{T, \mathbf{p}} dz \quad (33.4)$$

und ein Vergleich mit Gl. 33.3 zeigt beispielsweise, dass

$$S = \left(\frac{\partial \mathcal{G}}{\partial T} \right)_{\mathbf{p}, z}.$$

Für Prozesse, die bei konstanter Temperatur $dT = 0$ und konstantem Druck ($d\mathbf{p} = 0$) ablaufen, also z.B. im Druck- und Temperaturgleichgewicht mit der Umgebung, resultiert die Änderung der Gibbs-Energie alleine aus dem Beitrag der neuen Energieform, d.h.

$$d\mathcal{G} = d\mathcal{H} - TdS = Zdz \quad (33.5)$$

mit

$$Z = \left(\frac{\partial \mathcal{G}}{\partial z} \right)_{T, \mathbf{p}}. \quad (33.6)$$

34. Chemische Energie

Neben der thermischen Energie ist die CHEMISCHE ENERGIE \mathcal{G}_{chem} ein weiteres wichtiges Beispiel nicht-mechanischer Energieformen.

34.1. Heizwert

Die chemische Energie ist sozusagen eine in der chemischen Substanz „eingefrorene Energie“. Beispiele für Substanzen, die Energie abgeben können, sind TREIBSTOFFE, mit denen man ein Kraftfahrzeug antreibt, oder BRENNSTOFFE, die man verbrennt, um Wärme zu gewinnen oder mit denen man eine Brennstoffzelle betreibt, um elektrische Energie zu gewinnen. Sie konstituiert einen neuen Freiheitsgrad der Energie, der in diesem Abschnitt erst einmal etwas oberflächlich eingeführt wird. Die chemische Energie

$$\Delta\mathcal{G}_{chem} = \Gamma_{chem}\Delta m = \rho\Gamma_{chem}\Delta V \quad (34.1)$$

ist der umgesetzten Stoffmenge des Brenn- bzw. Treibstoffes proportional. Für eine homogene Substanz kann man die Stoffmenge durch die Masse und wegen $m = \rho V$ alternativ durch das Volumen V angeben. Hier ist ρ die Massendichte und Γ_{chem} die für eine spezifische chemische Reaktion charakteristische CHEMISCHE ENERGIEDICHTE bzw. der HEIZWERT. Beim chemischen Umsatz der üblichen Treibstoffe (Benzin, Diesel o.ä.) mit Luft ist beispielsweise $\Gamma_{chem} \approx 4 \times 10^4$ kJ/kg. Diese Zahl soll als grobe Vorstellung von der Größenordnung der bei chemischen Umwandlungen freiwerdenden Energien erst einmal genügen.

34.2. Grundgesetze der Chemie

Mit der Zunahme der Kenntnisse der Alchemie kristallisierte sich zunächst einmal der Gedanke der CHEMISCHEN ELEMENTE heraus. Das sind Stoffe, die sich mit den damaligen Methoden nicht mehr in elementarere Stoffe zerlegen ließen. Schon diese Definition lässt erahnen, wie schwer es war, Elemente als solche zu identifizieren. Alle anderen Stoffe sind CHEMISCHE VERBINDUNGEN dieser Elemente. Den chemischen Elementen wurden schon bald Symbole zugeordnet. Heute ordnet man z.B. dem Wasserstoff das Symbol H (Hydrogenium), dem Sauerstoff das Symbol O (Oxygenium) und dem Stickstoff das Symbol N (Nitrogenium) zu. Die Unterscheidung von Elementen und Verbindungen schuf eine erste qualitative Ordnung für das Verständnis des Naturgeschehens, aber noch keine quantitative. Die kam erst durch die Verwendung der Waage zur Bestimmung von Stoffmengen (s. Kap. 18.2). Dabei kommt zugute, dass in chemischen Reaktionen i.A. nur kleine Geschwindigkeiten und kinetische Energien auftreten, so dass der Newtonsche Grenzfall vorliegt. Der bei chemischen Reaktionen auftretende relative Massendefekte $\Delta m/m$ liegt in einer Größenordnung von ca. 10^{-9} . Wenn man nicht zu höchst aufwändigen Methoden mit außergewöhnlich hoher Massenauflösung greift, sind die Massendefekte im Rahmen der in der Chemie üblichen Messgenauigkeit nicht feststellbar und werden daher im Folgenden ignoriert. Grundlage der Chemie sind folgende Erkenntnisse:

1. Lavoisiersches Gesetz der „Massenerhaltung“

Der von Lavoisier (1789) aufgestellte „MASSENERHALTUNGSSATZ“ besagt, dass die Summe der Massen aller Stoffe vor und nach einer chemischen Reaktion gleich sind. Wenn sich die Ausgangsstoffe A und B, zu einem Produkt G verbinden, so bleibt die Masse

$$m_G \approx m_A + m_B \quad (34.2)$$

in chemischer Näherung erhalten.

2. Gesetz der konstanten Proportionen

Bei einer chemischen Reaktion setzen sich die Stoffe in charakteristischen Massenverhältnissen vollständig um. Es verbinden sich z.B. 1 kg Wasserstoff und 35.45 kg Chlor zu 36.45 kg Chlorwasserstoff. Die Massenverhältnisse des Gesetzes der konstanten Proportionen sind in der Regel nicht ganzzahlig.

3. Regel der multiplen Proportionen

Wenn sich mehrere unterschiedliche Verbindungen aus den gleichen chemischen Elementen zusammensetzen, dann stehen deren Massen im Verhältnis kleiner ganzer Zahlen. Beispielsweise sind mehrere Verbindungen des Stickstoffs mit Sauerstoff bekannt, die man als Stickstoffoxide bezeichnet. Wenn man sich für die fünf wichtigeren Stickstoffoxide stets auf die gleiche Stickstoffmasse bezieht, dann verhalten sich deren Sauerstoffmassen im Rahmen der Messunsicherheit wie 1 : 2 : 3 : 4 : 5. Daraus lässt sich die Hypothese ableiten, dass die Sauerstoffmassen in exakt ganzzahligen Verhältnissen vorliegen. In heutiger Notation schreibt man diese Verbindungen deshalb als N_2O , NO , N_2O_3 , NO_2 und N_2O_5 . Diese Bezeichnung für die fünf Stickstoffoxide wurde so gewählt, dass sie zugleich Auskunft darüber gibt, in welchen Massenverhältnissen sie sich aus ihren beiden Elementen zusammensetzen.

4. Das Avogadrosche Gesetz

Avogadro untersuchte die chemische Umsetzung von Gasen. Wenn man Gase bei gleicher Temperatur und gleichem Druck miteinander vergleicht, dann stehen die Volumina der Ausgangsstoffe sowie der Endprodukte einer chemischen Umsetzung im Verhältnis kleiner ganzer Zahlen (AVOGADROSCHES GESETZ). Wenn sich beispielsweise Wasserstoff und Chlor zu Chlorwasserstoff verbinden, dann verhalten sich die Volumina wie 1:1 und wenn Wasserstoff und Sauerstoff sich zu Wasser verbinden, wie 1:2 (s. auch Kap. 35.5.2).

34.3. Der chemische Stoffmengenbegriff

Im Rahmen des Kontinuumsmodells der Materie legen die vier genannten chemischen Gesetzmäßigkeiten nahe, dass sich homogene Stoffe bei chemischen Reaktionen in ganzzahligen Verhältnissen ihrer STOFFMENGE n umsetzen. Die stärkste Stütze dieser Hypothese ist zweifellos das Gesetz der multiplen Proportionen.

Gesetz der ganzzahligen Stoffmengenverhältnisse

Bei physikalischen oder chemischen Reaktionen setzen sich die Mengen der beteiligten Stoffe in ganzzahligen Verhältnissen um.

In dieser neuen Formulierung gelten die historisch ursprünglich mit der Waage erforschten Gesetze nun auch dann, wenn, wie in der Teilchenphysik, enorm große Massendefekte bei den Reaktionen auftreten. Wenn man nämlich die als „Massenerhaltung“ bezeichnete Lavoisiersche Erkenntnis als ein Gesetz über Stoffmengenumsätze interpretiert, dann

erübrigt sich auch die oben kurz angesprochene Diskussion über den Massendefekt. Im historischen Rückblick könnte man sogar salopp sagen, dass es ein großes Missverständnis war, die Lavoisiersche Erkenntnis als „Massenerhaltung“ zu bezeichnen, denn sie beruht letztlich auf dem Gesetz der ganzzahligen Stoffmengenverhältnisse.

Wir lösen nun die Masse als unser bis hierhin vorläufiges Maß der Stoffmenge (s. Kap. 18.2) durch das aus dem Gesetz der multiplen Proportionen folgende, also chemisch begründete Maß der Stoffmenge ab. Aus den chemischen Experimenten gehen nur Verhältnisse dieser Stoffmengen hervor. Daher muss man die Masse irgendeines Stoffes als Maß festlegen, das die Einheit definiert. Ursprünglich war das 1 g elementarer Wasserstoff. Heute bezieht man sich auf ^{12}C . Die Einheit der Stoffmenge nennt man MOL:

$$[n] = 1 \text{ Mol.}$$

Masse und Stoffmenge eines homogenen chemischen Stoffes sind einander proportional:

$$m = nM \quad (34.3)$$

Den stoffspezifischen Proportionalitätsfaktor M bezeichnet man als MOLARE MASSE bzw. MOLMASSE. Für elementaren Wasserstoff ist der Proportionalitätsfaktor beispielsweise $M(\text{H}) = 1.00 \text{ g/Mol}$ und für Chlor $M(\text{Cl}) = 35.5 \text{ g/Mol}$. Bei einer chemischen Reaktion zu Chlorwasserstoff (HCl) setzt sich eine Stoffmenge n des Wasserstoffs mit der exakt gleichen Stoffmenge n des Chlors zu einer Stoffmenge n des Stoffes HCl um, also im Verhältnis $\text{H}:\text{Cl}:\text{HCl}$ von 1:1:1 bezüglich der (chemischen) Stoffmengen, aber im Verhältnis von 1.00 : 35.5 : 36.5 bezüglich ihrer physikalischen Stoffmengen bzw. Massen. Im Folgenden wird der Begriff der Stoffmenge im Sinne des chemischen Stoffmengenbegriffs aufgefasst. Mit ihm gelangt man zu einer erstaunlichen Erkenntnis über die spezifischen Wärmekapazitäten und die inneren Energien von Gasen, was schließlich zum Modell des idealen Gases führt.

34.4. Das ideale Gas

Für chemisch unterschiedliche Stoffe ergeben sich keine erkennbaren Regelmäßigkeiten für die spezifischen Wärmen c_p bei konstantem Druck bzw. c_V bei konstantem Volumen, und auch keine für die Differenz $c_p - c_V$. Mit der Beziehung $m = nM$ zwischen Masse m , Stoffmenge n und Molmasse M kann man die spezifischen Wärmen von der Stoffmengeneinheit in Kilogramm auf die von den Chemikern favorisierte Stoffmengeneinheit in Mol umrechnen. Für diese MOLWÄRMEN $C_V = Mc_V$ und $C_p = Mc_p$ bei konstantem Volumen bzw. Druck kommt man zur verblüffenden Feststellung, dass die Differenz $C_p - C_V$ mit abnehmender Dichte der Gase gegen den gleichen Grenzwert

$$\lim_{\rho \rightarrow 0} (C_p - C_V) \rightarrow R = 8.3 \text{ J/Mol} \cdot \text{K} \quad (34.4)$$

konvergieren, und zwar unabhängig von der chemischen Beschaffenheit des Gases! Man bezeichnet diesen Grenzwert R daher als UNIVERSELLE GASKONSTANTE. Dieses extrapolierte Verhalten von Gasen modelliert man, indem man durch

$$pV = nRT \quad (34.5)$$

die Zustandsgleichung eines Modellgases definiert. Es wird als IDEALES GAS bezeichnet, weil in seine Zustandsgleichung keine stoffspezifischen Eigenschaften mehr eingehen. Dass die Zustandsgleichungen realer Gase unterschiedlicher chemischer Beschaffenheit im Grenzfall tatsächlich einheitlich durch Gl. 34.5 modelliert werden können, ist ein äußerst erstaunlicher Sachverhalt.

Wenn man die Ergebnisse empirischer Untersuchungen zur Wärmekapazität auf den Grenzfall des verdünnten Gases hin extrapoliert, so konvergiert die Molwärme von Edelgasen und Metallgasen stoffunabhängig (!) auf den Wert

$$C_{V0} = \frac{3}{2}R.$$

Der Index „0“ soll hier darauf hinweisen, dass hier die Molwärme gemeint ist, die man dem idealen Gases zuschreibt. Für die innere Energie, welche für verdünnte Gase allein eine Funktion der Temperatur ist, folgt mit Gl. 30.6 für ein ideales Gas somit

$$\mathcal{U} = mc_V T = nC_{V0}T = \frac{3}{2}nRT. \quad (34.6)$$

34.5. Chemisches Potential

Mit der Stoffmenge n als extensiver Zustandsvariabler kann man nun die bisherige im Standardmodells der Thermodynamik betrachtete inneren Energie um die chemische Energie ergänzen:

$$d\mathcal{U} = TdS - pdV + \mu dn. \quad (34.7)$$

Die zur Substanzmenge n konjugierte intensive Zustandsvariable

$$\mu = \left(\frac{\partial \mathcal{U}}{\partial n} \right)_{S,p} \quad (34.8)$$

heißt CHEMISCHES POTENTIAL.

Mit Gl. 34.7 kann man jedoch überhaupt nichts anfangen, denn die chemische Energie unterscheidet sich fundamental von den anderen beiden Beiträgen. Das liegt daran, dass das Volumen V überhaupt keiner Einschränkung unterliegt und die Entropie S nur durch einen halben Erhaltungssatz beschränkt ist, denn sie darf anwachsen. Die Substanzmenge aber unterliegt einem vollständigen Erhaltungssatz: Substanz kann man nicht wegzaubern. Das bedeutet, dass wegen der Substanzerhaltung nur $dn = 0$ gelten kann.

Der einfachst mögliche Fall, für den das chemische Potential überhaupt von Relevanz ist, liegt vor, wenn mindestens zwei Stoffe A und B existieren, welche untereinander eine Umwandlung



ausführen können. Das ist die elementarste chemische Reaktion. Umwandlungen dieser einfachsten Art liegen beispielsweise bei Übergängen zwischen verschiedenen Aggregatzuständen (fest, flüssig, gasförmig), unterschiedlichen Kristallstrukturen (z.B. ferromagnetisches α -Eisen und antiferromagnetisches γ -Eisen) oder unterschiedlichen allotropen Modifikationen (z.B. Sauerstoff und Ozon) vor. Umwandlungen von Substanzen sind deshalb erlaubt, weil sie nicht die Substanz selbst ändern, sondern nur ihre Form. Die in verschiedenen Formen vorliegenden Stoffmengen unterliegen dabei der Bedingungsgleichung

$$dn_A = -dn_B \quad (34.9)$$

der Substanzerhaltung.

Das einfachst mögliche Energiemodell für zwei Stoffe ist also

$$d\mathcal{U} = TdS_A + TdS_B + p dV_A + p dV_B + \mu_A dn_A + \mu_B dn_B. \quad (34.10)$$

Da sich die beiden Reaktionspartner A und B i.A. in innigem Kontakt miteinander befinden, sind Druck und Temperatur der beiden aus A und B bestehenden Phasen gleich. Mit s_A und s_B seien nun die spezifischen Entropien der beiden Phasen bezeichnet. Da die Entropie eine extensive Größe ist, ist die jeweilige Entropie der Phasen $S_A = s_A n_A$ bzw. $S_B = s_B n_B$. Unter Beachtung von Gl. 34.9 gilt für eine chemische Stoffumwandlung

$$d\mathcal{U} = (Ts_A + \mu_A)dn_A + (Ts_B + \mu_B)dn_B + p dV. \quad (34.11)$$

Wenn man eine chemische Reaktion bei konstantem Volumen ($dV = 0$) ablaufen lassen könnte, wäre $d\mathcal{U}$ gleich der Energie $\delta Q + \delta W$, die das Zweiphasensystem mit der Umgebung handeln kann, indem es die Stoffmenge von A ändert. Chemische Reaktionen laufen aber i.A. in der freien Atmosphäre ab und damit unter konstantem Druck. Daher ist es praxisnäher, wenn man anstatt Gl. 34.11 die entsprechende Formulierung für die Enthalpie verwendet:

$$d\mathcal{H} = \mathcal{h}_A dn_A + \mathcal{h}_B dn_B + V dp = \delta Q + \delta W, \quad (34.12)$$

wobei

$$\mathcal{h}_A = \mu_A + Ts_A$$

die spezifische Enthalpie für die Phase A ist. Der Ausdruck für \mathcal{h}_B ist entsprechend. Für eine Reaktion unter konstantem Druck ($dp = 0$), bei der eine Stoffmenge Δn_A umgesetzt wird, ist daher die Enthalpieänderung

$$\Delta\mathcal{H} = (\mathcal{h}_A - \mathcal{h}_B)\Delta n_A = \Delta Q + \Delta W. \quad (34.13)$$

gleich der Energie $\Delta Q + \Delta W$, die an die Umgebung abgegeben oder von ihr aufgenommen wird, wenn die sich der Stoff A um Δn_A ändert. Die spezifischen Umwandlungs- bzw. Reaktionsenthalpien \mathcal{h}_A bzw. \mathcal{h}_B sind für Standardbedingungen (s. Kap. 27.2) tabelliert.

Gl. 34.13 ist für die Reaktion $A \rightarrow B$ folgendermaßen zu lesen: Wenn die Stoffmenge von A um Δn_A abnimmt (Δn_A negativ), dann resultiert die „Vernichtung“ dieser Stoffmenge

von A in einer Abnahme der Enthalpie um $\bar{h}_A \Delta n_A$. Die mit der Reaktion verbundene „Erzeugung“ von B um $\Delta n_B = -\Delta n_A$ lässt die Enthalpie um $-h_B \Delta n_A$ anwachsen. Durch den Vernichtungs-/Erzeugungsprozess hat sich die Enthalpie insgesamt um $\Delta \mathcal{H}$ geändert. Die Enthalpieänderung für die umgekehrte Reaktion $B \rightarrow A$ ist dann gerade $-\Delta \mathcal{H}$.

Um einen Festkörper (Phase A) bei konstantem Druck zu schmelzen oder eine Flüssigkeit (Phase A) zu verdampfen, muss man gemäß Gl. 34.13 eine der umgesetzten Stoffmenge n_A proportionale Energie zuführen. Die spezifische Enthalpie $\bar{h}_{AB} = \bar{h}_A - \bar{h}_B$ heißt im ersten Fall SPEZIFISCHE SCHMELZENTHALPIE (B flüssige Phase) und im zweiten Fall (B gasförmige Phase) SPEZIFISCHE VERDAMPFUNGSENTHALPIE.

34.6. Phasenumwandlungen und chemische Reaktionen

Als Beispiel für die Umwandlungen zwischen zwei Stoffen soll die Umwandlung zwischen zwei Aggregatzuständen diskutiert werden, wie sie beispielsweise bei der Umwandlung des Wassers von der gasförmigen in die flüssige Phase auftritt. Die Ursache des Phänomens, dass zwei Phasen auftreten können, erkennt man bereits aus Gl. 30.3: Wenn man bei gleichem Druck und Volumen die Stoffmenge und somit die Dichte erhöht, wird die isotherme Kompressibilität für $\mathbf{p} = a\rho^2$ unendlich. Das bedeutet, dass der Materiezustand, hier die Gasphase, instabil wird und ein PHASENÜBERGANG zu einer anderen Phase auftritt, hier zur flüssigen Phase. Die van-der-Waals-Gleichung (Gl. 30.2) beschreibt sowohl die flüssige als auch die gasförmige Phase in akzeptabler Näherung, allerdings unterscheiden sich die Parameter a , V_{min} und $c_p - c_V$ für die Flüssigkeit und das Gas i.A. erheblich. Für Gase kann die Kompressibilität oberhalb einer bestimmten Temperatur T_{crit} , der kritischen Temperatur, bzw. oberhalb kritischen kritischen Drucks \mathbf{p}_{crit} nicht mehr negativ werden. Oberhalb des KRITISCHEN PUNKTS ($T_{crit}, \mathbf{p}_{crit}$) im T - \mathbf{p} -Zustandsdiagramm kann kein Phasenübergang mehr stattfinden.

Es sei nun der Phasenübergang zwischen Eis und flüssigem Wasser betrachtet. Eis soll die Phase A sein und flüssiges Wasser die Phase B. Die spezifische Bildungsenthalpie \bar{h}_A für Eis ist kleiner als die für flüssiges Wasser, d.h. $\bar{h}_A < \bar{h}_B$. Die Schmelzenthalpie ist die Differenz dieser beiden spezifischen Bildungsenthalpien. Wenn man Eis schmelzen will, so muss man gemäß Gl. 34.13 von außen Energie zuführen, d.h. $\Delta Q + \Delta W$ muss positiv sein. Beim umgekehrten Prozess wird die gleiche Energiemenge als Wärme und/oder Arbeit nach außen abgegeben.

Nun soll die Frage untersucht werden, wieviel davon maximal als Arbeit nach außen abgegeben werden kann. Dabei ist zu beachten, dass die spezifische Entropie s_A von Eis niedriger als die spezifische Entropie s_B von Wasser ist. Beim Erstarrungsvorgang wird zwar Energie frei, aber es nimmt auch die Entropie ab. Das geht nur, wenn ein Teil der aus der Enthalpieabnahme zur Verfügung stehenden Energie in Form von Wärme an die Umgebung abgegeben wird, und zwar mindestens so viel, wie im Zweiphasensystem an Entropie bei der Umwandlung $B \rightarrow A$ generiert wird. Wenn der Erstarrungsprozess bei einer Temperatur T stattfindet, so kann maximal die Energie $\bar{h}_{AB} \Delta n_A - T(s_A - s_B) \Delta n_A$ als Arbeit abgegeben werden. Weniger ist immer möglich, weil die Summe der Entropie

des Zweiphasensystems und seiner Umgebung immer anwachsen darf. Dieser Sachverhalt lässt sich am elegantesten durch die Gibbs-Energie:

$$\Delta\mathcal{G} = (\hbar_{AB} - Ts_{AB})\Delta n_A \quad (34.14)$$

mit der spezifischen Entropie $s_{AB} = s_A - s_B$ beschreiben: Die Erstarrung kann nur dann „spontan“ ablaufen, d.h. ohne Verletzung des zweiten Hauptsatzes, wenn die Temperatur so niedrig ist, dass $\hbar_{AB} > Ts_{AB}$ gilt. Nur in diesem Fall kann das Zweiphasensystem so viel Energie abgeben, dass die durch Erstarrung erniedrigte Entropie des Zweiphasensystems durch Abgabe von Wärme nach außen kompensiert werden kann.

Ist nun umgekehrt die Temperatur so hoch, dass $\hbar_{AB} < Ts_{AB}$, so kann das in thermischem Kontakt mit der Umgebung stehende Zweiphasensystem Wärme aus der Umgebung abziehen und einen Schmelzvorgang einleiten. Auch dieser entropiegetriebene Vorgang kann „spontan“ ablaufen, denn er verletzt den zweiten Hauptsatz der Thermodynamik nicht.

Die Gibbs-Energie stellt also ein klares Kriterium bereit, wann eine chemische Umwandlung spontan abläuft, wann Gleichgewicht herrscht und wann sie gar nicht abläuft:

$\Delta\mathcal{G} < 0$	Reaktion läuft spontan ab
$\Delta\mathcal{G} = 0$	Chemisches Gleichgewicht
$\Delta\mathcal{G} > 0$	Reaktion kann nicht ablaufen

Aus Gl. 34.14 ergeben sich damit die folgenden vier Alternativen für den Ablauf von Phasenübergängen oder Reaktionen:

$\Delta\mathcal{G} = \Delta\mathcal{H} - T\Delta S$	$\Delta S < 0$	$\Delta S > 0$
$\Delta\mathcal{H} < 0$	bei niedriger Temperatur spontan (enthalpiegetrieben)	immer spontan (instabil)
$\Delta\mathcal{H} > 0$	unmöglich	bei hoher Temperatur spontan (entropiegetrieben)

Die Gibbsenergie

$$\mathcal{G} = \mathcal{G}(T, \mathbf{p}, n_A) = \mathcal{H}_A + \mathcal{H}_B - T(S_A + S_B) = \hbar_A n_A + \hbar_B n_B - T(s_A n_A + s_B n_B)$$

eines Zweiphasensystems ändert sich (für fixiertes T und \mathbf{p}) bei einem Schmelzvorgang bezüglich ihrer extensiven Größen gemäß

$$\Delta\mathcal{G}(T, \mathbf{p}, n_A) = (\hbar_{AB} - Ts_{AB})\Delta n_A = (\mu_A - \mu_B)\Delta n_A. \quad (34.15)$$

Hier sind $\hbar_{AB} = \hbar_A - \hbar_B$ die spezifische Schmelzenthalpie und $s_{AB} = s_A - s_B$ die spezifische Schmelzentropie, und die rechte Seite folgt aus einer Anwendung von Gl. 33.5 auf Zweiphasensysteme bei konstantem Druck und konstanter Temperatur. Im Grenzfall der Annäherung einer spontanen Reaktion an das chemische Gleichgewicht, also im

Grenzfall $\Delta\mathcal{G} \rightarrow 0$, müssen die Ausdrücke in den runden Klammern in Gl. 34.15 verschwinden. Das bedeutet zum einen, dass im thermischen Gleichgewicht eine Enthalpiezunahme

$$\cancel{\ell}_{AB}\Delta n_A = T s_{AB}\Delta n_A = \Delta Q$$

allein durch Zufuhr von Wärme vollzieht. Die Arbeit ist hier aus dem Spiel. Ferner folgt dass

$$\mu_A(T, \mathbf{p}) - \mu_B(T, \mathbf{p}) = 0. \quad (34.16)$$

Wenn A daher eine feste Phase ist, die durch Schmelzen in eine flüssige Phase B übergehen kann, so gibt es für einen gegebenen Druck \mathbf{p} eine ganz bestimmte Temperatur T_m , für welche Gl. 34.16 erfüllt werden kann. Diese Temperatur heißt SCHMELZTEMPERATUR. Wird sie überschritten, geht A vollständig in B über, wird sie unterschritten, erstarrt Phase B vollständig zur Phase A. Wie aus Gl. 34.16 ersichtlich, ist die Schmelztemperatur eine Funktion $T_m = T_m(\mathbf{p})$ des Drucks. Die inverse Funktion $\mathbf{p} = \mathbf{p}(T_m)$ heißt SCHMELZKURVE. Aus den analogen Gründen gibt es für den Übergang von der flüssigen in die gasförmige Phase eine Siedekurve auf der beide Phasen koexistieren können. Wenn alle drei Phasen, fest, flüssig und gasförmig, koexistieren sollen, dann kann das nur an einem Punkt stattfinden, der auf beiden Kurven liegt und daher ist das nur bei einer ganz bestimmten Temperatur und einem ganz bestimmten Druck möglich. Das ist der TRIPELPUNKT. Da für den Tripelpunkt keine Freiheit der Zustandsvariablen mehr existiert, ist er für die Festlegung von Bezugspunkten für die Temperaturskala besonders geeignet.

Da die Volumina wieder extensive Größen sind, führt man wieder spezifische Volumina v_A bzw. v_B für A bzw. B ein, so dass $V_A + V_B = v_A n_A + v_B n_B$. Das Differential der Funktion $\Delta\mathcal{G}(T, \mathbf{p}, n_A)$ im chemischen Gleichgewicht mit $\mu_A = \mu_B$ ist dann

$$d\Delta\mathcal{G} = (dh_{AB} - T_m s_{AB})\Delta n_A = [-s_{AB}dT_m + (v_A - v_B)d\mathbf{p}]\Delta n_A = 0. \quad (34.17)$$

Aus dieser Gleichung folgt unmittelbar die CLAUSIUS-CLAPEYRON-GLEICHUNG:

$$\frac{d\mathbf{p}}{dT_m} = \frac{\Delta s_{AB}}{(v_A - v_B)} = \frac{\Delta \cancel{\ell}_{AB}}{T_m(v_A - v_B)} \quad (34.18)$$

welche die Steigung der Schmelzkurve beschreibt. Sie ist für den Übergang fest-flüssig wesentlich steiler als für den Übergang flüssig-gasförmig, weil sich die spezifischen Volumina der festen und der flüssigen Phase nur wenig unterscheiden.

Die Koexistenzbereiche der festen, flüssigen und gasförmigen Phase können in einem p-T-Diagramm dargestellt werden (Abb. 51). Folgt man der Koexistenzgrenze zwischen flüssiger und gasförmiger Phase mit zunehmender Temperatur, so endet sie am kritischen Punkt. Wegen der hohen Kompressibilität am kritischen Punkt können dort enorme Dichteschwankungen auftreten. Wenn sich ein transparentes Medium dem kritischen Punkt nähert, tritt daher eine sehr starke Lichtstreuung auf (kritische Opaleszenz). Oberhalb des kritischen Punktes kann man nicht mehr von einer flüssigen oder gasförmigen Phase sprechen. Der Punkt des Phasendiagramms, in dem alle Koexistenzlinien zusammenlaufen ist der TRIPELPUNKT. Hier können alle drei Phasen (die feste, flüssige und gasförmige) koexistieren.

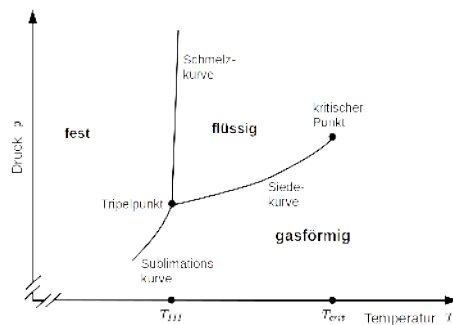


Abbildung 51: Phasendiagramm. Aufgetragen ist der Druck p als Funktion der Temperatur T . Sublimation findet an der Grenze zwischen fester und gasförmiger Phase statt, Schmelzen an der flüssig-festen und Verdampfung bzw. Sieden an der flüssig-gasförmigen Zweiphasenlinie. Letztere endet im kritischen Punkt bei der Temperatur T_{crit} . Am Tripelpunkt liegen alle drei Phasen gleichzeitig vor.

35. Elektrische Energie

35.1. Elektrizität

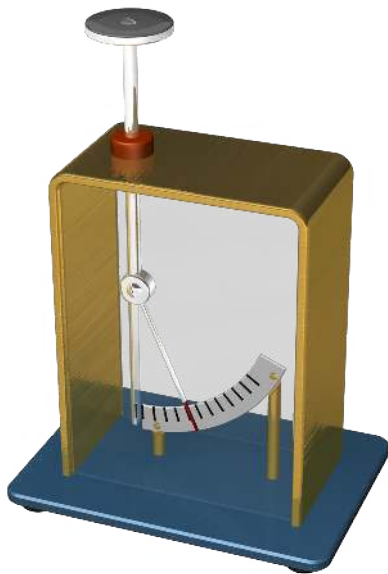
Die Tribologie untersucht Phänomene, bei denen Grenzflächen eng miteinander in Kontakt gebracht werden. Hierzu gehören die Grenzflächenenergie (Kap. 18.7.1), das Haften (Kap. 18.7.2) und die Gleitreibung (Kap. 18.7.3). Sie treten nur so lange in Erscheinung, wie die Grenzflächen in engem Kontakt zueinander verbleiben. Vor mehr als 2500 Jahren berichtete Thales von Milet jedoch erstmals, dass an geriebenem Bernstein auch nach der Trennung des Grenzflächenkontakts eine fortdauernde Wirkung beobachtet wird. Sie äußert sich u.a. in einer über große Distanzen wirksamen Anziehung bzw. Abstoßung anderer Gegenstände äußert. Diese Erscheinung wurde nach dem altgriechischen Wort „Elektron“ für den Bernstein als ELEKTRIZITÄT bezeichnet. Wenn man betonen will, dass sie durch Reibung hervorgerufen wurde, spricht man von REIBUNGSELEKTRIZITÄT bzw. TRIBOELEKTRIZITÄT.

Reibt man einen Luftballon an der Kleidung oder einen Glas- oder Kunststoffstab an einem Stück Fell, kann man mit den so ELEKTRISIERTEN Objekten Wirkungen über große Entfernungen auslösen. In dieser Hinsicht ähnelt das Phänomen der Gravitation, jedoch treten nicht nur anziehende, sondern auch abstoßende Wirkungen auf. Geht man von einem Glasstab aus, der mit einem Fell elektrisiert wurde, so lässt sich die Elektrizität aller anderen elektrisierten Objekte danach einteilen, ob sie vom elektrisierten Glasstab abgestoßen oder angezogen werden. Die Elektrizität der ersten Gruppe wird als POSITIV und die der zweiten als NEGATIV klassifiziert. Diese Konvention geht auf Benjamin



Abbildung 52: Benjamin Franklin auf der 100-Dollar-Note

Der Physiker Franklin (1706-1790) führte die Bezeichnungen „positive Elektrizität“ und „negative Elektrizität“ ein.



Die wesentlichen Elemente der gezeigten Elektrometers sind ein unbeweglicher Metallstab, der durch eine nichtmetallische Durchführung gehalten ist. Daran ist ein kleiner Metallzeiger drehbar befestigt, dessen Ausschlag man auf einer Skala ablesen kann. In seinem natürlichen Gleichgewichtszustand hängt der Zeiger aufgrund der Schwerkraft nach unten. Das entspricht der Auslenkung null auf der Skala. Nach Elektrisierung wird ein permanenter Ausschlag angezeigt.

Abbildung 53: Elektroskop

Franklin (Abb. 52) zurück.²⁹

Einige Aspekte der Elektrizität lassen sich mit einem sehr einfachen Instrument untersuchen, dem ELEKTROSKOP. Abb. 53 zeigt eine der vielfältigen technischen Realisierungsmöglichkeiten. Das Elektroskop spielt für die empirische Aufklärung elektrischer Phänomene eine ähnliche Funktion wie Thermoskope für die empirische Untersuchung thermischer Phänomene (s. Kap. 26.1).

Wenn sich ein elektrisiertes Objekt, z.B. ein elektrisierter Stab, einem Elektroskop nä-

²⁹Benjamin Franklin setzte sich für Demokratie, Bürgerfreiheiten und die amerikanische Unabhängigkeitsbewegung ein. Als Freigeist und Freimaurer war er ein Vorkämpfer gegen die Sklaverei und religiöse Intoleranz. Durch die Erfindung des Blitzableiters und seine Initiative zur Gründung der freiwilligen Feuerwehr verbannte er die Brandgefahr aus den Städten. Der erste, der zwei Ladungsarten unterschied, war Charles du Fay (1733). Die Elektrizität eines geriebenen Glasstabs nannte er *électricité vitreuse* (Glaselektrizität).

hert, schlägt der Zeiger aus. Entfernt man es wieder, dann geht auch der Zeigerausschlag wieder auf null zurück. Dieses Phänomen demonstriert augenfällig, dass elektrische Wechselwirkungen durch die Luft hindurch über eine große Entfernung wirksam sein können, d.h. über eine große Reichweite verfügen. Ferner zeigt das Experiment aber auch, dass die Stärke dieser langreichweitigen Wechselwirkung mit der Entfernung abnimmt. Weitere Details sollen erst später weiter untersucht werden. In den nächsten Kapiteln soll es aber zunächst einmal um das Phänomen gehen, dass ein permanenter Zeigerausschlag zurückbleibt, wenn der der nach außen geführte Metallstab durch einen elektrisierten Stab berührt wird. Er bleibt bestehen, auch wenn man den elektrisierten Stab anschließend so weit entfernt, dass keine von ihm herrührenden entfernungsabhängigen Wirkungen mehr feststellbar sind. Der Zeigerausschlag kann im Beobachtungszeitraum von einigen Stunden scheinbar unverändert bleiben. Mit einem zweiten Elektroskop kann man nachweisen, dass vom ersten Elektroskop anschließend die gleichen Phänomene hervorgerufen werden, wie durch jedes andere elektrisierte Objekt. Daraus folgt, dass ein Elektroskop durch die Berührung mit einem elektrisierten Objekt selbst elektrisiert wird.

35.2. Elektrische Energie

An einer Reihe von Indikatoren kann man ablesen, dass Elektrisierung ein energetisches Phänomen ist: Wenn ein Elektroskop elektrisiert wird, setzt sich beispielsweise der Zeiger in Bewegung (kinetische Energie), dabei ändert seine potentielle Energie der Schwere, oszilliert und kommt anschließend zur Ruhe (Dissipation). Es kann also kein Zweifel daran bestehen, dass die innere Energie \mathcal{U} des Elektroskops durch das Elektrisieren zugenommen hat. Man hat es mit einem neuen Freiheitsgrad der Energie zu tun, der ELEKTRISCHEN ENERGIE. Man beschreibt ihn in der bewährten Weise durch eine extensive und eine dazu konjugierte intensive Zustandsvariable. Erstere wird als ELEKTRISCHE LADUNG q bezeichnet und letztere (analog zum chemischen Potential μ) als ELEKTRISCHES POTENTIAL φ . Wenn man ein energetisches Modell mit thermischer, elastischer, chemischer und elektrischer Energie konzipiert, so hätte die Änderung der Energie $\mathcal{U} = \mathcal{U}(S, V, n_1, n_2, \dots, q_1, q_2, \dots)$ die Form

$$d\mathcal{U} = T dS - p dV + \sum \mu_i n_i + \sum \varphi_i dq_i \quad (35.1)$$

mit dem neuen Beitrag $\varphi_i dq_i$ für die elektrische Energie der i -ten geladenen Stoffspezies. Wenn man ein elektrisches System mit nur einer einzigen Ladungsspezies bei konstanter Temperatur und konstantem Druck betrachtet, ist die Änderung der Gibbsenergie gerade

$$d\mathcal{G} = \varphi dq$$

mit dem elektrischen Potential

$$\varphi = \frac{\partial \mathcal{G}}{\partial q}. \quad (35.2)$$

35.3. Ladungserhaltung

Das zentrale Axiom der LADUNGSEHALTUNG hat für die Elektrodynamik gleichen bedeutenden Rang wie z.B. die Impulserhaltung für die Mechanik.

Axiom der Ladungserhaltung

Die Ladung ist eine invariante additive Erhaltungsgröße. In einem abgeschlossenen System laufen alle physikalischen Vorgänge so ab, dass die Ladung q erhalten bleibt.

Trotz der gewaltigen Zahl an Experimenten und Anwendungen, in denen Elektrizität eine Rolle spielt, und auch nach mehr als einem Jahrhundert der Prüfung hat sich nie ein Widerspruch zur Hypothese der Ladungserhaltung ergeben. Vergleicht man die Ladung mit der Masse (invariant, aber nicht erhalten) und der Energie (erhalten, aber nicht invariant), so sind die Anforderungen an die Ladung strenger: Sie ist sowohl invariant als auch erhalten!

Wenn die Summe der Impulse zweier Teilchen am Anfang am Anfang einer Wechselwirkung null war, dann sind die Impulse am Ende betragsmäßig gleich, haben aber entgegengesetztes Vorzeichen. Genauso verhält es sich mit der Ladung: Wenn A und B zwei Teilsysteme mit den Ladungen q_A und q_B sind, dann besitzt das Gesamtsystem G wegen der Additivität, die jeder extensiven Größe zukommt, die Ladung $q_G = q_A + q_B$. Wenn irgendein physikalischer Prozess stattfindet und die Teilsysteme anschließend die Ladungen \tilde{q}_A und \tilde{q}_B besitzen, so verlangt die Ladungserhaltung:

$$q_A + q_B = q_G = \tilde{q}_A + \tilde{q}_B \quad (35.3)$$

Das Axiom der Ladungserhaltung ist weitreichend: Egal aus wieviel Einzelteilen ein abgeschlossenes System vor und nach dem Prozess besteht und völlig egal, ob im Prozess neue Teile generiert oder vernichtet werden: Wenn man die Summe der Ladungen aller vor dem Prozess vorhandenen Teile bildet und jene aller Teile nach dem Prozess, dann ist die Summe gleich. So weit ist die Ladungserhaltung ähnlich zur Impulserhaltung. Führt man die Beobachtungen in einem System S' aus, das sich relativ zu S mit der Geschwindigkeit v bewegt, so ist im Unterschied zum Impuls wegen der Invarianz der Ladung $q_A = q'_A$ usw., d.h. die Ladungen sind die gleichen – egal in welchem Bezugssystem man sie misst.

Makroskopische Objekte tragen üblicherweise eine Ladung $q = 0$, d.h. sie begegnen uns in der Regel zunächst als ungeladene Objekte. Wurde ein zuvor ungeladenes Objekt elektrisiert, sagt man auch, dass es anschließend elektrisch geladen ist. Werden Ladungen aus ungeladener Materie „generiert“, so muss bei der Elektrisierung stets genauso viel positive wie negative Ladung entstehen, d.h. der Prozess der „Ladungserzeugung“ ist letztendlich immer eine LADUNGSTRENNUNG.

Beispielsweise sei A ein zunächst ungeladenes Stück Fell und B ein ungeladener Glasstab. Die Summe der Ladungen, d.h. die Gesamtladung, ist anfangs $q_A + q_B = q = 0$, weil sich auf beiden Objekten die Ladung $q_A = q_B = 0$ befindet. Nachdem man die beiden Objekte in Kontakt gebracht hat, hat der Glasstab infolge der triboelektrischen

Wechselwirkung anschließend eine positive Ladung $\tilde{q}_A = +q_0$. Folglich muss auf dem Fell anschließend eine gleich große entgegengesetzte Ladung $\tilde{q}_B = -q_0$ sein. Das kann man durch folgendes Experiment auf die Probe stellen: Auf ein Elektroskop steckt man einen Metallbecher auf. Man nimmt einen ungeladenen Glasstab und ein ungeladenes Fell und elektrisiert sie, indem man sie in Kontakt bringt. Anschließend steckt man beide Objekte in den Becher. Der Zeigerausschlag bleibt null, denn obwohl durch das Elektrisieren Ladungen getrennt wurden, ist die im Becher steckende Summe der getrennten Ladungen null. Wenn man den Glasstab herauszieht und dadurch eine positive Ladungsmenge q_0 entfernt, schlägt der Zeiger des Elektroskop auf einen bestimmten Wert aus, weil das Elektroskop durch die verbleibende negative Ladung $-q_0$ elektrisiert wird. Steckt man den Stab wieder zurück, geht der Ausschlag wieder auf null zurück. Nun kann man ja auch umgekehrt das Fell herausziehen und damit zugleich die daran haftende negative Ladung. Das führt wieder zu einem bestimmten Zeigerausschlag. Der interessante Punkt ist nun, dass beide Zeigerausschläge gleich sind, wenn man das Experiment ausreichend perfekt durchführt, d.h. man stellt fest, dass die Zeigerausschläge für gleich große positive wie negative Ladung gleich groß sind.

Aus diesem Resultat kann man auch noch eine interessante Einsicht zur funktionalen Abhängigkeit des Potentials von der Ladung gewinnen: Zwischen einem Elektroskop, auf dem eine positive Ladung sitzt und einem Elektroskop, auf dem eine betragsmäßig gleich große negative Ladung sitzt, lässt sich kein Unterschied feststellen: Sie weisen die gleichen Zeigerausschläge auf. Das legt den Schluss nahe, dass auch die auf ihnen deponierte elektrische Energie gleich groß ist, d.h. wenn man die Ladung auf dem positiv geladenen Elektroskop um dq verändert und entsprechend die Ladung auf einem negativ geladenen Elektroskop um $-dq$, dann sind die Energieänderungen $d\mathcal{G}_+$ bzw. $d\mathcal{G}_-$ auf beiden Elektroskopen gleich. Aus

$$d\mathcal{G}_+ = \varphi(q)dq = d\mathcal{G}_- = \varphi(-q)(-dq)$$

folgt dann, dass das Potential eine ungerade Funktion der Ladung ist, d.h.

$$\varphi(q) = -\varphi(-q). \quad (35.4)$$

Ferner zeigt eine Prüfung, dass das elektrische Potential unabhängig von Druck und Temperatur der Umgebung ist, d.h. dass das Potential $\varphi = \varphi(q)$ allein eine Funktion der Ladung ist.

35.4. Thermoelektrische Energieumwandlungen

35.4.1. Richtungsvorgabe für den Ladungstransport

Wenn ein physikalisches System aus zwei Teilsystemen A und B besteht, so gilt für die Änderung der elektrischen Energie

$$d\mathcal{G} = \varphi_A dq_A + \varphi_B dq_B$$

und wegen der Ladungserhaltung muss noch gelten

$$dq_A = -dq_B.$$

Da für spontan ablaufende Prozesse $d\mathcal{G} < 0$ sein muss, ist die Abgabe einer positiven Ladung dq_A an das Teilsystem B nur dann möglich, wenn das Potential φ_A höher ist als φ_B , d.h. wenn die als ELEKTRISCHE SPANNUNG

$$U = U_{AB} = \varphi_A - \varphi_B \quad (35.5)$$

bezeichnete Potentialdifferenz zwischen A und B positiv ist. Die Tatsache, dass positive Ladungen nur vom höheren Potential zum niedrigeren Potential abgeschoben werden können und negative Ladungen nur vom niedrigeren Potential zum höheren, ist also letztendlich dem zweiten Hauptsatz der Thermodynamik geschuldet, der keine Entropieerniedrigung duldet.

Wenn man drei Systeme mit den Potentialen $\varphi_A > \varphi_B > \varphi_C$ vorliegen hat und eine positive Ladung dq von A nach B nach C transportiert wird, so wird im ersten Schritt die elektrische Energie $d\mathcal{G}_{AB} = U_{AB}dq$ abgegeben und im zweiten Schritt die elektrische Energie $d\mathcal{G}_{BC} = U_{BC}dq$ und somit insgesamt die Energie $d\mathcal{G}_{AC} = d\mathcal{G}_{AB} + d\mathcal{G}_{BC}$. Das ist gleichbedeutend mit der Feststellung, dass sich die elektrischen Spannungen addieren:

$$U_{AC} = U_{AB} + U_{BC}.$$

35.4.2. Potentialausgleich

Im elektrischen Gleichgewicht ist $d\mathcal{G} = 0$ und somit

$$\varphi_A = \varphi_B. \quad (35.6)$$

Thermodynamische Prozesse laufen demnach so lange ab, bis ein POTENTIALAUSGLEICH erreicht ist und eine anfängliche Spannung U zwischen A und B auf null gefallen ist.

Im elektrischen Gleichgewicht sind alle miteinander wechselwirkenden Körper auf gleichem elektrischen Potential.

Beispiel: A und B seien zwei gleich aufgebaute Elektroskope. Zu Anfang sei B ungeladen (Zeigerausschlag null), während auf A eine positive Ladung q_0 aufgebracht wird, was einen entsprechenden Zeigerausschlag hervorruft. Bringt man die Elektroskope miteinander in Kontakt, erkennt man an der Abnahme des Zeigerausschlags von A und der Zunahme des Zeigerausschlags von B, dass ein thermodynamischer Prozess abläuft. Am Ende wird der Zeigerausschlag beider Elektroskope gleich hoch sein. Das bedeutet, dass sich auf beiden Geräten die gleiche Ladung, nämlich jeweils $q_0/2$, und die gleiche elektrische Energie befindet sowie dass beide Elektroskope gleiches Potential haben. Beide Elektrometer befinden sich im gleichen Zustand.

Die Summe der beiden Gibbs-Energien nach dem Prozess ist allerdings kleiner als die Summe vor dem Prozess. Die Differenz wurde in Form von Arbeit und/oder Wärme nach außen abgegeben. Bei dieser Betrachtung ist zu beachten, dass Druck und Temperatur mit der Umgebung im Gleichgewicht sind, d.h. wenn die elektrische Energie im Systeminneren dissipiert wird, wird sie durch den thermischen Kontakt in Form von Wärme nach außen geführt.

Die Geschwindigkeit der Aufnahme bzw. Abgabe von Ladungen durch ein Teilsystem kann man in weiten Grenzen kontrollieren: Man kann den Potentialausgleich sehr rasch ablaufen lassen oder auch extrem langsam. Es ist sogar nahezu möglich, ihn kinetisch nahezu zu blockieren, d.h. so weit hinauszögern, dass innerhalb der experimentellen Beobachtungszeit sogar keine Potentialänderung beobachtet wird. Dann spricht man von ELEKTRISCHER ISOLATION. Wenn ein geladenes Elektroskop beispielsweise von Luft umgeben ist, so ist es hervorragend gegen einen Ladungsverlust isoliert und kann somit auch nichts von seiner elektrischen Energie abgeben.

Man kann aber auch den Kontakt zwischen zwei Körpern mit einer Spannung $U \neq 0$ herstellen, dass U sehr rasch auf null fällt. Das geschieht beispielsweise, wenn ein metallischer Kontakt hergestellt wird. Die Geschwindigkeit hängt u.A. von den Eigenschaften des Mediums zwischen den Teilsystemen und der Spannung ab. Nähere Einzelheiten zum Ladungstransport und damit zu den kinetischen Aspekten werden im Teil 35.7.4 untersucht. In diesem Teil soll die Untersuchung der ELEKTROSTATIK im Vordergrund stehen. Interessant ist noch der Fall, wo sich zu Anfang auf zwei Elektroskopen A und B eine betragsmäßig gleich große Ladung $q_A = -q_B$ entgegengesetzten Vorzeichens befindet. Die Potentials sind dann entgegengesetzt gleich, d.h. $\varphi_A = -\varphi_B$, und beide Instrumente zeigen den gleichen Zeigerausschlag an. Auch die Potentiale sind entgegengesetzt gleich: $\varphi_A = -\varphi_B$. Lässt man beispielsweise durch einen metallischen Kontakt einen Austausch von Ladungen zu, so fällt die Spannung auf null und die Ladungen kompensieren sich dabei vollständig zu null. Die elektrische Energie, die auf den beiden entgegengesetzt geladenen Elektrometern zu Anfang gespeichert war, „verpufft“ bei diesem Prozess vollständig, indem sie z.B. nach außen als Wärme abgegeben wird. Im Extremfall können die entgegengesetzten Ladungen gewissermaßen zu Wärme zerstrahlen.

35.4.3. Elektrische Heizung

Innerhalb einer elektrischen Heizung, beispielsweise einem Tauchsieder, ändert sich die innere Energie nicht:

$$dU = T dS + \varphi_+ dq_+ + \varphi_- dq_- = T dS + U dq_+ = 0. \quad (35.7)$$

Beim elektrischen Umsatz einer positiven Ladung dq_+ auf dem Potential φ_+ mit einer negativen Ladung $dq_- = -dq_+$ auf dem Potential φ_- zur Ladung null verschwindet die elektrische Energie $U dq_+$ durch Rekombination der Ladungen (dq_+ nimmt dabei auf null ab) und erhöht die thermische Energie um $T dS$. Da die Temperatur eine monoton steigende Funktion der Temperatur ist, erhöht sich mit der Entropiezunahme zugleich

die Temperatur des Tauchsieders. Bei einem Tauchsieder, der beidseitig an eine elektrische Ladungsquelle angeschlossen ist, wird die durch Rekombination „verbrauchte“ positive bzw. negative Ladung fortlaufend nachgeliefert, so dass durch den Tauchsieder permanent elektrische Energie in thermische Energie umgesetzt wird. Diese Umsetzung geschieht in der lokalen Umgebung der Rekombination und daher erhöht sich auf die Temperatur quasi-lokal. Der umgekehrte Vorgang, dass nämlich die Entropie im Tauchsieder abnimmt und dafür elektrische Energie produziert wird, also positive und negative Ladungen auf unterschiedlichem Potential erzeugt werden, ist durch den zweiten Hauptsatz ausgeschlossen.

35.4.4. Thermoelektrische Generatoren und Thermoelemente

Ähnlich man durch isotherme Wärmezufuhr von außen eine Druckerhöhung bewirken und damit Arbeit gewinnen kann, kann man, allerdings nur in bestimmten Materialien, durch Wärmezufuhr eine Ladungstrennung bewirken.

und damit eine elektrische Spannung. Dieses Phänomen heißt PYROELEKTRIZITÄT und kann zur Erzeugung elektrischer Energie aus Wärmezufuhr herangezogen werden. Pyroelektrische Sensoren finden beispielsweise technische Anwendung bei Bewegungsmeldern. Mit diesem Phänomen kann man aber wegen des zweiten Hauptsatzes genausowenig wie in Wärmekraftmaschinen eine fortdauernde bzw. periodische Umwandlung von Wärme in elektrische Energie bzw. Arbeit erzielen. Das elektrische Pendant zu den Wärmekraftmaschinen, (Kap. 32) sind die THERMOELEKTRISCHEN GENERATOREN. Mit ihnen kann fortlaufend elektrische Energie aus Wärmeenergie gewonnen werden, indem Wärme aus einem Reservoir A mit höherer Temperatur T_A so aufgenommen und an ein Reservoir B bei der Temperatur T_B abgegeben wird, dass ein Kreisprozess stattfindet, bei dem keine Entropie im Generator verbleibt. Ein einfaches Beispiel dafür ist ein THERMOELEMENT. Es besteht aus zwei unterschiedlichen metallischen Materialien, die miteinander in engem Kontakt sind, beispielsweise miteinander verschweißt sind. Hält man eine Kontaktstelle A auf der Temperatur T_A und die andere auf der Temperatur T_B , so tritt zwischen den Enden eine elektrische Spannung auf (Thermospannung, Seebeck-Effekt) und man kann elektrische Energie entnehmen. Zugleich wird Wärme vom Bad mit höherer Temperatur zum Bad mit niedrigerer Temperatur transportiert. Da die Wärmebäder aus elektrischen Gründen metallisch verbunden sind, sind sie thermisch nur schlecht voneinander isoliert. Daher sind die Wirkungsgrade thermoelektrischer Generatoren deutlich kleiner als der Carnot'sche Wirkungsgrad. Dennoch werden sie unter bestimmten Umständen eingesetzt. In der Raumfahrt wird beispielsweise die elektrische Energie für Sonden, die außerhalb des Sonnensystems operieren, durch thermoelektrische Generatoren erzeugt, bei denen die höhere Temperatur durch den Zerfall eines radioaktiven Mediums gehalten wird.

Führt man umgekehrt einem thermoelektrischen Generator elektrische Energie zu, so kann Wärme von einem Wärmereservoir niedrigerer Temperatur an eines bei höherer Temperatur transportiert werden, d.h. er arbeitet als Wärmepumpe (Peltier-Effekt). Solche Peltier-Elemente kann man z.B. zur Kühlung von Prozessoren in Computern einsetzen. Die Thermospannung eines Thermoelements kann man für die Messung der

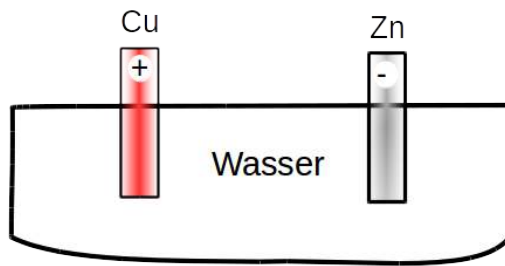


Abbildung 54: Galvanische Zelle: Zwischen einer in Wasser eintauchenden Kupfer- und Zinkelektrode tritt eine elektrische Spannung auf.

Temperatur T_A eines Systems A relativ zur Temperatur eines Systems B heranziehen. Kontaktstelle B kann z.B. in Wasser am Tripelpunkt eintauchen.

35.5. Elektrochemische Energieumwandlungen

Um elektrochemische Energieumwandlungen zu verstehen, müssen einige Begriffe vorausgeschickt werden, die thematisch eigentlich zu den Transportphänomenen gehören (Teil 35.7.4). Der Transport von Ladungen zwischen zwei Systemen kann auf zwei Arten erfolgen, nämlich verbunden mit einem chemischen Stofftransport, dann heißt er ELEKTROLYTISCH, oder ohne chemischen Stofftransport, dann heißt er ELEKTRONISCH. Elektrolytischen Ladungstransport beobachtet man z.B. in Säuren, Basen, Salzlösungen oder Salzschnmelzen. Diese Leiter bezeichnet man als ELEKTROLYTE. Elektronische Ladung ist Ladung, die nicht an einen chemischen Stoff gebunden ist. Der Ladungstransport in Metallen erfolgt ausschließlich durch elektronischen Ladungstransport. Metalloberflächen wirken sich im Elektrolyt wie eine semipermeable Membran aus, welche für elektronische, nicht aber für elektrolytische Ladung durchlässig ist. Elektronische Leiter mit dieser Eigenschaft der Grenzschicht bezeichnet man als ELEKTRODEN.

35.5.1. Galvanische Zelle

Die Umwandlung von chemischer Energie in elektrische Energie geschieht in Galvanischen Zellen, BATTERIEN, AKKUMULATOREN oder BRENNSTOFFZELLEN. Dabei laufen elektrochemische Reaktionen ab, d.h. Reduktions- oder Oxidationsvorgänge.

Hier soll nur das Grundprinzip an einem möglichst einfachen Beispiel besprochen werden: Eine galvanische Zelle erhält man, indem man zwei chemisch unterschiedliche Metallstäbe in ein Medium steckt, das als Elektrolyt fungieren kann. Man kann sie beispielsweise dadurch realisieren, dass man ein Kupferblech (Cu) und ein Zinkblech (Zn) einfach in Wasser steckt (Abb. 54) [12]. An jeder der beiden Elektrodenoberflächen findet eine elektrochemische Reaktion statt, die zu einem elektrochemischen Gleichgewicht führt. Damit liegt zwischen Elektrolyt (hier das Wasser) und Elektrode jeweils eine für den elektronischen Leiter Cu bzw. Zn charakteristische Potentialdifferenz vor. Da der Elektrolyt dafür

sorgt, dass beiden Seiten praktisch auf gleichem Potential sind, führt das schlussendlich zu einer Spannungsdifferenz zwischen den Elektroden – vorausgesetzt, dass die elektronischen Leiter aus chemisch unterschiedlichen Materialien bestehen. Verbindet man die beiden Elektroden mit einem Draht, so dass ein elektronischer Ladungstransport stattfinden kann, wird Zink oxidiert und an der Kupferelektrode entsteht durch Reduktion Wasserstoff.

Mehrere solcher galvanischen Zellen lassen sich durch Hintereinanderschaltung zu einer Batterie kaskadieren und ergeben so eine leistungsstarke Ladungsquelle. Galvanische Zellen, an denen ein technisch nicht umkehrbarer Vorgang abläuft, nennt man PRIMÄRZELLEN. Eine galvanische Zelle, bei der die Umwandlung von elektrischer Energie in chemische Energie bidirektional vonstatten gehen kann, nennt man eine SEKUNDÄRZELLE bzw. einen AKKUMULATOR. Die Elektrode, bei welcher eine Reduktion stattfindet, ist per definitionem die Kathode und diejenige, bei welcher eine Oxidation stattfindet, ist die Anode. Daher wechseln Kathode und Anode die Seiten, wenn man vom Ladevorgang zum Entladevorgang des Akkumulators übergeht.

35.5.2. Elektrolyse

Mit der in Abb. 55 gezeigten Apparatur kann man die ELEKTROLYSE von Wasser beobachten, d.h. seine chemische Zersetzung. Die beiden Elektroden können aus dem chemisch gleichen Metall bestehen. An Kontaktflächen zwischen dem Wasser und der metallischen Zuführung kommt es zu einer chemischen Reaktion. Das an der positiven Elektrode gebildete Gas kann man als Sauerstoff und das an der negativen Elektrode gebildete Gas als Wasserstoff identifizieren. Die beiden chemischen Produkte kann man z.B. wieder miteinander reagieren lassen und mit der dabei frei werdenden chemischen Energie z.B. heizen oder einen Verbrennungsmotor laufen lassen. Kurzum: bei der Elektrolyse wird also elektrische Energie in chemische Energie umgewandelt. Da Wasser das chemische Produkt einer Verbrennung bzw. Oxidation des Wasserstoffs ist, an der negativ angeschlossenen Elektrode aber der umgekehrte Vorgang stattfindet, wird der Wasserstoff dort reduziert. Die Elektrode, an der Reduktion stattfindet, heißt KATHODE. Die positiv angeschlossene Gegenelektrode, an welcher der Oxidationsvorgang abläuft, heißt ANODE. Diese Bezeichnungen für die Elektroden wurden von Michael Faraday eingeführt, der als einer der ersten diesen unerwarteten Zusammenhang zwischen Chemie und Elektrizität untersuchte. Faraday entdeckte bei der Abscheidung von Stoffen an den Elektroden, dass deren Menge bzw. Masse proportional zur transportierten Ladungsmenge zunimmt. Das kann man bei der Elektrolyse des Wassers gut beobachten, wo die gebildete Gasmenge mit der transportierten Ladungsmenge q zunimmt, d.h. der chemische Umsatz ist proportional zur Ladungsmenge.³⁰

Ähnliche Beobachtungen macht man, wenn man Metallstäbe in eine Salzsäure oder eine wässrige Salzlösung eintaucht und eine bestimmte Ladungsmenge hindurchfließen

³⁰Da man durch Verbrennen von Wasserstoff und Sauerstoff nutzbare Energie gewinnen kann, muss diese elektrisch aufgebracht worden sein. Daraus folgt, dass der Vorgang erst abläuft, wenn eine bestimmte Mindestspannung an den Elektroden überwunden wird.



Apparatur zur Untersuchung der Elektrolyse von Wasser. Wenn man Wasser ein wenig ansäuert, wird es zu einem guten Elektrolyten. Schließt man die in das Wasser eintauchenden Metallelektroden an eine Gleichstromquelle an, kann man an den Elektroden Gasbläschen aufsteigen sehen. An der positiven Elektrode (rotes Kabel) entsteht Sauerstoff, an der negativen (schwarzes Kabel) Wasserstoff. Den Wasserstoff identifiziert man z.B. durch die Knallgasprobe, den Sauerstoff durch das Aufflammen eines glimmenden Holzspans. Die Gase entstehen im H_2/O_2 -Volumensverhältnis von 2 : 1.

Abbildung 55: Elektrolyse von Wasser.

lässt. An den Elektroden läuft wieder eine REDOX-REAKTION³¹ ab. Leitet man beispielsweise durch eine Silbernitratlösung einen elektrischen Strom, dann wird an der Kathode Silber abgeschieden.

35.6. Elektrische Maßeinheiten

In diesem Abschnitt wird das alte auf der Basiseinheit Coulomb beruhende ELEKTRISCHE SI-SYSTEM als provisorisches Einheitensystem für elektrische Erscheinungen festgelegt.

Basiseinheit Coulomb. Wenn man eine Ladung q über eine Silbernitratlösung entlädt, dann wird an der Kathode Silber abgeschieden. Die abgeschiedene Stoffmenge des Silbers ist der abgegebenen Ladungsmenge proportional. Durch diesen Zusammenhang kann man die Einheit der Ladung definieren, die man als COULOMB bezeichnet:

$$[q] = 1 \text{ Coulomb} = 1 \text{ C} .$$

³¹Reduktions-Oxidations-Reaktion

Ein Coulomb ist diejenige Ladungsmenge, bei der eine Menge von 1.118 mg Silber aus einer Silbernitratlösung abgeschieden wird.

Die hier angegebene Definition der Einheit Coulomb war bis 1948 in Gebrauch. Sie soll vorläufig im Rahmen einer konsistenten Einführung in die Elektrizitätslehre akzeptiert sein. Die heute grundlegende SI-Einheit ist jedoch das Ampere. Die Gründe für diese Wahl wird aber erst verständlich, wenn man sich auch einen Überblick über die magnetischen Erscheinungen verschafft hat und zu einem Verständnis der Vereinigung der elektrischen und der magnetischen Theorie zur Theorie des Elektromagnetismus vorgestoßen ist (s. „*Romanos Physikkurs 2*“).

Coulombmeter. Das Messverfahren einer Ladungsmenge durch Wägung der an der Kathode abgeschiedenen Silbermenge stellt zugleich ein COULOMBMETER dar, also ein Messgerät für Ladungsmengen. Mit diesem kann man auf anderen Prinzipien beruhende Coulombmeter kalibrieren. Beispielsweise könnte man die Anzeige eines Elektrometers entsprechend kalibrieren und es so zu einem Coulombmeter für die auf ihm gespeicherte Ladung machen. Dazu bräuchte man nur das auf einen bestimmten Wert geladene Elektrometer, dessen Ladungsmenge qualitativ durch die Zeigeranzeige gegeben ist, jeweils über eine Silbernitratlösung zu entladen und die abgeschiedene Silbermenge bestimmen.³²

Amperemeter. Wenn man ein Coulombmeter mit einer Uhr ausstattet und die dort gemessene Ladungsmenge auf die Zeiteinheit bezieht, hat man auch ein primäres Strommessgerät bzw. AMPEREMETER zur Verfügung. Schaltet man ein zweites Strommessgerät in den Stromkreis hinein (SERIENSCHALTUNG), so kann man dieses mit dem ersteren kalibrieren, denn sowohl durch das erste wie das zweite Amperemeter fließt der gleiche Strom I . Ausgehend von einem primärem Amperemeter kann man also jedes auf anderen Prinzipien beruhende Amperemeter kalibrieren und daher soll ab hier davon ausgegangen werden, dass Amperemeter zur Verfügung stehen.

Volt. Die Einheit der Spannung heißt VOLT:

$$[U] = 1 \text{ Volt} = 1 \text{ V} = 1 \text{ J/C}.$$

Ein primäres Messprinzip für die Spannung U einer Spannungsquelle lässt sich nach Gl. 35.2 folgendermaßen etablieren: Man befüllt einen thermisch gut isolierenden Dewar mit einer Wassermenge der Masse m der spezifischen Wärmekapazität c_p und stellt einen Tauchsieder hinein. Diesen schließt man an die Spannungsquelle an und bestimmt die von der Spannungsquelle abgegebene Ladungsmenge Δq . Bei diesem Vorgang wird die elektrische Energie $U\Delta q$ vollständig in eine Enthalpieänderung des Wassers umgesetzt, welche gleich derjenigen ist, die einer Wärmezufuhr von $mc_p\Delta T$ entspricht. Hierbei ist ΔT die Temperaturerhöhung des mit dem Tauchsieder erwärmten Wassers. Setzt man die elektrische Energiezufuhr und die

³²Das ist ein Gedankenexperiment, um das Prinzip zu etablieren, kein praktisches Verfahren.

Wärmezufuhr gleich, ergibt sich die Spannung gemäß

$$U = \frac{mc_p \Delta T}{\Delta q}.$$

Voltmeter. Mit dem gerade geschilderten Messprinzip ist ein primäres Voltmeter gegeben. Durch PARALLELSCHALTEN eines primären Voltmeters und eines zweiten zur Spannungsmessung geeigneten Anzeigeelements kann man letzteres kalibrieren und so VOLTMETER realisieren, die auf anderen Prinzipien beruhen. Beispielsweise kann man das mit einem Elektroskop (Abb. 53) tun. Dieses wird damit zu einem ELEKTROSTATISCHEN VOLTMETER. Meist hat man den Wunsch, Voltmeter so zu bauen, dass sie die zu messende Spannung möglichst wenig beeinflussen (s. Kap. 35.7.3).

Sobald erst einmal ein primäres Messgerät für eine elektrische Messung zur Verfügung hat, ist die Kalibrierung jedes anderen elektrischen Messgeräts möglich. Die Messprinzipien der Geräte und die Kalibrationsmethoden unterscheiden sich in der Genauigkeit, mit der sie durchgeführt werden können. Sie beruhen oft auf ausgeklügelten messtechnischen Überlegungen. Darauf soll hier nicht weiter eingegangen und auf Speziallehrbücher dazu verwiesen werden. Die Messverfahren, die in diesem Kapitel exemplarisch diskutiert wurden, sind weit von der messtechnischen Eleganz käuflich erwerbbarer Coulomb-, Ampere- und Voltmeter entfernt. Sie wurden hier nur geschildert, um das prinzipielle Vorgehen zu erläutern und eine Ausgangsbasis für das Verständnis der Elektrizitätslehre bieten. Für welche technische Lösung, d.h. für welches Messgerät, man sich bei einem konkreten Messproblem dann tatsächlich entscheidet, kann von vielen Faktoren abhängen, z.B. von der Zweckmäßigkeit, der Bequemlichkeit, der erforderlichen Genauigkeit oder den Kosten. Wir wollen ab hier einfach davon ausgehen, dass uns geeignete Messgeräte für die grundlegenden Messungen von Ladung, Strom und Spannung im Labor zur Verfügung stehen.

35.7. Kapazität

Gemäß Gl. 35.2 ist das Potential $\varphi = \varphi(q)$ eines thermodynamischen Systems eine Funktion der auf ihm vorhandenen Ladung q . Wir entwickeln die Umkehrfunktion $q = q(\varphi)$ in eine Potenzreihe und beachten Gl. 35.4:

$$q(\varphi) = C\varphi + D\varphi^3 + \dots$$

mit Koeffizienten C, D, \dots . Empirisch zeigt sich, dass für metallische Leiter in Luft nur der erste Koeffizient von null verschieden ist.³³ Ladung und Potential sind dann proportional zueinander:

$$q = C\varphi \tag{35.8}$$

³³Von Spezialfällen bei denen nichtlineare Materialeigenschaften eine Rolle spielen und somit der Koeffizient $D \neq 0$ ist, wollen wir zunächst absehen.

Die Proportionalitätskonstante heißt ELEKTRISCHE KAPAZITÄT³⁴ und ist grundsätzlich eine positive Größe. Ihre physikalische Einheit heißt FARAD:³⁵

$$[C] = 1 \text{ Farad} = 1 \text{ F} = 1 \text{ C/V}$$

35.7.1. Metallkugel

Die einfachsten Untersuchungsobjekte für die Bestimmung der Kapazität sind Metallkugeln im Vakuum. Wegen der Kugelsymmetrie kann das Resultat nur vom Kugelradius r_K abhängen. Man bringt eine Ladung q auf, misst die Spannung U gegen Erde und berechnet die Kapazität $C = q/U$.

Wir verwenden dabei eine Metallkugel, die aus einer dünnen Kugelschale besteht, die sich öffnen und mit verschiedenen Materialien, z.B. ein Kupferdrahtknäuel oder Metallkügelchen befüllen lässt. Dabei stellt sich verblüffenderweise heraus, dass es bei gegebenem Radius r_K völlig egal ist, womit sie gefüllt ist. Es ist egal, ob es sich um eine Vollkugel handelt oder um eine beliebig dünne Kugelschale. Daraus ist zu schließen, dass sich die Ladung nur auf der Kugeloberfläche $4\pi r_K^2$ befindet. Jegliches im Inneren befindliche metallisch leitende Material spielt keine Rolle. Die Kapazität eines metallischen Objekts wächst also einerseits proportional zu seiner Oberfläche an. Andererseits findet man empirisch, dass die Kapazität proportional zum Radius ist (Abb. 56). Das legt nahe, dass die Kapazität proportional zum Verhältnis von Kugeloberfläche und Kugelradius ist:

$$C = \epsilon \frac{4\pi r_K^2}{r_K} = 4\pi\epsilon r_K \quad (35.9)$$

Die Proportionalitätskonstante ϵ bezeichnet man als DIELEKTRIZITÄTSKONSTANTE. Sie hängt vom Material im Außenraum der Kugel ab. Das weist auf eine Wechselwirkung zwischen der Kugel und der Materie im Außenraum hin, die noch zu untersuchen ist.

Für verdünnte Gase erweist sich ϵ als eine Materialkonstante. Aus Messungen erhält man

$$\epsilon = \frac{q}{4\pi r_K \varphi} \approx 8.9 \cdot 10^{-12} \text{ F/m}. \quad (35.10)$$

Diesen Wert kann man auch für Luft unter Atmosphärendruck verwenden. Das Potential

$$\varphi = \frac{1}{4\pi\epsilon r_K} q$$

einer mit q geladenen Kugel ist umgekehrt proportional zum Kugelradius r_K . Wenn man den Radius größer und größer macht, geht das Potential gegen Null. Macht man den Radius hingegen immer kleiner, so wird das Potential schließlich unendlich groß. Die Referenzspannung $U = 0$ ist hier also die Spannung eines sehr großen (genau genommen unendlich großen) Kugelkondensators. Dem kommt die Erdkugel recht nahe.

³⁴Analog zur Bezeichnung Wärmekapazität (s. Kap. 28.3).

³⁵Vorsicht! Das Symbol C für die physikalische Größe Kapazität und das Symbol C für die Einheit Coulomb.

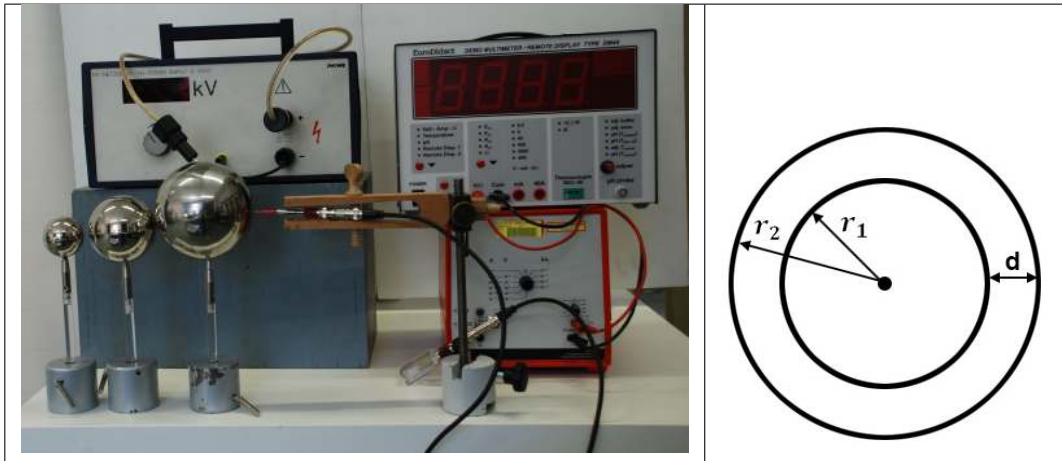


Abbildung 56: (a) Kugeln mit unterschiedlichem Radius r_K werden bei gleicher angelegter Spannung U aufgeladen. Anschließend wird mit einem Coulombmeter die Ladungsmenge q bestimmt. Trägt man q gegen r_K auf, erhält man eine Gerade, d.h. empirisch ergibt sich der Zusammenhang $C \propto r_K$. (b) Skizze zum Kugelkondensator.

35.7.2. Kondensator

Ein KONDENSATOR besteht aus zwei leitenden Flächen, die sich in einem isolierenden Medium in einem gewissen Abstand voneinander gegenüberstehen.

Uns interessiert der häufige Fall, bei dem auf der einen Leiterfläche die Ladung $q = +q_0$ deponiert ist und auf der anderen Leiterfläche die betragsmäßig gleich große Ladung $q = -q_0$. Die Kapazität eines Kondensators ist grundsätzlich eine positive Größe. Sie ist definiert durch das Verhältnis des Betrags der Ladung auf einer der beiden Platten zum Betrag der Spannungsdifferenz.

Wiederum wählen wir den einfachen Fall zweier kugelförmiger leitender Flächen mit den Radien r_1 bzw. r_2 , von denen die eine, wie in Abb. 56b dargestellt, sich innerhalb der anderen befindet. Dies stellt einen KUGELKONDENSATOR dar. Auf den ersten mit Radius r_1 bringen wir die positive Ladung $q = +q_0$ und auf den zweiten mit dem Radius r_2 die Ladung $q = -q_0$.

Ich möchte nun den Fall des aus zwei Kugelflächen bestehenden Kugelkondensators auf den im letzten Abschnitt diskutierten Fall einer einzelnen Kugelfläche zurückführen. Läge die positiv geladene Kugel alleine vor, dann wäre das Potential auf diesem ersten Leiter

$$\varphi_1 = +q_0/C_1 = q_0/4\pi\epsilon r_1$$

und im umgekehrten Fall gilt

$$\varphi_2 = -q_0/C_2 = -q_0/4\pi\epsilon r_2$$

Die Spannung $U = U_{12}$ zwischen den beiden Flächen des Kugelkondensators ist so definiert, dass die Energieänderung gerade $dE = Udq$ beträgt, wenn man eine infinitesimale Ladungsmenge von der einen zur anderen Fläche überführt. Wenn man einen der beiden Radien gegen unendlich gehen lässt, wird dessen Beitrag Null und die Spannungsdifferenz wird $U = \varphi_1$ bzw. $U = \varphi_2$. Man hat noch einen weiteren Grenzfall: Wenn $r_1 = r_2$, d.h. wenn die beiden geladenen Schichten aufeinanderfallen und sich die Ladungen kompensieren, ist $U = 0$. Diese Grenzfälle sind konsistent erfasst, wenn die Spannungsdifferenz gleich der Summe der Potentiale des Falles einer einzigen Kugel ist, also

$$U = \varphi_1 + \varphi_2$$

Damit erhält man die Kapazität gemäß

$$\frac{1}{C} = \frac{U}{q_0} = \left| \frac{1}{C_1} - \frac{1}{C_2} \right| = \frac{1}{4\pi\epsilon} \left| \frac{1}{r_1} - \frac{1}{r_2} \right| = \frac{1}{4\pi\epsilon} \frac{|r_2 - r_1|}{r_1 r_2} = \frac{1}{\epsilon} \frac{d}{4\pi r^2}$$

Hier ist $d = |r_2 - r_1|$ der Abstand der Kondensatorplatten und $r = \sqrt{r_1 r_2}$ ist das geometrische Mittel der beiden Kugelradien. Die Kapazität eines Kugelkondensators mit der Kugeloberfläche $\mathcal{A} = 4\pi r^2$ ist also

$$C = \frac{q}{U} = \epsilon \frac{\mathcal{A}}{d} \quad (35.11)$$

Wenn man r sehr groß macht, und eine Teilfläche $\Delta\mathcal{A}$ von \mathcal{A} betrachtet, so stellt dieses Flächenstück $\Delta\mathcal{A}$ einen FLÄCHENKONDENSATOR dar. Gl. Wirkliche Flächenkondensatoren sind jedoch endlich und haben einen Rand. Am Rand entsprechen sie nicht mehr dem Modell, aus dem wir Gl. 35.11 hergeleitet haben. Daher kann man Gl. 35.11 nur so lange als gute Näherung für einen Flächenkondensator heranziehen, wie das Randgebiet gegenüber dem Flächenbereich vernachlässigt werden kann. Wenn man einen Flächenkondensator zu einem ZYLINDERKONDENSATOR verformt, dann gilt Gl. 35.11 immer noch für jedes Teilstück, das klein genug ist, um es als Flächenkondensator anzusehen, und somit auch für den Zylinderkondensator. Kurzum: Gl. 35.11 gilt für Kugelkondensatoren exakt und für Flächen- und Zylinderkondensatoren in guter Näherung, wenn die Bedingung $d \ll \sqrt{\mathcal{A}}$ erfüllt ist.³⁶

Wenn man zwei Kondensatoren mit den Kapazitäten C_1 und C_2 parallel schaltet, dann ist die zwischen ihren Endpunkten liegende Spannung gleich. Die Gesamtladung ist

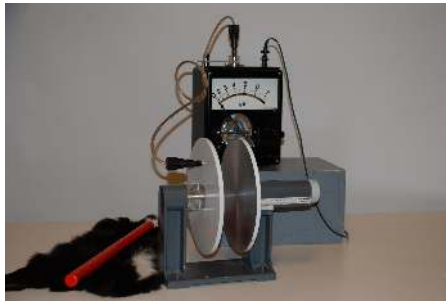
$$q = q_1 + q_2 = C_1 U + C_2 U$$

Für die Gesamtkapazität gilt daher:

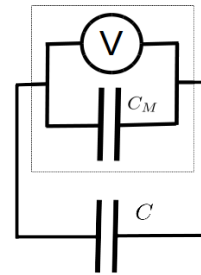
$$C = q/U = C_1 + C_2 \quad (35.12)$$

Wenn man hingegen zwei Kondensatoren in Serie schaltet, addieren sich hingegen die Spannungen:

³⁶Die exakte Formel für den Zylinderkondensator lautet $C = 2\pi\epsilon l / \ln(r_2/r_1)$, wobei l die Länge des Kondensators ist. Für kleines $d = r_2 - r_1 \ll r \approx r_1 \approx r_2$ ergibt sich Gl. 35.11.

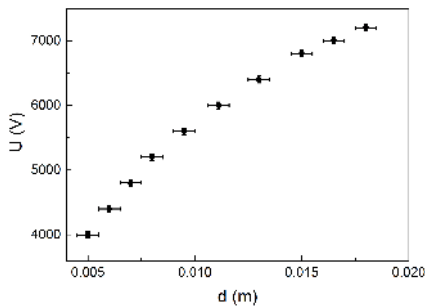


(a) Experimenteller Aufbau

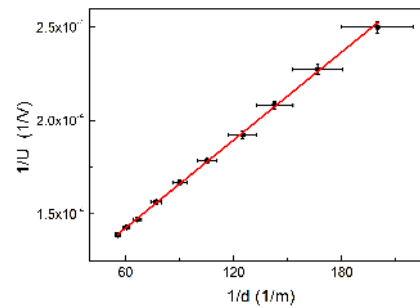


(b) Zugehöriges Schaltbild

Abbildung 57: Experimentelle Prüfung von Gl. 35.11 .



(a) Auftragung von U gegen d .



(b) Auftragung von $1/U$ gegen $1/d$.

Abbildung 58: Einfluss der Eigenkapazität eines elektrostatischen Voltmeters bei der Messung der Spannung.

$$U = U_1 + U_2 = q/C_1 + q/C_2$$

Für die Gesamtkapazität $C = Q/U$ folgt bei Serienschaltung daher

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2}. \quad (35.13)$$

35.7.3. Korrektur elektrischer Messungen.

Bei Messungen ist manchmal die Rückwirkung des Messinstruments auf das Messresultat nicht vernachlässigbar. Daher ist dies durch eine Korrektur zu berücksichtigen. Das Problem und seine Behebung soll anhand von Kapazitätsmessungen an einem Plattenkondensator exemplarisch erläutert werden.

Wenn man eine Ladung q auf einen Plattenkondensator aufgebracht hat und den Plattenabstand d verändert, sollte sich die Spannung U wegen Gl. 35.11 proportional zu d

verändern, d.h. man kann die Spannung eines Kondensators erhöhen, indem man den Plattenabstand vergrößert. Misst man U und trägt die Werte gegen d auf, dann sollte sich eine Gerade ergeben. Abb. 57a zeigt einen Plattenkondensator mit veränderbarem Plattenabstand und einem Voltmeter, mit dem man diese Konsequenz prüfen kann. Die in Abb. 58a gezeigten Messresultate zeigen aber, dass die Spannung mit d nicht linear, sondern sublinear ansteigt.

Der Grund hierfür ist die EIGENKAPAZITÄT eines elektrostatischen Voltmeters. Wie im Schaltbild gezeigt (Abb. 57b), kann man das reale Voltmeter durch ein Ersatzschaltbild (gepunktet in Abb. 57b) modellieren, bei dem die idealisierte reine Messfunktion des Voltmeters (mit Eigenkapazität null) durch das Symbol für ein Voltmeter dargestellt wird und seine Eigenkapazität durch das dazu parallelgeschaltete Kondensatorsymbol mit der Kapazität C_M deutlich gemacht wird. Für die Spannung am Messgerät gilt daher:

$$\frac{q}{U} = C_M + C = C_M + \epsilon \frac{\mathcal{A}}{d} \quad (35.14)$$

So lange die Kapazität C_M des Messgerätes vernachlässigbar klein bleibt, gilt näherungsweise $U \propto d$. Mit $d \rightarrow \infty$ wird C jedoch vernachlässigbar klein und man erhält $U = q/C_M = \text{const.}$ Der in Abb. 58a skizzierte prinzipielle Verlauf von U als Funktion von d zeigt, wie das Messgerät die Messung am Plattenkondensator um so mehr beeinflusst, je kleiner dessen Kapazität mit zunehmendem Plattenabstand wird. Wenn man jedoch, wie in Abb. 58b gezeigt, $1/U$ als Funktion von $1/d$ aufträgt, erhält man, wie nach Gl. 35.14 eine Gerade. Aus dem Achsenabschnitt kann C_M/q und aus der Steigung $\epsilon\mathcal{A}/q$ entnehmen und bei bekannter Plattenfläche \mathcal{A} sowohl C_M als auch q ermitteln. Wird der Plattenabstand sehr groß, so kommt noch ein weiterer Aspekt ins Spiel: Bis wurde nämlich angenommen, dass der Rand des Kondensators keine wesentliche Rolle spielt. Mit zunehmendem d wird diese Annahme jedoch immer problematischer.

35.7.4. Elektrische Energie eines Kondensators

Die elektrische Energie E_{el} , die in einem Kondensator der Kapazität C gespeichert ist, wenn sich darauf eine Ladung q_0 befindet, erhält man, indem man die Energiebeträge $dE_{el} = Udq = qdq/C$ aufintegriert, die mit einer Ladungsmenge dq zugeführt wird:

$$E_{el} = \int_0^{q_0} qdq/C = \frac{1}{2}q_0^2/C = \frac{1}{2}q_0U = \frac{1}{2}CU^2. \quad (35.15)$$

Hier ist $U = q_0/C$ die Spannung am Kondensator. Aus der Sicht der Relativitätstheorie hat sich durch die Aufnahme elektrischer Energie auch die Masse des Kondensators erhöht, d.h. ein geladener Kondensator ist ein bisschen schwerer als ein ungeladener. Doch diese Massenzunahme ist im Rahmen der heutigen experimentellen Möglichkeiten genauso wenig nachweisbar, wie sie es für mechanische Energieänderungen ist.

Transportphänomene

36. Strom

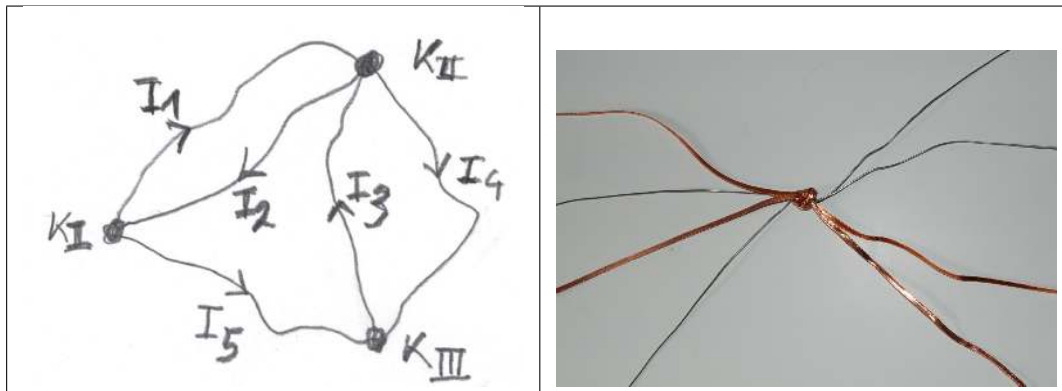


Abbildung 59: (a) Schematische Darstellung von Knoten, Pfaden und Strömen in einem Netzwerk. (b) Kirchhoffsche Knotenregel: Mehrere metallische Drähte sind durch einen Knoten leitend miteinander verbunden. Die Summe der über die Drähte zu- und abfließenden Ströme ist null.

Bis hier hin wurden die Grundzüge der Gleichgewichtsthermodynamik vorgestellt. Das Ziel dieses Kapitels ist es, einige Grundlagen der Nicht-Gleichgewichtsthermodynamik zu vermitteln. Die hier hauptsächlich zu klärenden Fragen sind, wie und mit welcher Geschwindigkeit der Übergang zum Gleichgewichtszustand abläuft. Der wesentlich neue Parameter, der hier ins Spiel kommt, ist die Zeit t .

Das einfachste Modellsystem, das man hierzu untersuchen kann, besteht aus einem aus einem abgeschlossenen Gesamtsystem, das aus zwei Teilsystemen besteht, die über einen Transportpfad miteinander verbunden sind. Die Teilsysteme sollen dabei zu Anfang nicht im Gleichgewicht sein, beispielsweise, weil sie eine elektrische Potentialdifferenz $\Delta\varphi$ oder eine Temperaturdifferenz ΔT aufweisen. Sobald man die Restriktionen beseitigt, die den Ausgleich bis da hin verhindert haben, dann treten Transportphänomene auf, die dazu führen, dass diese Differenzen verschwinden. Die Geschwindigkeit des Ausgleichsprozesses hängt von den Eigenschaften des Transportpfades ab. Hier sollen drei exemplarische Beispiele betrachtet werden: der ELEKTRISCHE LADUNGSTRANSPORT, der Wärmetransport und der Stofftransport.

Abb. 59a zeigt ein Netzwerk, das aus Knoten bzw. Teilsystemen K_I , K_{II} , ... besteht und verschiedenen Transportpfaden zwischen den Knoten, die technisch beispielsweise durch Straßen, Drähte, Eisenbahnschienen, Leiter, Rohre usw. realisiert sind.

Knoten K_I sei beispielsweise ein Elektroskop A, das geladen sei, und Knoten K_{II} ein anderes Elektroskop B, das ungeladen sei. Die Ladungszustände sind dann durch deren Zeigerstellungen unmittelbar ersichtlich. Man kann ein drittes mobiles Elektroskop T nehmen, den Träger. Mit ihm berührt man Elektroskop A. Dieses gibt Ladung ab, was unmittelbar an der abnehmenden Zeigerstellung zu erkennen ist, und T wird aufgeladen, denn seine Zeigerstellung nimmt zu. Dann trägt man T über irgendeinen Transportpfad

zu B und berührt das dortige Elektroskop. An den Zeigern wird unmittelbar ersichtlich, dass bei dem Vorgang insgesamt Ladung von A nach B transportiert worden ist. Ein materielles Objekt wie hier T, mit dem eine Ladung von A nach B getragen wird, bezeichnet man als LADUNGSTRÄGER. In unserem Beispiel wäre das Elektroskop als ein diskreter Ladungsträger anzusehen.

Im Rahmen der Vorstellung des Kontinuumsmodells der Materie kann der Ladungstransport aber auch kontinuierlich erfolgen. Wenn beispielsweise Wasser fließt, dann findet im Kontinuumsmodell ein kontinuierlicher Stofftransport statt. Wenn es sich um erwärmtes Wasser handelt, dann ist der Stofftransport mit einem Wärmetransport verknüpft und wenn das Wasser elektrisiert ist, also zugleich ein Ladungsträger ist, dann findet auch ein kontinuierlicher Ladungstransport statt.

Die einem System pro Zeiteinheit zufließende oder abfließende Stoff-, Wärme- oder Ladungsmenge stellt einen materiellen, thermischen oder elektrischen STROM dar. Wenn pro Zeiteinheit dt über die Transportpfade bestimmte Ladungsmengen dq , Wärmemengen dQ oder Stoffmengen dn oder Energiemengen (s. Kap. 12.6) von einem Knoten zum anderen fließen, werden extensive bzw. mengenartige Größen transportiert. Die sich im Transportpfad befindlichen Mengen können gewöhnlich in den Betrachtungen vernachlässigt werden. Dann ist die Summe der Mengen, die von allen Knoten aufgenommen (positiv) oder abgegeben (negativ) werden, null.

In elektrischen Netzwerken führt man zur Beschreibung des Transports die ELEKTRISCHE STROMSTÄRKE

$$I = -\frac{dq}{dt} \quad (36.1)$$

ein. Dabei ist dq die von einem Knoten abgegebene Ladungsmenge und I der aus dem Knoten herausfließende Ladungsstrom. Der Strom I ist positiv, wenn der Strom aus dem Knoten herausfließt und die Ladungsmenge im Knoten abnimmt, d.h. wenn dq/dt für den Knoten negativ ist. Analog definiert man den Strom eines Stoffes C durch

$$I_C = -\frac{dn_C}{dt},$$

wobei hier n_C die Menge des Stoffes C ist. Entsprechend ist

$$I_Q = -\frac{dQ}{dt},$$

der WÄRMESTROM.

Führen N Pfade mit den positiven oder negativen elektrischen Strombeiträgen I_i ($i = 1, \dots, N$) zu einem Knoten, so gilt für die Ladung q im Knoten die INTEGRALE KONTINUITÄTSGLEICHUNG

$$\frac{dq}{dt} + \sum_{i=1}^N I_i = 0. \quad (36.2)$$

Im zeitlich stationären Fall ist $\frac{dq}{dt} = 0$, d.h. die Ladung q bleibt dann zeitlich konstant, und es gilt

$$\sum_{i=1}^N I_i = 0, \quad (36.3)$$

d.h. die Summe aller Ströme ist null. Wenn die Strombilanz nicht ausgeglichen ist, nimmt die Ladung im Knoten ab oder sie nimmt zu. Der Knoten agiert im ersten Fall als Quelle und im zweiten Fall als Senke für die Ladung. Die Begriffe und Gesetze für andere Transportphänomene sind analog und eine Diskussion erübrigt sich daher.

Die zwischen der Zeit t_1 und t_2 zugeflossene Ladungsmenge Δq ergibt sich aus Gl. 36.1 durch Integration:

$$\Delta q = - \int_{t_1}^{t_2} I(t) dt \quad (36.4)$$

Analoge Begriffe kann man auch für die einem System zufließende Wärme oder Materie bilden. Beispielsweise ergibt sich die einem System zwischen der Zeit t_1 und t_2 zugeflossene Wärme ΔQ aus dem Wärmestrom I_Q gemäß

$$\Delta Q = - \int_{t_1}^{t_2} I_Q(t) dt \quad (36.5)$$

Beispiel 1: Die Drähte für verschiedene elektrische Bauelemente oder Geräte seien in einem Knotenpunkt miteinander verdrahtet (Abb. 59b). Für die zu- und abfließenden Ströme im Knotenpunkt gilt im stationären Fall dann die Kirchhoffsche Knotenregel, d.h. Gl. 36.3.

Beispiel 2: Der „Knotenpunkt“ habe eine Wärmekapazität $\Gamma = dQ/dT$ und es fließe über alle Pfade netto ein Wärmestrom $I_Q = -dQ/dt$ zu. Gemäß Gl. 36.2 gilt für die zugeführte Wärme

$$\frac{dQ}{dt} = \Gamma \frac{dT}{dt} = -I_Q$$

Folglich erhöht sich die Temperatur T mit der Rate

$$\frac{dT}{dt} = -I_Q/\Gamma.$$

Das geschieht um so schneller, je größer der Wärmezuffluss I_Q und je kleiner die Wärmekapazität Γ des Knotens ist.

Beispiel 3: Auf eine kalte Fensterscheibe fließe ein Strom von Luft, das gasförmiges Wasser enthält. Das Wasser kondensiert dort, d.h. es geht in die flüssige Phase über. Die Fensterscheibe ist hier der „Knoten“ und wirkt wegen der Umsetzung von gasförmigen zu flüssigem Wasser als Senke für das transportierte gasförmige Wasser. Der Partialstrom des gasförmigen Wassers (d.i. jener Anteil des Stromes, der sich spezifisch auf den Zustrom von Wasser und nicht Luft bezieht) sei $I(t)$. Für die in der Zeitspanne zwischen t_1 und t_2 auf der Fensterscheibe kondensierte Wassermenge m gilt gemäß Gl. 36.2

$$m = - \int_{t_1}^{t_2} I(t) dt$$

Beispiel 4: Durch einen Hahn fließe ein Wasserstrom I_1 in eine Badewanne und über einen Ausfluss ein Strom I_2 ab. So lange genauso viel zu- wie abfließt, bleibt (abgesehen von einer kleinen Durchflussmenge) kein Wasser in der Wanne. Wenn jedoch mehr zu- als abfließt, steigt der Wasserpegel. Damit die Wanne nicht überläuft, gibt es einen zweiten, höher gelegenen Ausfluss. Der sorgt für einen Strom I_3 , so dass $I_1 + I_2 + I_3 = 0$, d.h. die Wassermenge in der Wanne erreicht damit (hoffentlich) einen stationären Zustand.

37. Widerstand

Im vorhergehenden Kapitel wurden Ströme eingeführt; es wurde aber noch nicht dargestellt, warum sie auftreten. Ganz generell kann man sagen, dass sie genau dann auftreten, wenn sich zwei Teilsysteme bzw. „Knoten“ in einem Netzwerk hinsichtlich einer ihrer intensiven Größen unterscheiden. Ströme können also auftreten, wenn es beispielsweise zwischen einem Knotenpunkt K_I und K_{II} einen Temperaturunterschied oder einen elektrischen oder chemischen Potentialunterschied gibt. Dann befindet sich das aus beiden Teilsystemen bestehende Gesamtsystem nämlich nicht im thermodynamischen Gleichgewicht. Je nachdem, hinsichtlich welcher intensiven Größe sich die Teilsysteme unterscheiden, wird nämlich Energie frei, wenn ein Teil der dazu konjugierten extensive Größe von einem Knoten zum anderen transportiert wird. Die frei werdende Energie kann als Arbeit nach außen abgegeben oder dissipiert werden. Die Option der Abgabe von nutzbringender Arbeit besteht nur, wenn in den Transportpfad entsprechende Elemente (Motoren, Turbinen, usw.) eingebaut sind. Die Option der Dissipation besteht aber immer und in allen realen Systemen findet sie auch statt. Dadurch läuft das Gesamtsystem auf einen Zustand höherer Entropie und somit auf das thermodynamische Gleichgewicht zu. Für beide Optionen stellt sich am Ende ein Zustand ein, bei dem die intensiven Größen aller Knotenpunkte gleich sind. Die „Antriebskraft“ für den thermodynamischen Ausgleich ist daher naheliegenderweise die Differenz der intensiven Größen, d.h. die Ströme sind eine Funktion dieser Differenz.

Transporte erfolgen unterschiedlich schnell und manchmal scheinbar gar nicht. Die Materialgrenzfälle sind der IDEALE ISOLATOR und der IDEALE LEITER. Vorläufig wollen wir von der einfachen Vorstellung ausgehen, dass in elektrischen Leitern ein positives oder negatives elektrisches Fluid existiert und in einem idealen Isolator gar keins. Zu den wichtigsten Isolatoren gehört die Luft. Ohne ihre isolierenden Eigenschaften würden viele elektrostatischen Experimente nicht funktionieren. Wenn die Luftfeuchtigkeit zu hoch wird und die Luft dadurch besser leitet, schlagen auch tatsächlich manche Demonstrationsexperimente fehl. Sowohl für Wärme als auch für Elektrizität ist Vakuum der beste Isolator.

Wenn man beispielsweise adiabatische thermodynamische Prozesse untersuchen möchte (Kap. 28.1), so muss man dafür sorgen, dass das zu untersuchende System von einem guten Wärmeisolator umgeben ist. Eine entsprechende Rolle spielen elektrische Isolatoren für die Elektrostatik. Gase und evakuierte Zonen sind sowohl für Wärme als auch Elektrizität bemerkenswert gute Isolatoren. Wäre Luft nicht so ein extrem guter elektrischer

Isolator, könnte man die oben besprochenen elektrostatischen Experimente nicht so einfach vorführen. Bei einem Elektrometer sind beispielsweise Stab und Zeiger aus Metall, also einem Leiter, damit der elektrische Zustand rasch auf diese Elemente übertragen wird. Halterung bzw. Durchführung für den Metallstab und die Füße des Gehäuses sind aus isolierenden Materialien und sorgen mit der isolierenden Luft dafür, dass die auf das Instrument aufgebrachte Elektrizität nicht sofort wieder verloren geht.

In elektrischen Experimenten hängt die Kinetik des Vorgangs auch davon ab, ob das Fließen von Elektrizität an die Bewegung eines chemischen Stoffes geknüpft ist (elektrolytischer Ladungstransport) oder nicht (elektronischer Ladungstransport). Für die Wärmeleitung kann man analoge Beobachtungen machen. In Metallen findet die Leitung in beiden Fällen ohne Stofftransport statt, d.h. Metalle sind i.A. sowohl gute elektronische Leiter und als auch gute Wärmeleiter. Das Analogon zur elektrolytischen Leitung ist für den Wärmetransport die Wärmekonvektion.

Des weiteren hängt die Kinetik von den Materialeigenschaften des Transportpfades ab. Beispielsweise kann man verschiedene Stäbe in die Hand nehmen und mit ihnen ein geladenes Elektrometer berühren. Für Metallstäbe geht der Zeigerausschlag des Elektrometers in in Sekundenschnelle zurück, mit einem Stück Tafelkreide kann es Minuten dauern und bei einem Glasstab ist die Abnahme so langsam, dass man sie auch innerhalb eines Tages noch nicht wahrnehmen kann. Die überwiegende Zahl der Materialien sind zugleich gute Wärmeleiter und gute elektrische Leiter, wie z.B. die Metalle. Gute elektrische Isolatoren, wie z.B. Styrophor und andere Kunststoffe, Gläser und viele Nichtmetalle, sind zugleich auch gute Wärmeisolatoren. Es gibt aber durchaus Ausnahmen, also Stoffe, die exzellente elektrische Leiter, aber sehr schlechte Wärmeleiter sind und umgekehrt.

Wärmewiderstand: Wenn zwei Teilsysteme (Knoten) einen Temperaturunterschied $\Delta T = T_{II} - T_I$ aufweisen, dann kommt es zu einem Wärmetransport vom Zustand mit höherer Temperatur zu dem mit niedrigerer Temperatur. Empirische Daten zum Wärmestrom $I_Q = I_Q(\Delta T)$ werden i.A. durch eine Potenzreihe beschrieben. Wenn die lineare Näherung genügt, erhält man das empirische Gesetz

$$I_Q = \frac{1}{R_Q} \Delta T .$$

Die Größe R_Q heißt WÄRMEWIDERSTAND.

Elektrischer Widerstand: Wenn zwei Knoten eine elektrische Potentialdifferenz $U = \Delta\varphi = \varphi_{II} - \varphi_I$ aufweisen, so liegt ebenfalls ein Nichtgleichgewichtszustand vor, der durch einen Ladungstransport abgebaut werden kann. Auch hier ist die Ursache des Transportprozesses das Streben nach maximaler Entropie. Häufig findet man, dass die lineare Näherung

$$I = \frac{1}{R} \Delta\varphi \tag{37.1}$$

bzw. das OHMSCHE GESETZ

$$U = RI \tag{37.2}$$

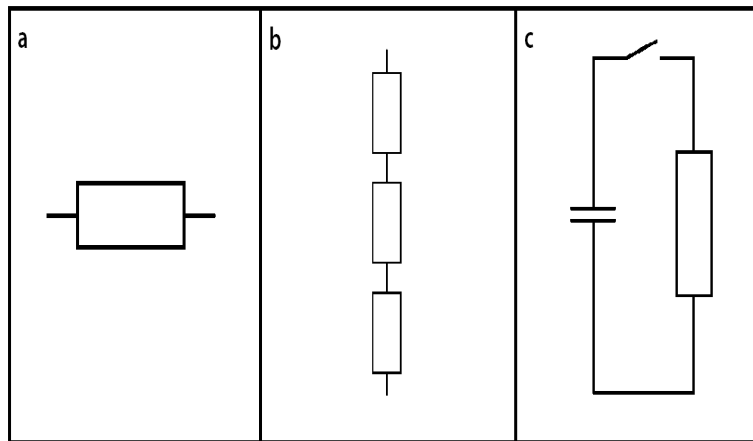


Abbildung 60: (a) Schaltsymbol für einen elektrischen Widerstand (b) Hintereinanderschaltung von drei Widerständen (c) Ein aus einem Widerstand und einem Kondensator bestehender Schaltkreis. Neben den Symbolen für Widerstand und Kondensator tritt hier ein Symbol für einen elektrischen Schalter auf.

für die Beschreibung des Zusammenhangs zwischen elektrischem Strom I und elektrischer Spannung $U = \Delta\varphi$ ausreicht. Die Maßeinheit des ELEKTRISCHEN WIDERSTANDS R heißt OHM:

$$[R] = 1 \text{ Ohm} = 1 \Omega = 1 \text{ V/A} \quad (37.3)$$

Die Größe R ist als positiv definiert. Daher ist das Vorzeichen des Stroms positiv, wenn positive elektrische Ladungen von einem Knoten höheren Potentials zu einem auf niedrigeren Potential wandern oder wenn negative Ladungen von einem Knoten niedrigeren Potentials zu einem auf höherem. Vorgänge können nur in dieser Richtung „freiwillig“ ablaufen, weil nur dann Energie zur Dissipation zur Verfügung steht bzw. die Entropie anwachsen kann.

Der Widerstand R_Q bzw. R ist ein Charakteristikum des Transportpfades. Jeder elektrische Transportpfad, für den das ohmsche Gesetz gilt, stellt einen OHMSCHEN WIDERSTAND R dar. Ein Transportpfad ist also ein ohmscher Widerstand, wenn der durch ihn fließende Strom der an den Knotenpunkten anliegenden Spannung proportional ist.

Wenn der Transportpfad durch einen Stirlingmotor oder einen Elektromotor mit idealen, nichtdissipativen Zufuhrleitungen realisiert ist, oder wenn Wasser über ein Druckgefälle durch eine Turbine fließt, dann wird zumindest ein Teil der frei werdenden Energie als Arbeit abgegeben. Ist das nicht der Fall, dann wird die Energie einfach dissipiert, was man meist anhand der Erwärmung des Transportpfades oder des Endknotens feststellen kann. Wenn man also beispielsweise eine Batterie einfach kurzschließt, dann erwärmen sich die Zuleitungen und die Batterie, denn irgendwo muss die durch den Transport frei werdende Energie ja hin.

Das Schaltkreissymbol für einen dissipativen elektrischen Widerstand als elektrisches (bzw. elektronisches) Bauelement ist in Abb. 60a gezeigt. Meist lässt man die nähere Bestimmung, elektrisch und dissipativ zu sein, i.A. weg und spricht einfach nur von einem Widerstand.

Wenn man mehrere Widerstände R_n mit $n = 1, 2, 3, \dots, N$ hintereinander schaltet, wie in Abb. 60b für drei gezeigt, dann steht jeder einzelne Widerstand dafür, dass in ihm pro Zeiteinheit eine gewisse Energie dissipiert wird. Da alle Widerstände vom gleichen Strom I durchflossen werden, ist die in jedem Widerstand dissipierte Leistung

$$P_n = -U_n I = -R_n I^2,$$

wobei U_n die zwischen seinen Anschlüssen liegende Spannung ist. Das negative Vorzeichen steht für die Leistungsabgabe. Die insgesamt abgegebene Leistung

$$P = -RI^2 = \sum_{n=1}^N P_n = -I^2 \sum_{n=1}^N R_n$$

ist die Summe aller dissipierten Teilleistungen und daraus folgt

$$R = \sum_{n=1}^N R_n.$$

Wenn sich ein Strom verzweigen kann und mehrere Widerstände von einem Verzweigungspunkt zu einem anderen führen, dann ist der Gesamtstrom

$$I = \frac{U}{R} = \sum_{n=1}^N I_n = U \sum_{n=1}^N \frac{1}{R_n},$$

denn die Spannung U zwischen den beiden Verzweigungspunkten ist die gleiche. Daraus folgt für parallel geschaltete Widerstände:

$$\frac{1}{R} = \sum_{n=1}^N \frac{1}{R_n}.$$

37.1. Messung elektrischer Widerstände.

Kalibrierte Messgeräte für elektrische Widerstände nennt man OHMMETER. Im Folgenden sollen zwei grundsätzlichen Messverfahren besprochen werden.

(a) **Kleine Widerstände:** Die einfachste und daher am häufigsten angewandte Messmethode beruht auf Gl. 37.2. Wie in Abb. 61 gezeigt, misst man dazu die über einem Widerstand abfallende Spannung mit einem Voltmeter und den durchlaufenden

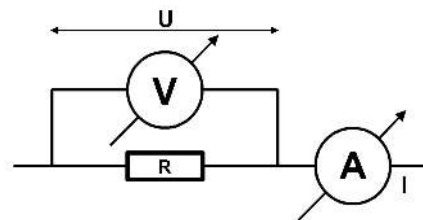


Abbildung 61: Schaltskizze zur Messung von Widerständen bzw. Strömen.

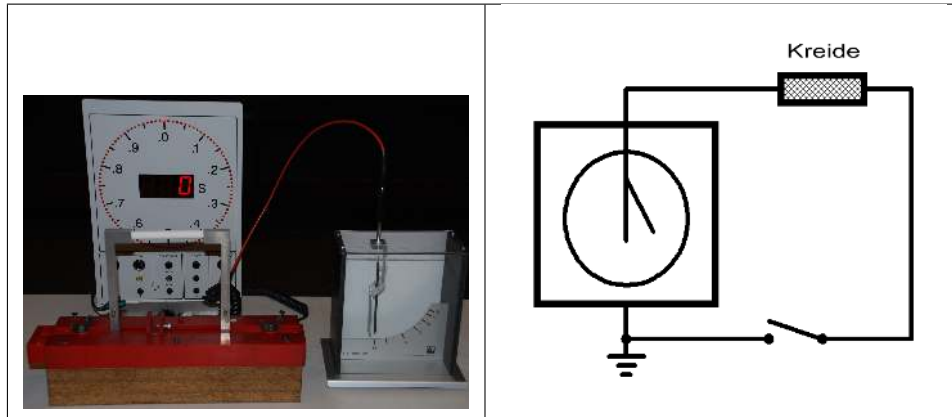


Abbildung 62: Messung des Widerstandes von Tafelkreide. a) Experimentelle Anordnung.
b) Schaltskizze.

Strom mit einem Amperemeter. Dieses Verfahren eignet sich für nicht allzu große Widerstände. Für extrem große Widerstände versagt sie, weil die Ströme dann unmessbar klein werden können.

(b) Große Widerstände: Um das Messverfahren für große Widerstände zu verstehen, betrachtet man die in Abb. 60c gezeigte aus einem Kondensator und einem Widerstand bestehende Schaltung. Der Spannungsabfall $-U_R = RI$ am Widerstand ist betragsmäßig gleich der Spannung U_C am Kondensator. Es gilt also:

$$U_C - U_R = q/C + RI = q/C + R\dot{q} = 0$$

Dies stellt eine lineare Differentialgleichung erster Ordnung dar, deren Lösung bereits bekannt ist (s. Kap. 13):

$$q(t) = q_0 \exp(-t/\tau) \quad (37.4)$$

Die Zeitkonstante

$$\tau = RC \quad (37.5)$$

der Entladung kann in elektronischen Schaltkreisen zur Steuerung zeitlicher Abläufe bzw. für die Zeitmessung herangezogen werden.

Um den elektrischen Widerstand von Tafelkreide zu messen, kann man ein Elektrometer bekannter Kapazität C mit einer Ladung q_0 aufladen und wie in Abb. 62a gezeigt, entladen. Mit Hilfe einer Stoppuhr schätzt man die Zeit ab, zu welcher die Ladung auf ungefähr 10% abgefallen ist. Mit Hilfe von Gl. 37.5 ergibt sich dann für das Kreidestück (Durchmesser 1 cm, Länge 8 cm) ein Widerstand R von größenordnungsmäßig $6 \times 10^{12} \Omega$.

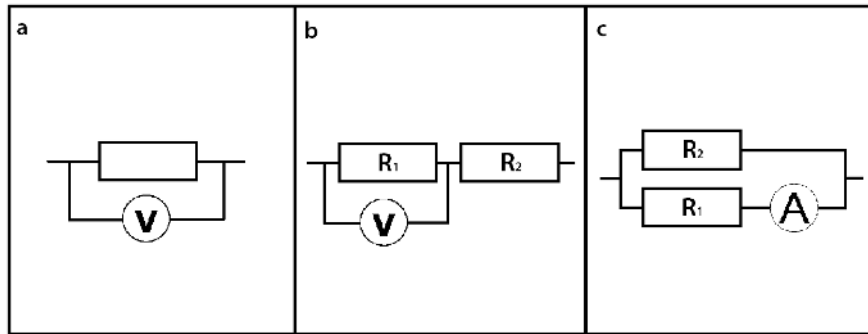


Abbildung 63: a) Realisierung eines Amperemeters durch ein Voltmeter und einen Messwiderstand. b) Messbereichserweiterung für ein Voltmeter. c) Messbereichserweiterung für ein Amperemeter.

37.2. Messbereichserweiterung

Der Messbereich eines Voltmeters ist beschränkt auf einen bestimmten Maximalaus Schlag. Wenn eine größere Spannung gemessen werden soll, als der Messbereich zulässt, kann man die in Abb. 63b gezeigte Schaltskizze verwenden. Man schließt zwei in Reihe geschaltete Widerstände R_1 und R_2 an und misst die Spannung (den Spannungsabfall) am Messwiderstand R_1 .

Mit hochgenau bekannten Widerständen (MESSWIDERSTÄNDEN) kann man aus einem Voltmeter ein Amperemeter machen. Wie in Abb. 63a gezeigt, misst man dazu den Spannungsabfall ΔU über dem Messwiderstand und eicht die Anzeige auf die Stromstärke. Wegen $\Delta U = RI$ erhält man bei bekanntem R den durch die Leitung geflossenen Strom I .

Der Messbereich eines Amperemeters lässt sich, wie in Abb. 63c gezeigt, erweitern, indem man durch einen zum Amperemeter parallel geschalteten Widerstand R_2 einen Teil des Stroms am Amperemeter vorbeifließen lässt und nur den durch R_1 fließenden Teilstrom misst. Aus den bekannten Werten der beiden Widerstände kann man auf den Gesamtstrom zurückschließen.

Teil V.

Mechanik im euklidischen Raum

38. Lineare Algebra (M1)

In diesem Kapitel werden Vektoren im Sinne der Mathematik definiert. In der Physik wird der Begriff des Vektors oft in einem enger gefassten Sinne verwendet, der auf den physikalischen Raum bezogen ist (s. Kap. 7.1).

38.1. Vektoren

VEKTOREN werden hier durch das Symbol „ $| \rangle$ “ gekennzeichnet. Vektoren wie $|a\rangle$ sind Elemente gleicher Art, für welche eine Addition $+$ und die Multiplikation mit einem Skalar α so definiert ist, so dass die LINEARKOMBINATION $|c\rangle = \alpha |a\rangle + |b\rangle$ ausgeführt werden kann und wieder einen Vektor ergibt. Die Operationen sind kommutativ, assoziativ und distributiv:

$$\begin{aligned}|a\rangle + |b\rangle &= |b\rangle + |a\rangle \\ |a\rangle + (|b\rangle + |c\rangle) &= (|a\rangle + |b\rangle) + |c\rangle \\ \alpha(|a\rangle + |b\rangle) &= \alpha |a\rangle + \alpha |b\rangle\end{aligned}$$

Die Menge aller Vektoren mit diesen linearen Operationen bildet einen LINEAREN RAUM, den *Vektorraum*.

Basis Ein Vektorraum besitzt eine kleinste Anzahl an Vektoren, durch die jeder seiner Vektoren als Linearkombination dargestellt werden kann. Ein Satz solcher Vektoren bildet die BASIS des Vektorraums. Die Anzahl der Basisvektoren bezeichnet die DIMENSION des Vektorraums.

Beispiel 1: Koordinatenvektoren in zwei Dimensionen. Genügen genau zwei Vektoren $|I\rangle$ und $|II\rangle$, so dass man jeden beliebigen Vektor $|A\rangle$ mit Hilfe zweier Skalare x_1 und x_2 durch

$$|a\rangle = x_1 |I\rangle + x_2 |II\rangle$$

ausdrücken kann, dann ist die aus den beiden Vektoren bestehende Menge $\{|I\rangle, |II\rangle\}$ die Minimalmenge, welche den Vektorraum aufspannt und somit eine Basis. Der Vektorraum ist zweidimensional. Die Skalare x_1 und x_2 sind die Koordinaten des Vektors $|a\rangle$ relativ zur Basis.

Aus den beiden Koordinaten lassen sich aus zwei Komponenten bestehende Objekte bilden und mit den Rechenregeln für Vektoren ausstatten. Man bezeichnet sie symbolisch mit

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

und nennt sie KOORDINATENVEKTOREN.

Beispiel 2: Lösungsraum von Differentialgleichungen. Die Anzahl der unabhängigen Lösungen einer linearen Differentialgleichung ist gleich seiner Ordnung. Die unabhängigen Lösungen bilden mit der linearen Struktur der Linearkombination einen Vektorraum.

Der Vektorraum der Lösungsfunktionen ist für alle linearen Differentialgleichungen zweiter Ordnung zweidimensional. Beispielsweise bilden die Vektoren $|1\rangle = \exp(\gamma x)$ und $|2\rangle = \exp(-\gamma x)$ eine Basis des Raums der Lösungen der Differentialgleichung

$$y'' - \gamma^2 y = 0 \quad (38.1)$$

Jede andere mögliche Lösung $|y\rangle = y(x)$ kann als Linearkombination

$$|y\rangle = a |1\rangle + b |2\rangle$$

mit zwei Skalaren a und b geschrieben werden. Insbesondere können die Vektoren $|1'\rangle = \cosh(\gamma x)$ und $|2'\rangle = \sinh(\gamma x)$ in dieser Basis ausgedrückt werden:

$$\begin{aligned} |1'\rangle &= \cosh(\gamma x) = \frac{1}{2} |1\rangle + \frac{1}{2} |2\rangle \\ |2'\rangle &= \sinh(\gamma x) = \frac{1}{2} |1\rangle - \frac{1}{2} |2\rangle \end{aligned}$$

Auch diese beiden Vektoren können als Basis des Lösungsraums von Gl. 38.1 herangezogen werden. Der Lösungsraum der Differentialgleichung

$$y'' + 2\gamma y' = 0 \quad (38.2)$$

wird von den beiden Vektoren $|1\rangle = 1$ und $|2\rangle = \exp(-2\gamma x)$ aufgespannt und die allgemeine Lösung

$$|y\rangle = a |1\rangle + b |2\rangle$$

ist wieder die Linearkombination der beiden Fundamentallösungen. Sie kann beispielsweise durch

$$y(x) = a + b \exp(-2\gamma x) = [a \exp(\gamma x) + b \exp(-\gamma x)] \exp(-\gamma x). \quad (38.3)$$

ausgedrückt werden.

38.2. Skalarprodukt

Zwischen zwei Vektoren $|a\rangle$ und $|b\rangle$ kann ein SKALARPRODUKT

$$\langle a|b\rangle = \lambda$$

erklärt sein, dessen Ergebnis ein Skalar λ ist. Es ist kommutativ und in jedem seiner beiden Faktoren linear. Fordert man $\langle a|a\rangle \geq 0$, dann kann man den Betrag bzw. die Norm

$$||a\rangle| = \sqrt{\langle a|a\rangle}$$

eines Vektors definieren. Vektoren mit Betrag 1 werden EINHEITSVEKTOREN genannt. Die sich aus

$$\cos \alpha = \frac{\langle a|b \rangle}{|a| |b|}$$

ergebende Zahl α wird als der Winkel zwischen den Vektoren $|a\rangle$ und $|b\rangle$ bezeichnet. Ist $\cos \alpha = 0$, dann stehen die Vektoren ORTHOGONAL zueinander. Besteht die Basis eines Vektorraums aus Einheitsvektoren, die orthogonal zueinander sind, dann bezeichnet man diese als eine ORTHONORMALBASIS.

38.3. Lineare Transformationen

T sei eine Abbildung, die einem Vektor $|x\rangle$ aus dem Vektorraum V einen Vektor $|y\rangle$ aus einem anderen Vektorraum W zuordnet: $T : |x\rangle \in V \mapsto |y\rangle \in W$. Wenn diese Abbildung die Eigenschaft hat, dass für zwei Vektoren $|a\rangle, |b\rangle \in X$ sowie einem beliebigen Skalar λ gilt:

$$T(\lambda |a\rangle + |b\rangle) = \lambda T(|a\rangle) + T(|b\rangle), \quad (38.4)$$

dann handelt es sich um eine *lineare Transformation*. Lineare Abbildungen lassen die Linearkombination als Struktur der beiden Vektorräume V und W unangetastet.

Beispiel: Differentiation Funktionen können als Elemente eines Vektorraums aufgefasst werden: Aus den Vektoren $|f\rangle = f(x)$ und $|g\rangle = g(x)$ lässt sich mit einem Skalar α durch Linearkombination ein anderer Vektor

$$|h\rangle = \alpha |f\rangle + |g\rangle$$

bilden, der die Funktion $h(x) = \alpha f(x) + g(x)$ repräsentiert. Die Differentiation stellt eine lineare Transformation auf dem Vektorraum der differenzierbaren Funktionen dar, d.h. es gilt

$$(\alpha |f\rangle + |g\rangle)' = \alpha |f\rangle' + |g\rangle'$$

38.4. Vektoren im euklidischen Raum

In der Physik sind Vektoren im dreidimensionalen euklidischen Raum von besonderer Bedeutung.

38.4.1. Die Richtungsbasis

Um Richtungen im dreidimensionalen euklidischen Raum zu spezifizieren, ist es üblich, von einer Orthonormalbasis der Richtungen als Bezugssystem auszugehen, im dreidimensionalen euklidischen Raum beispielsweise von einer aus drei Vektoren bestehenden

Orthonormalbasis $\{|1\rangle, |2\rangle, |3\rangle\}$. Diese orthogonalen Richtungs-Einheitsvektoren bilden die *Richtungsbasis*.

Orthogonal bedeutet, dass das Skalarprodukt zweier unterschiedlicher Einheitsvektoren null ist, wie z.B. $\langle 1|2\rangle = 0$. Das Skalarprodukt eines Einheitsvektors mit sich selbst ist hingegen 1, wie z.B. $\langle 1|1\rangle = 1$. Einheitsvektoren haben den Betrag 1 und sind somit dimensionslos.

Beispielsweise kann ein Punkt $P(x_1, x_2, x_3)$ im dreidimensionalen Raum durch den Vektor

$$|P\rangle = x_1 |1\rangle + x_2 |2\rangle + x_3 |3\rangle \quad (38.5)$$

angegeben werden. Relativ zur Richtungsbasis kann man den Punkt aber auch einfach durch den Koordinatenvektor

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

angegeben. Physikalische Größen, die man als Koordinatenvektoren relativ zu einer räumlichen Richtungsbasis angibt, werden nachfolgend durch einen übergestellten Pfeil „ $\vec{}$ “ gekennzeichnet. Wie man leicht nachrechnet,³⁷ gilt für das Skalarprodukt zwischen einem Vektor $|P\rangle$ und einem Vektor $|Q\rangle = y_1 |1\rangle + y_2 |2\rangle + y_3 |3\rangle$

$$\langle P|Q\rangle = x_1 y_1 + x_2 y_2 + x_3 y_3 = \vec{x} \cdot \vec{y}$$

Dies definiert das Standardskalarprodukt der Koordinatenvektoren und wird symbolisch durch das Malzeichen „ \cdot “ bezeichnet.

Wenn metrisierte Koordinaten vorliegen, dann ist

$$\vec{\ell} = \vec{x} - \vec{y}$$

der metrische Distanzvektor von Punkt P zu Punkt Q, Dies ist eine relativ zur Richtungsbasis angegebene physikalische Größe. Die Distanz zwischen den beiden Punkten ist gleich dem Betrag

$$\ell = \sqrt{\ell_1^2 + \ell_2^2 + \ell_3^2} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}$$

dieses Koordinatenvektors. Der Koordinatenvektor einer physikalischen Größe ist stets der Träger der physikalischen Einheit, während die Einheitsvektoren der Richtung, wie bereits erwähnt, stets dimensionslos sind.

38.4.2. Drehungen

Wenn man Drehbewegungen in Bezug auf ein Koordinatensystem beschreiben möchte, muss man zwischen rechts- und linkshändigen Koordinatensystemen unterscheiden. Sie

³⁷ $\langle P|Q\rangle = \langle x_1 |1\rangle + x_2 |2\rangle + x_3 |3\rangle |y_1 |1\rangle + y_2 |2\rangle + y_3 |3\rangle\rangle = x_1 y_1 \langle 1|1\rangle + x_1 y_2 \langle 1|2\rangle + \dots = x_1 y_1 + x_2 y_2 + x_3 y_3$

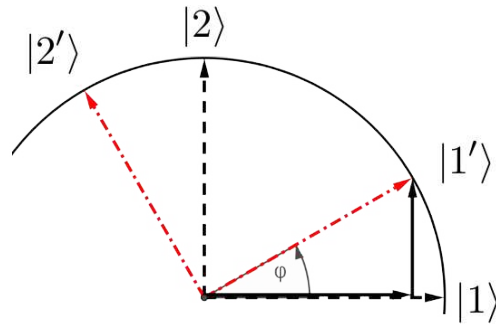


Abbildung 64: Skizze zur Ableitung der Gleichungen des Basiswechsels. Einheitsvektoren der beiden Richtungsbasen von S und S' haben den gleichen Betrag 1.

gehen durch Inversion ineinander über. Die Vereinbarung in diesem Buch ist, rechtshändige Koordinatensysteme zu verwenden. Ihre Orientierung ist in Abb. 28 skizziert.

Wenn S' ein Bezugssystem ist, das den gleichen Ursprung wie S hat, aber relativ zu S mit dem Winkel φ um die gemeinsame z-Achse verdreht ist, so ergibt sich seine aus den gedrehten Einheitsvektoren bestehende Richtungsbasis $\{|1'\rangle, |2'\rangle, |3'\rangle\}$ gemäß Abb. 64 zu

$$\begin{aligned} |1'\rangle &= \cos \varphi |1\rangle + \sin \varphi |2\rangle \\ |2'\rangle &= \sin \varphi |1\rangle - \cos \varphi |2\rangle \\ |3'\rangle &= |3\rangle \end{aligned}$$

Der gleiche durch den Vektor $|P\rangle$ beschriebene Punkt wird bezüglich S und bezüglich S' durch unterschiedliche Koordinatenvektoren beschrieben:

$$|P\rangle = x_1 |1\rangle + x_2 |2\rangle + x_3 |3\rangle = x'_1 |1'\rangle + x'_2 |2'\rangle + x'_3 |3'\rangle$$

Durch Bilden des Skalarprodukts $\langle 1'|P\rangle = x'_1 = x_1 \langle 1'|1\rangle + x_2 \langle 1'|2\rangle$ erhält man

$$x'_1 = \cos \varphi x_1 + \sin \varphi x_2$$

Das Resultat stimmt mit der Koordinatentransformation gemäß Gl. 5.2 überein.

Aus der eindimensionalen Komponente für die Orts- bzw. Impulskoordinate entlang der x-Achse erhält man die entsprechenden dreidimensionalen Komponenten der Koordinatenvektoren durch eine Drehung der Richtungsbasis.

38.4.3. Orientierung bzw. Händigkeit

Mit $\{|1\rangle, |2\rangle, |3\rangle\}$ sei eine gereichte Basis gegeben, d.h. Basisvektoren mit festgelegter Reihenfolge. Damit ist deren Orientierung (bzw. Händigkeit) festgelegt. Nun kann man

zwei der Basisvektoren vertauschen bzw. permutieren und erhält eine gereichte Basis $\{|1'\rangle, |2'\rangle, |3'\rangle\}$ mit $|1'\rangle = |2\rangle$, $|2'\rangle = |1\rangle$ und $|3'\rangle = |3\rangle$. Das ist eine anders orientierte Basis. Sie ist zur ersteren spiegelsymmetrisch. Analog zu 5.4.2 wird im Folgenden stets von einer rechtshändigen Richtungsbasis ausgegangen.

Skalare und Pseudoskalare. Physikalische Skalare sind einkomponentige Größen, die sich einerseits nicht ändern, wenn man zu einer gedrehten Richtungsbasis übergeht, und andererseits auch nicht, wenn man statt einer rechtshändigen eine linkshändige Richtungsbasis zugrunde legt. Beispielsweise ist das Skalarprodukt zweier Vektoren das Gleiche, egal ob es sich bei beiden Vektoren um Koordinatenvektoren handelt, die sich auf eine links- oder eine rechtshändige Richtungsbasis beziehen.

Neben den Skalaren gibt es einkomponentige Größen, die bei einem Wechsel der Händigkeit ihr Vorzeichen ändern. Sie werden als Pseudoskalare bezeichnet. Winkel sind beispielsweise Pseudoskalare (s. Kap. 1.6.1).³⁸

Polare und axiale Vektoren. Bei einem Wechsel der Händigkeit ändern die Richtungsvektoren ihr Vorzeichen, d.h. $|1\rangle \rightarrow -|1\rangle$ usw. Ein Vektor wie der Positionsvektor $|P\rangle$ aus Gl. 38.5, ändert sich durch einen Wechsel der Händigkeit nicht, denn die Position ist nach wie vor physikalisch die gleiche Position im Raum. Das bedeutet aber, dass der Koordinatenvektor sein Vorzeichen ändert, d.h. $\vec{x} \rightarrow -\vec{x}$. Koordinatenvektoren, die sich so verhalten, bezeichnet man als *polare Vektoren*.

Ein Vektor $|A\rangle$, der bei einem Wechsel der Händigkeit sein Vorzeichen ändert und dessen Koordinatenvektor folglich unverändert bleibt, d.h. $\vec{A} \rightarrow \vec{A}$, wird als *axialer Vektor* bezeichnet und durch einen nach links zeigenden Pfeil gekennzeichnet.

38.4.4. Kreuzprodukt

Sind \vec{a} und \vec{b} zwei bezüglich einer rechtshändigen Richtungsbasis gegebene polare Koordinatenvektoren, so erhält man durch das *Vektor-Kreuzprodukt*

$$\vec{c} = \vec{a} \times \vec{b} = \begin{pmatrix} a_y b_z - a_z b_y \\ a_z b_x - a_x b_z \\ a_x b_y - a_y b_x \end{pmatrix} = -\vec{b} \times \vec{a} \quad (38.6)$$

einen axialen Vektor \vec{c} . Das Kreuzprodukt ist antikommutativ! Aus dem Kreuzprodukt eines axialen und eines polaren Vektors geht ein polarer Vektor hervor:

$$\vec{c} = \vec{a} \times \vec{b}.$$

³⁸Ein anderes Beispiel für einen Pseudoskalar ist Spatprodukt dreier polarer Vektoren bzw. die Determinante. Die aus den Koordinatenvektoren einer Einheitsbasis gebildete Determinante hat je nach Händigkeit die Werte +1 oder -1. Wenn man daher einen Skalar mit der Determinante der Richtungsbasis multipliziert, erhält man einen Pseudoskalar.

38.4.5. Ableitung von Vektorfunktionen

Eine Vektorfunktion ist ein Komponentenvektor, dessen Komponenten von einem Parameter abhängen. Ein Beispiel ist ein bezüglich einer Richtungsbasis definierter Vektor

$$\vec{A}(t) = \begin{pmatrix} A_1(t) \\ A_2(t) \\ A_3(t) \end{pmatrix},$$

dessen Komponenten Funktionen der Zeit t sind. Die zeitliche Ableitung der Vektorfunktion ergibt sich aus der zeitlichen Ableitung nach jeder seiner Komponenten:

$$\dot{\vec{A}}(t) = \begin{pmatrix} \dot{A}_1(t) \\ \dot{A}_2(t) \\ \dot{A}_3(t) \end{pmatrix}$$

39. Relativitätstheorie im dreidimensionalen Raum

Ich möchte hier zuerst ein bekanntes Gedankenexperiment besprechen, das aufzeigt, dass die eindimensionale Relativitätstheorie, die in Kap. 8 dargestellt wurde, nur mit einer euklidischen räumlichen Metrik kompatibel ist, d.h. einen flachen Raum fordert.

39.1. Beschränkung aller Geschwindigkeiten durch c

Von einem Galileisystem S' aus gesehen bewege sich ein Objekt mit der Geschwindigkeit u' nach rechts. Wenn man annimmt, dass sich das Bezugssystem S' von S aus gesehen ebenfalls nach rechts bewegt, und zwar mit der Geschwindigkeit v , dann darf die Geschwindigkeit u des Objekts bezüglich S nicht kleiner ausfallen als u' bzw. v :

$$u \geq u', v \quad (39.1)$$

Um die Konsequenzen dieser Feststellung zu erkennen, betrachtet man den Ausdruck $c^2 - u^2$. Er ist ein Maß für die Abweichung des Betrags der Geschwindigkeit u von c . Setzt man Gl. 8.16 ein, so erhält man:³⁹

$$c^2 - u^2 = \frac{(c^2 - v^2)(c^2 - u'^2)}{c^2(1 + u'v/c^2)^2} \quad (39.2)$$

Der Nenner des rechten Ausdrucks ist stets positiv. Wenn Geschwindigkeitsbeträge möglich wären, die größer als c sind, dann würde aus $v, u' > c$ folgen, dass $u < c$. Das ist im Widerspruch zu Gl. 39.1. Ist hingegen $v, u' \leq c$, dann ist Gl. 39.1 stets erfüllt. Da die rechte Seite positiv ist, ist das Ergebnis auch konsistent, denn auch für u gilt stets $u \leq c$. Mit anderen Worten, wenn man die Geschwindigkeiten von einem Galileisystem

³⁹wegen $c^2 - u^2 = \frac{c^2(1+2u'v/c^2+u'^2v^2/c^4)-(u'+v)^2}{(1+u'v/c^2)^2} = \frac{c^4+u'^2v^2-v^2c^2-u'^2c^2}{c^2(1+u'v/c^2)^2} = \frac{(c^2-v^2)(c^2-u'^2)}{c^2(1+u'v/c^2)^2}$

in ein anderes transformiert, kann man durch Hintereinanderschaltung von Geschwindigkeitstransformationen niemals aus der Menge der Transformationen hinauskommen, bei denen die Geschwindigkeit kleiner als c ist. Folglich gibt es eine universelle Geschwindigkeitsbeschränkung:

Kein Geschwindigkeitsbetrag eines Signals oder Objekts kann die Raum/Zeit-Konstante c übersteigen.

39.2. Ulks

Das ideale relativistische Teilchen heißt ULK. Jedenfalls ist das der Name, den ich diesem Teilchen gegeben habe. Ulks sollen Teilchen sein, die sich mit einem Geschwindigkeitsbetrag von c bewegen, also mit der gemäß Gl. 39.2 gerade noch erlaubten Grenzggeschwindigkeit. Mehr braucht man dazu nicht zu sagen, denn wenn sie es bezüglich eines Galileisystems tun, dann tun sie es in allen. Hinsichtlich ihrer Masse haben sie allerdings eine etwas gewöhnungsbedürftige Eigenschaft: Wegen Gl.11.16 folgt aus $|v| = c$, dass $p_0 = |p_1|$ und aus Gl. 11.11, dass $P = 0$ und somit $m = 0$. Ulks sind masselose Teilchen!

Gibt es Ulks? Diese Frage gehört zur gleichen Kategorie wie die Frage, ob es ein ideales Gas gibt. Das ideale relativistische Teilchen ist genauso wie das ideale Gas zunächst einmal eine rein theoretische Abstraktion, die man für Gedankenexperimente eingesetzt. Sie stellen den gerade noch zulässigen Grenzfall dar, der durch Verbote gerade noch nicht ausgeschlossen ist. Daher auch ihre etwas merkwürdigen Eigenschaften, nämlich dass ideale Gase für $T = 0$ volumenlos werden ($V = 0$) und ideale relativistische Teilchen masselos ($m = 0$). Es ist also völlig egal, ob es ein Ulk oder ein ideales Gas tatsächlich gibt. Wichtig ist nur, dass weder durch das eine noch das andere Modell ein grundlegendes physikalisches Axiome verletzt wird. Tatsächlich gibt es ein Gas, das dem idealen Gas sehr nahe kommt, nämlich das Photonengas, und lange hat man auch geglaubt, dass Neutrinos, Photonen und Gluonen solche Ulks sind. Nach dem neuesten Stand der Experimente kann man aber inzwischen ausschließen, dass die Neutrinos zu diesem exklusiven Club dazugehören, denn man konnte inzwischen experimentell nachweisen, dass sie nicht masselos sind.

Aber völlig unabhängig davon, ob Ulks in der Natur real existieren oder nicht, dürfen Theoretiker jedes zulässige Modell verwenden, um theoretische Konsequenzen der Relativitätstheorie zu durchdenken. Hier wollen wir nun eine Ulk-Uhr betrachten. Das soll eine Uhr sein, die einen periodischen Vorgang dadurch realisiert, dass ein Ulk zwischen zwei relativ zueinander ruhenden Spiegeln hin- und herreflektiert wird (Gedankenexperiment!). Die Ulk-Uhr werde in einem Galileisystem S' beobachtet, das sich mit der Geschwindigkeit v relativ zu S bewege. Zur Zeit $t = t' = 0$ werde das Signal bei $x = x' = 0$ abgeschickt. Wenn der Weg für einen Rundtrip gleich L' ist, dann braucht das Signal für einen Rundtrip die Zeit $t' = L'/c$. Setzt man für das Ankunftsereignis die Koordinaten $t' = L'/c$ und $x' = 0$ in die Lorentzgleichungen (Gl. 8.21) ein, so erhält man für dieses Ereignis in S die Koordinaten:

$$t = \gamma(v)t' \tag{39.3}$$

Für Beobachter in S hat das Signal in dieser Zeit die Strecke

$$L = ct = c\gamma(v)t' = \gamma(v)L'$$

zurückgelegt, weil sich Ulk-Signale auch in S mit einer Geschwindigkeit c bewegen. Der zeitliche Abstand zwischen den beiden Ereignissen in S ist um $\gamma(v)$ größer als in S' (Zeitdilatation), aber wundern wird sich dort niemand darüber, weil das Signal dort einen um den gleichen Faktor $\gamma(v)$ längeren Weg zurückgelegt hat. Letzteres wird auch als LORENTZKONTRAKTION bezeichnet, weil man es auch umgekehrt sehen kann, dass nämlich der Weg in S' um den Faktor $1/\gamma(v)$ kleiner ausfällt als in S. Zeitdilatation und Lorentzkontraktion sind gewissermaßen zwei Seiten ein und derselben Medaille. Analog zu den thermodynamischen Schlüssen, die man für das ideale Gas ziehen kann, sind die Schlussfolgerungen korrekt, die hier für Ulks gezogen wurden, und zwar unabhängig davon, ob Ulks real existierende Teilchen sind oder nicht.

39.3. Warum ist der isotrope homogene Raum euklidisch?

In diesem Abschnitt soll zunächst einmal zugelassen sein, dass der Raum nichteuklidisch sein könnte. Eine Linie, welche die kürzeste Verbindungslinie zwischen allen auf ihr liegenden Punkten ist, bezeichnet man als Geodäte. Im zweidimensionalen Raum sollen sich S und S' relativ zueinander entlang der gleichen mit x - bzw. x' -Geodäten bewegen. Der Ursprung von S' werde von irgendeiner Geodäten geschnitten. Wenn das Ulk sich entlang dieser Geodäten hin und zurück bewegt, und wenn dieser Rundtrip in der gleichen Zeit t' ausgeführt wird, so wird gemäß der Definition der Längeneinheit jedes Mal der gleiche Weg $L' = t'/c$ zurückgelegt. Wegen der Isotropie des Raumes in S' wird jeder dieser Wege in der gleichen Laufzeit t' zurückgelegt und folglich ergibt sich auch bezüglich S die gleiche Laufzeit

$$t = \gamma(v)t' .$$

Man kann nun leicht zeigen, dass sich diese Beziehung nur für einen euklidischen Raum einhalten lässt. Man betrachtet dazu beispielsweise eine Ulk-Signal entlang der Geodäten $x' = 0$, d.h. entlang einer zur x' -Geodäten orthogonalen Geodäte, die hier als y -Geodäte bezeichnet werden soll. Bezüglich S gilt, dass das Signal vom Nullpunkt zum Spiegel mit den Koordinaten $(vt/2, L'/2)$ und zurück gelaufen ist. Wenn der Raum euklidisch ist, dann gilt $L^2 = L'^2 + v^2t^2$ bzw. $L = L'/\sqrt{1 - v^2/c^2}$ bzw. $t = \gamma(v)t'$. Daraus folgt: Wenn der Raum eine euklidische Metrik hat (für Spezialisten: damit ist er flach), dann ist die Zeitdilatation auch in S isotrop und für jede Richtung im Zweidimensionalen genauso groß wie im Eindimensionalen. Aus Konsistenzgründen wird daher im Folgenden das euklidische Raummodell für die Raumzeit zugrunde gelegt.

39.4. Relativistische Kinematik und Dynamik in drei Dimensionen

Die Erweiterung der eindimensionalen Gleichungen auf drei Dimensionen ist damit ganz einfach: Alle skalaren Größen wie die Zeit bleiben Skalare, alle Vektoren werden dreidimensionale Vektoren und erhalten einen Pfeil über dem entsprechenden Symbol. Dort wo

das Produkt zweier Vektoren auftritt, muss es durch ein Skalarprodukt von mit einem Pfeil geschriebenen Vektoren ersetzt werden. Einige Beispiele sollen das hier erläutern:

1. Gl. 8.16 ist durch

$$\vec{u} = \frac{\vec{v} + \vec{u}'}{1 + \vec{u}' \cdot \vec{v}/c^2}$$

zu ersetzen.

2. Für den Impuls erhält man

$$\vec{p} = \frac{m\vec{v}}{\sqrt{1-v^2/c^2}} \quad (\text{Impuls})$$

und für die Energie.

$$E = \sqrt{m^2c^4 + \vec{p}^2c^2} \quad (39.4)$$

Die Lorentzgleichungen müssen tensoriell formuliert werden. Das soll erst später nachgeholt werden (Romanos Physikkurs III).

40. Bewegung im Raum

40.1. Bewegung im zweidimensionalen Raum

Eine Bewegung im Zweidimensionalen ist dadurch gegeben, dass man für jeden Punkt der Bewegung für einen gegebenen Zeitpunkt t zwei unabhängige Koordinaten angibt, beispielsweise die beiden kartesischen Koordinaten $x(t)$ und $y(t)$ des Ortsvektors

$$\vec{r}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}. \quad (40.1)$$

Die Gleichung einer allgemeinen, durch den Ursprung verlaufenden gleichförmigen geradlinigen Bewegung mit der Geschwindigkeit v ist in zwei Dimensionen durch

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} vt \cos \alpha \\ vt \sin \alpha \end{pmatrix} \quad (40.2)$$

gegeben. Eliminiert man aus 40.2 die Zeit t , so erhält man die Gerade

$$y = x \tan \alpha \quad (40.3)$$

mit dem Steigungswinkel α . Die Geschwindigkeit einer gleichförmigen und geradlinigen Bewegung ist konstant. Man kann sie aus Gl. 40.2 berechnen, indem man Gl. 6.3 auf jede der beiden Koordinaten anwendet. Es ergibt sich Vektor

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_x \\ v_y \end{pmatrix} = \begin{pmatrix} \Delta x / \Delta t \\ \Delta y / \Delta t \end{pmatrix} = \begin{pmatrix} v \cos \alpha \\ v \sin \alpha \end{pmatrix} = v \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix} = v \vec{e} \quad (40.4)$$

der Geschwindigkeit. Der Einheitsvektor (*Richtungseinheitsvektor*)

$$\vec{e} = \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix}$$

gibt die Bewegungsrichtung zur Richtungsbasis des Koordinatensystems an. Die Konstante $v = \sqrt{v_1^2 + v_2^2}$ ist der Betrag der Geschwindigkeit. Für eine Bewegung entlang der x -Achse ist $\alpha = 0$ und es gilt $v_x = v$ und $v_y = 0$.

40.2. Bewegung im dreidimensionalen Raum

Bei der Beschreibung der Bewegung eines Objekts im dreidimensionalen Raum müssen drei Ortskoordinaten bezüglich eines Koordinatensystems als Funktion der Zeit angegeben werden. Für kartesische Koordinaten ist das z.B. der Vektor

$$\vec{r}(t) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix} \quad (40.5)$$

Gl. 40.5 beschreibt die Ortskurve des Objekts durch einen Vektor, der durch die Zeit parametrisiert ist.

Wenn man einen parametrisierten Vektor differenzieren möchte, so muss man die Differentiation nach dem Parameter für jede Vektorkomponente durchführen. Beispielsweise ergibt sich die Geschwindigkeit aus Gl. 40.5 durch Ableitung jeder der drei Komponenten nach der Zeit:

$$\vec{v}(t) = \begin{pmatrix} v_x(t) \\ v_y(t) \\ v_z(t) \end{pmatrix} = \dot{\vec{r}}(t) = \frac{d\vec{r}}{dt} = \begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \\ \dot{z}(t) \end{pmatrix} \quad (40.6)$$

Analog erhält man die höheren Ableitungen von Vektoren, beispielsweise die Beschleunigung

$$\vec{a}(t) = \dot{\vec{v}}(t) = \frac{d\vec{v}}{dt} = \begin{pmatrix} \dot{v}_x(t) \\ \dot{v}_y(t) \\ \dot{v}_z(t) \end{pmatrix} \quad (40.7)$$

Nur wenn alle drei Komponenten verschwinden, ist die Bewegung geradlinig und gleichförmig und somit nicht beschleunigt.

Wie Sie sehen, ist es keine prinzipielle Schwierigkeit, statt einer eindimensionalen eine zwei- oder dreidimensionale Bewegung zu beschreiben. In diesem Buch wird daher, wo es möglich ist, der einfachere Fall eindimensionaler Bewegungen herangezogen, um grundlegende physikalische Konzepte möglichst einfach und übersichtlich darzustellen.

41. Dreidimensionale Newtonsche Dynamik

Auch für die Newtonsche Physik erhält man die Gleichungen für den dreidimensionalen Raum, indem man alle Skalare belässt und alle eindimensionalen Vektoren (die Grö-

ßen bei Rauminversion das Vorzeichen wechseln) durch einen Pfeil kennzeichnet und als dreikomponentige Vektoren schreibt.

41.1. Newtonscher Impuls

In dreidimensionaler Formulierung lautet die Newtonsche Bewegungsgleichung, Gl. 15.7,

$$\dot{\vec{p}} = \begin{pmatrix} \dot{p}_1 \\ \dot{p}_2 \\ \dot{p}_3 \end{pmatrix} = m\vec{v} = \begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix} = \vec{F}. \quad (41.1)$$

Die Kraft

$$\vec{F} = \sum_{i=1}^N \vec{F}_i. \quad (41.2)$$

setzt sich vektoriell aus den Einzelbeiträgen zusammen. Koordinatenvektoren unterschiedlicher physikalischer Größen gehören i.A. unterschiedlichen Vektorräumen an. Das trifft zum Beispiel auf die Koordinatenvektoren \vec{F} der Kraft und \vec{a} der Beschleunigung zu. Durch die lineare Transformation (Multiplikation mit dem Skalar m , der Masse)

$$\vec{F} = m\vec{a}$$

wird die Beschleunigung auf die Kraft abgebildet.

Wenn die Position \vec{x} die Zustandsvariable ist, muss Gl. 16.2 entsprechend zum Wegintegral

$$\Delta W = \int_{\vec{x}_0}^{\vec{x}_1} \vec{F}_{ext}(\vec{x}) \cdot d\vec{x} \quad (41.3)$$

erweitert werden. Ein infinitesimal kleiner Beitrag $\delta W = \vec{F}_{ext} \cdot d\vec{x}$ zur Arbeit ist positiv, wenn der differentielle Weg $d\vec{x}$ parallel zur angreifenden Kraft \vec{F}_{ext} ist. Dann wird Arbeit zugeführt und die innere Energie erhöht sich um $d\mathcal{U} = \delta W$. Das Symbol "δ" bedeutet, dass es sich um einen infinitesimal kleinen Arbeitsbeitrag δW handelt, während das Symbol „d“ für dU nicht nur bedeutet, dass die Größe infinitesimal klein ist, sondern dass sie auch ein Differential darstellt.

Beispiel: Als Anwendung der Newtonschen Näherung sei hier der (näherungsweise) elastischen Stoß zweier Teilchen A und B gleicher Masse m betrachtet. Das können z.B. der Stoß zweier gleicher Münzen sein. Für die Beschreibung sei dasjenige Laborsystem gewählt, wo die Münze B zu Anfang ruht. Dort lautet die Energie- und Impulserhaltung

$$\frac{p_A^2}{2m} + 0 = \frac{\tilde{p}_A^2}{2m} + \frac{\tilde{p}_B^2}{2m} \quad (41.4)$$

$$\vec{p}_A + 0 = \vec{\tilde{p}}_A + \vec{\tilde{p}}_B \quad (41.5)$$

Quadriert man Gl. 41.5, so folgt aus Gl. 41.4:

$$\tilde{\vec{p}}_A \cdot \tilde{\vec{p}}_B = 0 \quad (41.6)$$

Das bedeutet, dass die Impulse der beiden Teilchen nach dem Stoß orthogonal zueinander sind. Wenn man zwei gleiche Münzen gegeneinander schnippt, kann man tatsächlich beobachten, dass sie nach dem Zusammenstoß im rechten Winkel auseinander gleiten.

41.2. Newtonscher Drehimpuls und Drehmoment

Wenn die Summe aller Kräfte in der Newtonschen Bewegungsgleichung verschwindet, also $\vec{F} = 0$ gilt, ändert sich der Impuls des Systems nicht, d.h. der Impuls ist dann eine Erhaltungsgröße. Außer der Impulserhaltung gibt es aber noch andere Größen, die bei physikalischen Prozessen erhalten sein können. Das soll hier an einem Beispiel aufgezeigt werden.

Wenn man die Newtonsche Bewegungsgleichung vektoriell mit dem Ortsvektor \vec{r} multipliziert, dann erhält man

$$\vec{r} \times \dot{\vec{p}} = \vec{r} \times \vec{F} \quad (41.7)$$

Da Impuls und Geschwindigkeit parallel zueinander sind, gilt

$$\frac{d}{dt} (\vec{r} \times \vec{p}) = \dot{\vec{r}} \times \vec{p} + \vec{r} \times \dot{\vec{p}} = 0 + \vec{r} \times \dot{\vec{p}}$$

und somit

$$\frac{d}{dt} (\vec{r} \times \vec{p}) = \vec{r} \times \vec{F} \quad (41.8)$$

Wenn die Kräfte immer radial an einem Objekt angreifen, ist die rechte Gleichungsseite null. Dann ist $\vec{r} \times \vec{p}$ eine zeitlich unveränderliche Größe, also eine Erhaltungsgröße, und wegen der Additivität des Impulses darüber hinaus eine additive Erhaltungsgröße. Es liegt daher nahe, hierfür eine eigene Größe einzuführen. Man nennt diese additive Erhaltungsgröße den DREHIMPULS

$$\tilde{\vec{L}} = \vec{r} \times \vec{p} \quad (41.9)$$

Der Drehimpuls ist stets orthogonal zu der von \vec{r} und \vec{p} aufgespannten Bahnebene. Die Größe

$$\tilde{\vec{T}} = \vec{r} \times \vec{F} \quad (41.10)$$

bezeichnet man als DREHMOMENT. Damit lässt sich Gl. 41.8 analog zur Newtonschen Bewegungsgleichung formulieren:

$$\frac{d\tilde{\vec{L}}}{dt} = \tilde{\vec{T}} \quad (41.11)$$

Wenn das Drehmoment $\tilde{\vec{T}}$ verschwindet, bleibt der Drehimpuls erhalten. Wenn eine Kraft eine von einem Punkt M ausgehende Radialkraft ist, d.h. bezüglich M in radialer

Richtung wirkt, so verschwindet das Drehmoment bezüglich M und der Drehimpuls bezüglich M ist eine Erhaltungsgröße. Mit den etwas unglücklich gewählten historisch gewachsenen Bezeichnungen Drehimpuls und Drehmoment für diese Erhaltungsgrößen wird man wohl leben müssen, obwohl sie nicht notwendigerweise etwas mit Drehungen zu tun haben.

Beispiel 1: Drehimpulserhaltung für eine geradlinige gleichförmige Bewegung Für eine geradlinige Bewegung zerlegen wir den Ortsvektor $\vec{r} = \vec{r}_{\parallel} + \vec{r}_{\perp}$ in eine Komponente \vec{r}_{\parallel} parallel zur Bewegungsrichtung und eine Komponente \vec{r}_{\perp} orthogonal dazu. Da kein Drehmoment angreift, bleibt der Drehimpuls erhalten. Er steht orthogonal zur von \vec{r}_{\perp} und dem konstanten Impuls \vec{p} aufgespannten Ebene und hat den Betrag $L = r_{\perp}p$. Wie man sieht, hängt Drehimpuls vom Bezugspunkt ab, auf den man ihn bezieht. Im vorliegenden Fall sind alle Bezugspunkte äquivalent, die auf einer Parallelen zur Bahn im Abstand r_{\perp} liegen.

Beispiel 2: Drehimpulserhaltung für Eisenbahner Während der ganzen Bahnfahrt sei die Bahngeschwindigkeit konstant. Wir fahren eine zunächst gerade Schienenstrecke. Irgendwo gehe diese in eine Kurve mit dem Krümmungsradius r_{\perp} über und verlaufe anschließend wieder gerade aus. Dann ist nur der Drehimpuls bezüglich Krümmungsmittelpunktes eine Erhaltungsgröße.

42. Winkelgeschwindigkeit und -beschleunigung

Zur Beschreibung von Drehbewegungen geht man im Zweidimensionalen am Zweckmäßigsten von Polarkoordinaten aus und im Dreidimensionalen von Zylinderkoordinaten, wobei man die Drehachse üblicherweise entlang der z -Achse wählt.

Die Winkelkoordinate $\varphi = \varphi(t)$ ist eine Funktion der Zeit. Ihre Ableitung ist die WINKELGESCHWINDIGKEIT

$$\omega(t) = \dot{\varphi}(t) = \frac{d\varphi}{dt} \quad (42.1)$$

Die Größe

$$\dot{\omega} = \frac{d\omega}{dt} = \frac{d^2\varphi}{dt^2} = \ddot{\varphi} \quad (42.2)$$

heißt WINKELBESCHLEUNIGUNG. Sie ist gleich der ersten Ableitung der Winkelgeschwindigkeit nach der Zeit und gleich der zweiten Ableitung des Winkels nach der Zeit.

42.1. Gleichförmige Kreisbewegung in zwei Dimensionen

Wenn die Winkelgeschwindigkeit ω und der Radius r bei der Bewegung konstant sind, liegt eine GLEICHFÖRMIGEN KREISBEWEGUNG vor. Aus Gl. 42.1 folgt

$$\varphi(t) = \omega t \quad (42.3)$$

In gleichen Zeiten werden also gleiche Winkel durchlaufen. Der Ortsvektor für das Durchlaufen der Kreisbahn ist

$$\vec{r}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} r \cos \varphi(t) \\ r \sin \varphi(t) \end{pmatrix} = r \begin{pmatrix} \cos \omega t \\ \sin \omega t \end{pmatrix} \quad (42.4)$$

Das kann man leicht einsehen, indem man aus den beiden Gleichungen für die Koordinaten t eliminiert und so die Beziehung zwischen den Koordinaten x und y erhält. Man quadriert dazu die erste und die zweite Koordinate und summiert auf:⁴⁰

$$x^2 + y^2 = r^2 \cos^2 \omega t + r^2 \sin^2 \omega t = r^2 \quad (42.5)$$

Das ist die KREISGLEICHUNG. Die durch Gl. 42.4 beschriebene Bewegung auf der Kreisbahn ist periodisch, denn sie wiederholt sich, wenn $\omega t = 2n\pi$ ist, wobei $n = 0, 1, 2, \dots$ die Zahl der Umläufe bezeichnet. Die Dauer einer Periode ist

$$T = 2\pi/\omega \quad (42.6)$$

Aus Gl. 42.4 folgt für die Geschwindigkeit:

$$\vec{v}(t) = \dot{\vec{r}}(t) = r\omega \begin{pmatrix} -\sin \omega t \\ \cos \omega t \end{pmatrix} \quad (42.7)$$

Obwohl es sich um eine gleichförmige Bewegung handelt, d.h. um eine Bewegung mit konstanter Bahngeschwindigkeit $v = |\vec{v}| = \omega r$, ist die Geschwindigkeit \vec{v} nicht konstant, weil sich die Bewegungsrichtung ständig ändert.⁴¹ Der in radialer Richtung zeigende Einheitsvektor

$$\vec{e}_r = \begin{pmatrix} \cos \omega t \\ \sin \omega t \end{pmatrix} \quad (42.8)$$

der Ortsposition steht orthogonal auf dem in tangentialer Richtung zeigenden Einheitsvektor

$$\vec{e}_t = \begin{pmatrix} -\sin \omega t \\ \cos \omega t \end{pmatrix} \quad (42.9)$$

der Geschwindigkeit.

⁴⁰Zur Notation: $\cos \omega t$ ist die unter Physikern übliche schlampige Schreibweise für $\cos(\omega t)$ und nicht für $\cos(\omega)t$. Wenn letzteres gemeint wäre, würde ein Physiker das so schreiben: $t \cos \omega$. Im vorliegenden Fall kommt das schon deshalb hier nicht in Frage, weil t die physikalische Dimension einer Zeit hat (z.B. in Sekunden angegeben ist) und im Argument einer Funktion nur *dimensionslose* Größen stehen dürfen. Man muss daher *im Argument* mit einer Konstanten ω multiplizieren, welche die Dimension $1/s$ bzw. s^{-1} hat (1 durch Sekunde).

⁴¹Die englische Sprache ist hier im Vorteil: Mit „speed“ ist der Betrag der Geschwindigkeit gemeint bzw. die Bahngeschwindigkeit v , mit „velocity“ die vektorielle Größe \vec{v} . Sowohl bei der gleichförmig-geradlinigen Bewegung als auch bei der gleichförmigen Kreisbewegung ist die Bahngeschwindigkeit bzw. „speed“ konstant, die Geschwindigkeit bzw. „velocity“ nur bei ersterer.

Da die gleichförmige Kreisbewegung nicht geradlinig ist, zählt sie –im Gegensatz zur gleichförmig geradlinigen Bewegung– zur Klasse der beschleunigten Bewegungen. Die Beschleunigung

$$\vec{a}(t) = \dot{\vec{v}}(t) = -r\omega^2 \begin{pmatrix} \cos \omega t \\ \sin \omega t \end{pmatrix} \quad (42.10)$$

ist zum Zentrum hin gerichtet (ZENTRIPETALBESCHLEUNIGUNG). Ihr Betrag ist

$$a = |\vec{a}| = r\omega^2 = \frac{v^2}{r}. \quad (42.11)$$

m

42.2. Kreisbewegung in drei Dimensionen

Ausgehend von der Beziehung

$$\begin{aligned} x &= r \cos \varphi \\ y &= r \sin \varphi \\ z &= z \end{aligned}$$

zwischen den Zylinderkoordinaten (r, φ, z) und den kartesischen Koordinaten (x, y, z) eines Punktes \vec{r} erhält man für konstanten Radius r , konstantes z und einen zeitlich veränderlichen Winkel $\varphi = \varphi(t)$:

$$\vec{v}(t) = \frac{d}{dt} \vec{r}(t) = r\dot{\varphi} \begin{pmatrix} -\sin \varphi(t) \\ \cos \varphi(t) \\ 0 \end{pmatrix} = \vec{\omega}(t) \times \vec{r}(t) \quad (42.12)$$

bzw. unter Auslassung des Zeitarguments

$$\vec{v} = \vec{\omega} \times \vec{r}. \quad (42.13)$$

Der Geschwindigkeitsvektor steht zu jedem Zeitpunkt normal zur Ebene, die vom Ortsvektor $\vec{r}(t)$ und vom Vektor $\vec{\omega} = \omega \vec{e}_a$ der Winkelgeschwindigkeit aufgespannt wird. Der Pseudoskalar ω ist die Winkelgeschwindigkeit. Die Richtung des axialen Vektors $\vec{\omega}(t)$ ist parallel zur Drehachse. Sie ist relativ zur gewählten Richtungsbasis durch einen Einheitsvektor \vec{e}_a spezifiziert. In dem von uns betrachteten einfachen Fall, wo \vec{e}_a parallel zur z -Achse liegt, gilt speziell

$$\vec{\omega} = \begin{pmatrix} 0 \\ 0 \\ \omega \end{pmatrix}. \quad (42.14)$$

Beispiel: Drehimpuls bei einer Kreisbewegung Bei einer Kreisbewegung ist $\vec{v} = \vec{\omega} \times \vec{r}$ und somit gilt für den Drehimpuls eines Massenpunktes bezüglich des Mittelpunkts der Kreisbewegung: $\vec{L} = m\vec{r} \times \vec{v} = mr^2\vec{\omega}$. Mit einem Drehschemel und einer Hantel kann man qualitativ veranschaulichen, wie sich eine Veränderung des Abstands der Hantel auf die Winkelgeschwindigkeit auswirken.

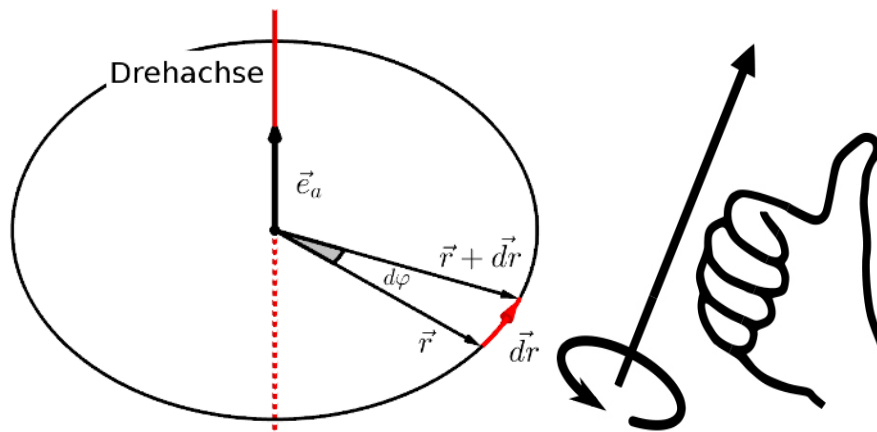


Abbildung 65: Skizze zur infinitesimalen Drehung. Der unterhalb der Ebene der Drehung befindliche Teil der Achse ist gestrichelt gezeichnet. Der Drehsinn ist positiv, wenn die Drehung relativ zum Blickrichtungsvektor \vec{e}_a gegen den Uhrzeigersinn erfolgt (Rechte-Hand-Regel).

42.3. Pseudoskalare und polare Vektoren

Offensichtlich hängt es von der Blickrichtung ab, ob man sagen würde, dass etwas links herum dreht, also gegen den Uhrzeigersinn, oder rechts herum (Abb. 65). Um den Drehsinn koordinatenunabhängig zu beschreiben, definiert man willkürlich einen in der Drehachse liegenden Einheitsvektor \vec{e}_a entlang einer der beiden möglichen Richtungen. Er legt sozusagen die Blickrichtung fest. Wenn man auf die Spitze des Vektors \vec{e}_a schaut und die differentielle Drehung um den Winkel $d\varphi$ gegen den Uhrzeigersinn verläuft, dann wird sie positiv gezählt. Man kann sich diese Konvention auch an der Regel der rechten Hand merken, wobei der Daumen in die Richtung des Achsenvektors zeigt und die gekrümmte Hand die positive Drehrichtung angibt (Abb. 65).

Bei einer Rauminversion dreht sich der die Blickrichtung beschreibende Vektor \vec{e}_a um und zugleich kehrt sich das Vorzeichen des infinitesimalen Winkels $d\varphi$ um. Das ist etwas Neues, denn wir haben mit $d\varphi$ nun eine einkomponentige Größe, die sich bei einer Inversion so verhält wie die Komponente eines Vektors und das Vorzeichen ändert. Andererseits bleibt $d\varphi$ bei einer Drehung der Richtungsbasis unverändert – im Gegensatz zur Komponente eines Vektors. Eine Größe, die sich in dieser Weise verhält, bezeichnet man als PSEUDOSKALAR.⁴²

Da sich Skalare um die Inversion einen Teufel scheren, kann man einen polaren Vektor mit einem Skalar multiplizieren, ohne dass er seinen Charakter als polarer Vektor ändert. Wenn man einen Vektor jedoch mit einem Pseudoskalar multipliziert wie das

⁴²Pseudoskalare haben also ein bisschen was von einem Skalar und ein bisschen was von einem Vektor, aber sie sind weder das eine noch das andere.

Tabelle 8: Verhalten physikalischer Größen unter Inversion und Drehung (dt Differential der Zeit, $d\varphi$ Differential des Winkels, \vec{v} Geschwindigkeit, $\vec{\omega}$ Winkelgeschwindigkeit).

	Skalar	Vektoren		Pseudovektor
		Pseudoskalar	polarer Vektor	
Beispiel	dt	$d\varphi$	\vec{v}	$\vec{\omega}$
Invarianz bei Inversion	ja	nein	nein	ja
Verhalten bei Inversion	$dt \longrightarrow dt$	$d\varphi \longrightarrow -d\varphi$	$\vec{v} \longrightarrow -\vec{v}$	$\vec{\omega} \longrightarrow \vec{\omega}$
Invarianz bei Drehung	ja	ja	nein	nein

beispielsweise bei der Definition der Winkelgeschwindigkeit

$$\vec{\omega} = \frac{d\varphi}{dt} \vec{e}_a, \quad (42.15)$$

der Fall ist, dann ist das Resultat zwar eine dreikomponentige Größe. Sie erfüllt auch die Axiome eines Vektors. Bei Rauminversion verhält sie sich jedoch anders als die bislang betrachteten (polaren) Vektoren.

42.4. Verhalten physikalischer Größen unter Drehung und Inversion

Die mathematischen Objekte, welche physikalische Größen darstellen, verhalten sich unterschiedlich, wenn man sie auf eine gedrehte bzw. invertierte Richtungsbasis bezieht. Als eigentliche SKALARE bzw. skalare Größen bezeichnet man solche, die sich weder unter Drehung noch Rauminversion ändern. Beispiele hierfür sind Invarianten wie das Zeitdifferential dt , die Masse m oder Größen, welche aus einem Skalarprodukt hervorgehen wie die kinetische Energie $E_{kin} = p^2/2m = \vec{p} \cdot \vec{p}/2m$.

Wenn man die Bezeichnung einer Sache mit der Vorsilbe „Pseudo“ versieht, dann meint man, dass sie nur so aussieht, als sei sie das Bezeichnete, es aber nicht ist. Pseudoskalare sind also keine Skalare und Pseudovektoren keine Vektoren im physikalischen Sinne. Vektoren wurden als Größen definiert, die durch Rauminversion ihr Vorzeichen wechseln (Kap. 7.1). Physikalische Größen, die sich unter einer Drehung wie Skalare verhalten (also einkomponentig darstellbar sind), bei einer Inversion sich aber als Vektor zu erkennen geben, werden als PSEUDOSKALARE bezeichnet.

Die Winkelgeschwindigkeit (Gl. 42.15), die sich aus dem Pseudoskalar $\omega = d\varphi/dt$ und dem polaren Vektor \vec{e}_a der Blickrichtung zusammensetzt, ist ein AXIALER VEKTOR bzw. ein Pseudovektor. Eine Übersicht über das unterschiedliche Verhalten dieser Größen bei Drehung oder Inversion eines physikalischen Systems gibt Tabelle 8.

Es ist offensichtlich, dass sowohl durch das Skalarprodukt als auch das Kreuzprodukt zweier polarer bzw. axialer Vektoren Größen entstehen, die invariant unter Inversion sind: im Fall des Skalarprodukts ist das Ergebnis ein Skalar und im Fall des Kreuzprodukts ein axialer Vektor. Bildet man hingegen das Skalarprodukt bzw. Kreuzprodukt eines polaren Vektors mit einem axialen Vektor, so entsteht ein Pseudoskalar bzw. ein polarer Vektor.

42.5. Die Vektoren der Winkelgeschwindigkeit und -beschleunigung

Wenn ein Partikel eine Kreisbewegung um die Achse a ausführt, so bewegt es sich in der Zeitspanne dt von seinem momentanen Punkt \vec{r} um $d\vec{r} = d\tilde{\varphi} \times \vec{r}$ zum Punkt $\vec{r} + d\vec{r}$ weiter. Seine Geschwindigkeit ist daher

$$\vec{v} = \dot{\vec{r}} = \frac{d\vec{r}}{dt} = \frac{d\tilde{\varphi}}{dt} \times \vec{r} = \vec{\omega} \times \vec{r} \quad (42.16)$$

Der Geschwindigkeitsbetrag $v = |\vec{\omega} \times \vec{r}|$ stellt die BAHNGESCHWINDIGKEIT dar, d.i. die Geschwindigkeit auf der Kreisbahn. Der axiale Vektor

$$\vec{\omega} = \frac{d\tilde{\varphi}}{dt} = \frac{d\varphi}{dt} \vec{e}_a \quad (42.17)$$

ist der Vektor der WINKELGESCHWINDIGKEIT. Seine zeitliche Änderung ist der Vektor der WINKELBESCHLEUNIGUNG

$$\dot{\vec{\omega}} = \frac{d^2\tilde{\varphi}}{dt^2} \quad (42.18)$$

43. Gravitations- und Trägheitskräfte in der Newtonschen Physik

Dieses Kapitel erläutert Gravitations- und Trägheitskräfte. Ihnen ist gemeinsam, dass in ihren Gesetzmäßigkeiten allein die Masse auftritt. Diese Gemeinsamkeit weist auf einen Zusammenhang zwischen Gravitations- und Trägheitskräften hin. Für Albert Einstein wurde diese Äquivalenz von Trägheits- und Gravitationskräften zum Schlüssel für die Ableitung einer relativistischen Gravitationstheorie, der Allgemeinen Relativitätstheorie. In diesem Kapitel soll erst einmal nur den Newtonschen Teil, d.h. den nichtrelativistischen Limes der Theorie nachgezeichnet werden.

43.1. Das Gravitationsgesetz

Johannes Kepler stellte drei berühmte Gesetze über die Planetenbahnen auf. Es handelt sich um empirische Gesetze, d.h. um phänomenologische Beschreibungen astronomischer Beobachtungen, die ursprünglich von Tycho Brahe gemacht wurden:

1. Die Planetenbahnen sind Ellipsen.
2. Die Verbindungslinie zwischen Sonne und Planet überstreicht in gleichen Zeiten gleiche Flächen.
3. Die Quadrate der Umlaufzeiten verhalten sich wie die Kuben der großen Halbachsen.

Um daraus das Newtonsche Gravitationsgesetz abzuleiten, wird ein punktförmiger Planet betrachtet, der als einziger Himmelskörper auf einer Kreisbahn um eine unendlich schwere Sonne läuft. Selbstverständlich gibt es so etwas nicht. Bei einem Modell kommt es nicht darauf an, ob es die im Modell gemachten Annahmen in der Wirklichkeit gibt, sondern nur darauf, dass das Modell den Keplergesetzen nicht widerspricht: Der Kreis ist ein Sonderfall der Ellipse und wie groß oder schwer Planet und Sonne sind, darüber setzen die Keplergesetze nichts voraus. Wir setzen daher ein Energiemodell mit kinetischer Energie $E_{kin} \propto v^2$ und einer potentieller Energie an, die für Kreise nur von dessen Radius r als phänomenologischem Parameter abhängen kann, d.h. $E_{pot} = E_{pot}(r)$. Aus dem zweiten Keplergesetz folgt, dass die Winkelgeschwindigkeit ω bzw. die Bahngeschwindigkeit $v = r\omega$ konstant ist. Die Summe von kinetischer Energie und der zu bestimmenden potentiellen Energie E_{pot} ist die konstante Gesamtenergie E_0 , d.h.

$$E_0 - E_{pot}(r) = E_{kin} \propto v^2$$

Nach Keplers drittem Gesetz verhalten sich die Quadrate der Umlaufzeiten $T = 2\pi/\omega$ wie die Kuben der Radien. Hieraus folgt, dass $r^3\omega^2$ eine Konstante ist bzw. $v^2 \propto r^2\omega^2 \propto 1/r$. Daraus folgt

$$E_{pot}(r) = E_0 - \frac{k_s}{r}$$

mit einem Proportionalitätsfaktor k_s , der nicht mehr von r abhängt. Da man in der Wahl der potentiellen Energie die Freiheit einer Konstanten hat, wählen wir sie so, dass $E_{pot}(r \rightarrow \infty) = 0$ und somit

$$E_{pot}(r) = -\frac{k_s}{r} \quad (43.1)$$

Indem man Gl. 43.1 gemäß Gl. 15.3 in radialer Richtung differenziert, erhält man das Newtonsche Gravitationsgesetz

$$F(r) = -\frac{dE_{pot}(r)}{dr} = -\frac{k_s}{r^2} \quad (43.2)$$

Die Sonne zieht also den Planeten an, und zwar mit einer in radialer Richtung wirkenden Anziehungskraft F mit einem quadratischen Abstandsgesetz. Wegen des Wechselwirkungsprinzips (actio=reactio, Kap. 15.2) übt der Planet umgekehrt eine gleich große Anziehungskraft auf die Sonne aus. Es ist naheliegend, die gleiche Gesetzmäßigkeit für die Bewegung des aus Mond und Erde bestehenden Systems einzufordern. Newtons geniale Idee war, diesen Gedanken noch radikaler weiterzuspinnen und so das „Gesetz des Himmels“ auf die Erde zu holen. Er postulierte, dass nicht nur der Mond, sondern alle Körper gemäß demselben Gesetz von der Erde angezogen werden, ja dass sogar jeder Körper A jeden beliebigen anderen Körper B gemäß Gl. 43.2 anzieht. Wenn A den Körper B anzieht und B gemäß dem Wechselwirkungsprinzip A , dann werden die Eigenschaften der beteiligten Körper, welche die Stärke der gegenseitigen Anziehung bestimmen, in symmetrischer Weise in die Konstante k_s eingehen. Das wird durch ein Gesetz der Form

$$F(r) = -G \frac{S^A S^B}{r^2} \quad (43.3)$$

mit $k_s = G s_A s_B$ geleistet, wobei die Größen s_A und s_B die für die Anziehung relevanten Stoffparameter beinhalten und G eine stoffunabhängige und noch frei wählbare Konstante ist. Die Stoffparameter s_A und s_B bezeichnet man als die SCHWERE MASSE von A bzw. B .

Wir betrachten nun die Anziehung zwischen der Erde und einem Körper A der Masse m_A . Er befinde sich in einer Entfernung $r = r_0 + h$ vom Mittelpunkt der Erde, wobei $h \ll r_0$ seine Höhe relativ zu einer Referenzfläche im Abstand r_0 vom Mittelpunkt der Erde sei. Seine potentielle Energie relativ zur Referenzfläche ist

$$E_{pot}(h) = -\frac{k_s}{(r_0 + h)} \approx -\frac{k_s}{r_0}(1 - h/r_0) = E_{pot}(0) + \frac{k_s}{r_0^2}h$$

mit $E_{pot}(0) = -k_s/r_0$. Ein Vergleich mit Gl. 12.35 zeigt, dass

$$g = g(r_0) = \left(\frac{s_A}{m_A}\right) \frac{G s_B}{r_0^2} \quad (43.4)$$

Die Abhängigkeit von r_0 zeigt an, dass g mit dem Abstand vom Erdmittelpunkt variiert.

43.2. Alle Körper fallen gleich schnell

Wir hatten gesagt, dass es denkbar wäre, dass $g \propto (s_A/m_a)$ eine stoffabhängige Größe ist, d.h. ein Materialparameter. Das kann man durch Experimente überprüft werden. Dazu kann man lokal (r_0 konstant) Experimente zum freien Fall ausführen (das System B ist dabei die Erde mit gleichbleibendem s_B). Um g beim freien Fall zu bestimmen, braucht man nur die durchfallene Höhe als Funktion der Zeit zu messen. Heute stehen uns Uhren zur Verfügung, mit denen die hier auftretenden kurzen Zeiten ausreichend genau gemessen werden können. Wenn man g in Wien durch hinreichend gute Messverfahren bestimmt, erhält man $g \approx 9.8 \text{ m/s}^2$.⁴³

Wenn es bei der experimentellen Überprüfung des freien Falls gelingt, alle Einflussfaktoren bis auf die gravitative Anziehung so weit zu reduzieren, dass sie das Messergebnis nicht mehr beeinflussen, dann zeigt sich, dass alle Körper gleich schnell fallen – egal aus was sie bestehen oder wie sie aufgebaut sind: Gl. 14.6 ist unabhängig von der Masse! Daraus folgt das bemerkenswerte Resultat, dass g überhaupt nicht von den Eigenschaften des fallenden Körpers A abhängt. Das geht nur, wenn $s_A \propto m_A$, also die schwere Masse proportional zur Masse ist. Wenn man denselben Körper auf der Erde und dann auf dem Mond bzw. anderen Planeten fallen lässt, hält man s_A fest und variiert s_B . Dabei erweist sich, dass $s_B \propto m_B$. Durch entsprechende Wahl des noch freien Parameters G kann man den Proportionalitätsfaktor zwischen der Masse und der schweren Masse gleich 1 setzen und erhält so das Newtonsche Gravitationsgesetz in seiner endgültigen Form:

$$\vec{F} = -G \frac{m_A m_B}{r^2} \vec{e}_r \quad (43.5)$$

⁴³Für Abschätzungen genügt es oft, mit $g = 10 \text{ m/s}^2$ zu rechnen. Das ist nur um 2% weniger genau.

Die Kraft \vec{F} wirkt auf die Masse m_A in radialer Richtung, d.h. in Richtung des radialen Einheitsvektors \vec{e}_r zur anziehenden Masse hin. Die GRAVITATIONSKONSTANTE G wurde von Cavendish und später erheblich präziser von Eötvös bestimmt.⁴⁴ Ihr Wert ist

$$G \approx 6.7 \times 10^{-11} \text{m}^3 \text{kg}^{-1} \text{s}^{-2}. \quad (43.6)$$

Für den Schwereparameter eines Körpers der Masse m_B , der andere Körper im Abstand r_0 gravitativ anzieht, ergibt sich:

$$g = \frac{Gm_B}{r_E^2}. \quad (43.7)$$

Beispiel: In Gl. 43.7 kann man die Masse $m_B = \frac{4}{3}\pi r_E^3 \rho_E$ der Erde durch ihr Volumen $\frac{4}{3}\pi r_E^3$ und ihre mittlere Dichte ρ_E ausdrücken, wobei $r_E \approx 6 \times 10^6 \text{m}$ der Erdradius ist. An der Erdoberfläche ist

$$g = G \frac{m_B}{r_E^2} = \frac{4\pi}{3} G \rho_E r_E \approx 10 \text{m/s}^2.$$

folgt $\rho_E \approx 6 \text{t/m}^3$ (Zum Vergleich: Die spezifische Dichte von Eisen beträgt ca. 8t/m^3).

43.3. Trägheitskräfte

Sei B irgendein Bezugssystem, das nicht notwendigerweise ein Inertialsystem sein muss. Nach Newton rufen dort auf ein Objekt der Masse m einwirkende Kräfte \vec{F}^B eine zeitliche Änderung des Impulses $\vec{F}^B = d\vec{p}^B/dt$ hervor. Wenn man der Einfachheit halber eine konstante Masse m voraussetzt, gilt:

$$m \frac{d^2 \vec{x}^B}{dt^2} = \vec{F}^B \quad (43.8)$$

Gl. 43.8 ist bei Newton als ein allgemeines Gesetz vorausgesetzt, wobei der hochgestellte Index B daran erinnern soll, dass sich alle Angaben dabei zunächst auf ein spezifisches Bezugssystem B beziehen.

Für jedes andere sich geradlinig und gleichförmig zu B bewegende Bezugssystem A folgt jedoch aus dem Galileischen Geschwindigkeitsadditionsgesetz, dass die Beschleunigung in A die gleiche wie in B ist. Damit sind die Kräfte Invarianten bezüglich dieser Transformation zwischen allgemeinen, aber gleichförmig zueinander bewegten Bezugssystemen.

Sei nun A aber ein Bezugssystem, das stattdessen relativ zu B mit der Beschleunigung \vec{a} beschleunigt ist.⁴⁵ Dann gilt dort das entsprechende Gesetz

$$m \frac{d^2 \vec{x}^A}{dt^2} = \vec{F}^A \quad (43.9)$$

⁴⁴Diese Experimente werden im Band „Einführung in die Physik 2“ besprochen.

⁴⁵Entsprechend ist dann B mit $-\vec{a}$ relativ zu A beschleunigt.

Aus der Relativbeschleunigung

$$\frac{d^2 \vec{x}^B}{dt^2} = \frac{d^2 \vec{x}^A}{dt^2} + \vec{a}$$

folgt nun aber, dass für die Kräfte folgendes Transformationsgesetz gelten muss:

$$\vec{F}^B = \vec{F}^A + m\vec{a} \quad (43.10)$$

Die Kräfte für zwei relativ zueinander beschleunigte Bezugssysteme unterscheiden sich. So weit gilt das Gesagte generell.

Mit dem ersten Newtonschen Gesetz, dem TRÄGHEITSGESETZ, zeichnet Newton eine Klasse von Bezugssystemen S dadurch aus, dass sich in ihnen ein Körper, der mit keinem anderen Körper wechselwirkt, geradlinig und gleichförmig bewegt. Dies definiert die KLASSE DER NEWTONSCHEN INERTIALSYSTEME. In ihnen ist die Kraft auf einen Körper, der an keiner Wechselwirkung teilnimmt, null. In allen anderen Bezugssystemen, die sich relativ zu den Inertialsystemen mit \vec{a} beschleunigt bewegen, wirkt eine zusätzliche Kraft $m\vec{a}$, die man TRÄGHEITSKRAFT nennt. In einem beschleunigten Bezugssystem bewegt sich ein ansonsten an keiner Wechselwirkung teilnehmender Körper unter dem Einfluss der Trägheitskraft. Die Ursache dieser Kraft ist – salopp gesagt – eine gravitative Wechselwirkung aufgrund einer Beschleunigung relativ zu den Massen des Kosmos.

43.4. Trägheitskräfte bei Drehbewegungen

B und B' seien zwei Bezugssysteme mit den Richtungsbasen $\{|1\rangle, |2\rangle, |3\rangle\}$ bzw. $\{|1'\rangle, |2'\rangle, |3'\rangle\}$ und den Koordinatenvektoren \vec{A} bzw. \vec{A}' für eine allgemeine physikalische Größe $|A\rangle$ bezüglich der jeweiligen Richtungs-Orthonormalbasen. Die durch die Koordinatenvektoren $\{\vec{e}_{x'}, \vec{e}_{y'}, \vec{e}_{z'}\}$ bezeichneten Einheitsvektoren sollen die Richtungsbasis von B' relativ zur Richtungsbasis von B darstellen. Es ist also:

$$\begin{aligned} |A\rangle &= A_1 |1\rangle + A_2 |2\rangle + A_3 |3\rangle = \sum_{i=1}^3 A_i |i\rangle \\ |A\rangle &= A'_1 |1'\rangle + A'_2 |2'\rangle + A'_3 |3'\rangle = \sum_{i=1}^3 A'_i |i'\rangle \end{aligned} \quad (43.11)$$

und beispielsweise für die x' - bzw. $1'$ -Achse

$$|x'\rangle = e_{x',1} |1\rangle + e_{x',2} |2\rangle + e_{x',3} |3\rangle = \sum_{j=1}^3 e_{x',j} |j\rangle$$

Daraus ergibt sich die Darstellung

$$|A\rangle = \sum_{i=1}^3 A'_i \sum_{j=1}^3 e_{x',j} |j\rangle + \sum_{i=1}^3 A'_i \sum_{j=1}^3 e_{y',j} |j\rangle + \sum_{i=1}^3 A'_i \sum_{j=1}^3 e_{z',j}(t) |j\rangle \quad (43.12)$$

und durch einen Vergleich mit Gl. 43.11 folgt für die Koordinatenvektoren:

$$\vec{A} = A'_1 \vec{e}_{x'} + A'_2 \vec{e}_{y'} + A'_3 \vec{e}_{z'} \quad (43.13)$$

Nehmen wir nun an, dass die Koordinatenvektoren zeitabhängig sind und sich auch die Achsen von B' relativ zu jenen von B zeitlich ändern. Dann gilt für Gl. 43.13 also explizit hingeschrieben

$$\vec{A}(t) = A'_1(t) \vec{e}_{x'}(t) + A'_2(t) \vec{e}_{y'}(t) + A'_3(t) \vec{e}_{z'}(t) \quad (43.14)$$

Die Zeitabhängigkeit des Koordinatenvektors $\vec{A}(t)$ kann also i.A. zwei Ursachen: erstens kann sie auf der Zeitabhängigkeit des Koordinatenvektors \vec{A}' beruhen und zweitens kann sie von der Bewegung des Bezugssystems B' relativ zu B herrühren. Wegen der Produktregel drückt sich das für die Zeitableitung (und damit für die Rate) der Größe $\vec{A}(t)$ entsprechend in zwei Beiträgen aus:

$$\begin{aligned} \dot{\vec{A}} &= \dot{A}'_1 \vec{e}_{x'} + \dot{A}'_2 \vec{e}_{y'} + \dot{A}'_3 \vec{e}_{z'} \\ &\quad + A'_1 \dot{\vec{e}}_{x'} + A'_2 \dot{\vec{e}}_{y'} + A'_3 \dot{\vec{e}}_{z'} \end{aligned}$$

Für die Beschleunigung erhält man entsprechend die drei Beiträge

$$\begin{aligned} \ddot{\vec{A}} &= \ddot{A}'_1 \vec{e}_{x'} + \ddot{A}'_2 \vec{e}_{y'} + \ddot{A}'_3 \vec{e}_{z'} \\ &\quad + 2\dot{A}'_1 \dot{\vec{e}}_{x'} + 2\dot{A}'_2 \dot{\vec{e}}_{y'} + 2\dot{A}'_3 \dot{\vec{e}}_{z'} \\ &\quad + A'_1 \ddot{\vec{e}}_{x'} + A'_2 \ddot{\vec{e}}_{y'} + A'_3 \ddot{\vec{e}}_{z'} \end{aligned}$$

Betrachtet man den Spezialfall, bei dem sich das Bezugssystem B' mit einer konstanten Winkelgeschwindigkeit $\vec{\omega}$ relativ zu B dreht, wobei die Nullpunkte der beiden Bezugssysteme zusammenfallen mögen, so gilt Gl. 42.16:

$$\begin{aligned} \dot{\vec{e}}_{x'} &= \vec{\omega} \times \vec{e}_{x'} \\ \ddot{\vec{e}}_{x'} &= \vec{\omega} \times (\vec{\omega} \times \vec{e}_{x'}) \end{aligned}$$

Entsprechendes gilt für die anderen beiden Einheitsvektoren. Daraus folgt

$$\dot{\vec{A}} = \dot{\vec{A}}' + \vec{\omega} \times \vec{A}'$$

und

$$\ddot{\vec{A}} = \ddot{\vec{A}}' + 2\vec{\omega} \times \dot{\vec{A}}' + \vec{\omega} \times (\vec{\omega} \times \vec{A}')$$

Ist \vec{A} speziell der Ortsvektor \vec{r} , so ist

$$\dot{\vec{r}} = \dot{\vec{r}}' + \vec{\omega} \times \vec{r}'$$

Für die Beschleunigung gilt:

$$\ddot{\vec{r}} = \ddot{\vec{r}}' + 2\vec{\omega} \times \vec{v}' + \vec{\omega} \times (\vec{\omega} \times \vec{r}') \quad (43.15)$$

mit der Geschwindigkeit $\dot{\vec{v}}'(t) = \dot{\vec{r}}'$ im System B'. Der letzte Beitrag auf der rechten Seite ist die im System B konstatierte Zentripetalbeschleunigung (s. Kap. V).

Wenn in B' eine Masse m im Punkt \vec{r}' ruht, dann führt diese Masse relativ zu B eine beschleunigte Bewegung aus. Newton würde fordern, dass diese Bewegung durch eine in B wirkende und an m angreifende Kraft zustande kommt. Aber warum sollte sie in B auftreten, während sie in B' nicht wirksam ist? Da die unterschiedlichen Beschleunigungen relativ zueinander sind und ein bloßes Beschreibungsartefakt der Bezugssysteme, sind sie keine Physik. Wenn Sie sich eine „einsame“ Punktmasse im leeren Weltall vorstellen, dann gibt es keinen Grund dafür, dass unterschiedliche Kräfte auftreten sollten, nur weil Bezugssysteme um die Masse rotieren oder nicht bzw. weil man sich das Objekt von unterschiedlichen Bezugssystemen her anschaut.

Aber das Weltall ist eben nicht leer. Durch die Fixsterne sind fixe Richtungen vorgegeben. Die kosmische Massenverteilung legt gewissermaßen ein natürliches Bezugssystem B fest, relativ zu dem die Masse m eine Kreisbewegung ausführt. Für eine Gravitationskonstante G , die ungleich null ist, treten durch die relative Drehung zueinander Kräfte auf, die als Trägheitskräfte bezeichnet werden. Wenn eine Masse sich nicht relativ zum natürlichen Koordinatensystem der Fixsterne dreht, treten auch keine Trägheitskräfte auf. Dieses Koordinatensystem ist das Inertialsystem im Newtonschen Sinn, und in ihm gilt die Newtonsche Bewegungsgleichung, d.h. die Kreisbewegung kommt dann nach Newton durch die Zentripetalkraft $\vec{F} = m\vec{\omega} \times (\vec{\omega} \times \vec{r})$ zustande.

Im mitrotierenden Bezugssystem B' ist diese weiterhin vorhanden. Da die Masse bezüglich des mitrotierenden Bezugssystem ruht, muss dort nach Newton eine in entgegengesetzter Richtung wirkende Kraft wirken. Das ist die ZENTRIFUGALKRAFT

$$\vec{F}_Z = -m\vec{\omega} \times (\vec{\omega} \times \vec{r}).$$

Entsprechend tritt dort auch die CORIOLISKRAFT

$$\vec{F}_C = -2m\vec{\omega} \times \vec{v}$$

auf. Die Corioliskraft ist stets orthogonal zur Drehachse und zur Geschwindigkeit gerichtet, während die Zentrifugalkraft stets orthogonal zur Drehachse gerichtet ist und von ihr weg weist.

Wenn man sich vorstellt, dass die Masse m in einem gegebenen Bezugssystem B' ruht und der Kosmos relativ zu B in entgegengesetzter Richtung rotiert, würden wegen der Relativität der Beschleunigung genau dieselben Trägheitskräfte auftreten (Machsches Prinzip). Das Auftreten der Masse in den Newtonschen Bewegungsgleichungen verweist gemäß diesem Gedankengang auf den gravitativen Ursprung der Trägheitskräfte.

43.5. Das Äquivalenzprinzip

Wir vergleichen die auf eine Punktmasse m wirkenden Kräfte einmal für die Situation, wo sie in einem Inertialsystem der Gravitationskraft $F = mg$ unterliegt und zum zweiten für

die Situation, wo man ihre Bewegung in einem Nicht-Inertialsystem beschreibt, das relativ zu einem Inertialsystem mit $a = g$ beschleunigt wird, aber wo keine Gravitation wirkt. Experimentell zeigt sich kein Unterschied bezüglich der physikalischen Wirkungen, d.h. die beiden physikalischen Situationen sind äquivalent und die Bewegung wird in beiden Systemen genau gleich ablaufen. Daher kann man die relativistischen Gravitationsgesetze durch eine Analyse gravitationsfreier beschleunigter Bezugssysteme analysieren. In den Vorlesungen über Allgemeine Relativitätstheorie werden Sie mehr darüber erfahren.

Die in beschleunigten Bezugssystemen auftretenden Trägheitskräfte sind ein empirisches Faktum. Warum gibt es sie? Aufgrund ihrer Äquivalenz zur Gravitation kann man bereits vermuten, dass sie sehr eng mit der Gravitation zusammenhängen. Sie erweisen sich letztendlich als ein gravitativer Beitrag der kosmischen Massen. Die Allgemeine Relativitätstheorie führt uns zu einer Kosmologie, d.h. zu einem physikalischen Verständnis des Kosmos und zu einer physikalischen Version der „Schöpfungsgeschichte“.

43.6. Symmetrien

Isotropie und Homogenität des Raumes sorgen dafür, dass Drehungen und Verschiebungen SYMMETRIETRANSFORMATIONEN sind. Das bedeutet, dass zwei Systeme physikalisch deckungsgleich bzw. äquivalent bzw. symmetrisch sind, wenn sie sich nur durch eine Verschiebung bzw. eine Drehung voneinander unterscheiden.

Neben diesen beiden Symmetrien gibt es noch eine Reihe anderer Symmetrien, die für die Physik von großer Bedeutung sind. An dieser Stelle möchte ich nur auf das Beispiel der Spiegelsymmetrie kurz eingehen: Betrachtet man beispielsweise Systeme, die längen- und winkeltreu gleich sind, dann lassen sich einige davon durch Verschiebungen und Drehungen zur Deckung bringen, aber nicht alle. Die anderen lassen sich jedoch durch eine zusätzliche Spiegelung zur Deckung bringen. Im Rahmen der Mechanik kann man für spiegelsymmetrische Systeme keinen Unterschied bezüglich der Abläufe physikalischer Vorgänge feststellen. Daher ist die Mechanik bezüglich der Spiegelung symmetrisch.

44. Nichtlineare Koordinatentransformationen (M4)

Kugelkoordinaten und Zylinderkoordinaten teilen mit den kartesischen Koordinaten die Eigenschaft, dass sich ihre Koordinatenlinien, d.h. die Linien, bei denen alle Koordinaten bis auf eine konstant gehalten werden, sich in jedem Punkt lokal rechtwinklig schneiden. Hier werden nur einige Beziehungen für Kugelkoordinaten hergeleitet, denn die entsprechenden Beziehungen für Zylinderkoordinaten ergeben sich analog (sie sind sogar einfacher herzuleiten).

Transformation zwischen kartesischen Koordinaten und Kugelkoordinaten. Bis hier hin traten nur lineare Koordinatentransformationen auf. Ein Beispiel für eine nichtlineare Koordinatentransformation ist die Transformation zwischen kartesischen Koordinaten und Kugelkoordinaten. In der Mathematik geht man meist von der in Abb. 66 gezeigte

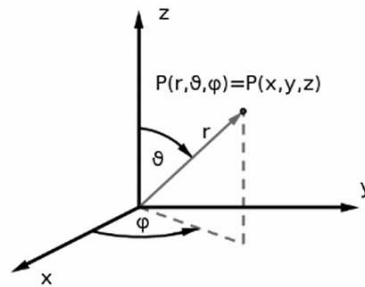


Abbildung 66: Beziehung zwischen kartesischen Koordinaten und Kugelkoordinaten

rechtshändige Standardanordnung der beiden Bezugssysteme aus, die allen drei wesentlichen Bezugselementen übereinstimmen:

1. Für das kartesische Koordinatensystem wird eine Bezugsrichtung als y -Achse gewählt. Der Nullpunkt wird so gewählt, dass er gemeinsamer Bezugspunkt für die y -Koordinate und die Radialkoordinate r ist.
2. Eine zur y -Richtung orthogonale Richtung wird als kartesische x -Achse gewählt und ist gemeinsame Bezugsrichtung für die x -Koordinate und den Azimutwinkel ϕ .
3. Die kartesische z -Achse wird orthogonal zur von der x - und y -Achse aufgespannten Horizontal- bzw. Äquatorialebene gewählt. Sie ist die gemeinsame Bezugsrichtung für die z -Koordinate und den Polarwinkel ϑ .
4. Die kartesischen Achsenrichtungen werden so gewählt, dass sie rechtshändig sind und alle Winkel mathematisch positiv.

Aus 28 ergibt sich für die Umrechnung der Kugelkoordinaten (r, ϑ, ϕ) eines Punktes P zu seinen kartesischen Koordinaten (x, y, z) die Koordinatentransformation

$$x(r, \vartheta, \phi) = r \sin \vartheta \cos \phi \quad (44.1)$$

$$y(r, \vartheta, \phi) = r \sin \vartheta \sin \phi \quad (44.2)$$

$$z(r, \vartheta, \phi) = r \cos \vartheta \quad (44.3)$$

Für Kreiskoordinaten in einer Ebene setzt man $\vartheta = 90^\circ$ und erhält die Koordinatentransformation

$$x = r \cos \phi \quad (44.4)$$

$$y = r \sin \phi \quad (44.5)$$

bezüglich eines Bezugssystems S. Wenn man von einem Bezugssystem S' ausgeht, das um den Winkel φ relativ zu S gedreht ist, dann ist hier $\phi' = \phi - \varphi$ einzusetzen, d.h.

$$x' = r \cos \phi' = r \cos(\phi - \varphi) = r [\cos \phi \cos \varphi + \sin \phi \sin \varphi] = \cos \varphi x + \sin \varphi y \quad (44.6)$$

$$y' = r \sin \phi' = r \sin(\phi - \varphi) = r [\sin \phi \cos \varphi - \cos \phi \sin \varphi] = \cos \varphi y - \sin \varphi x \quad (44.7)$$

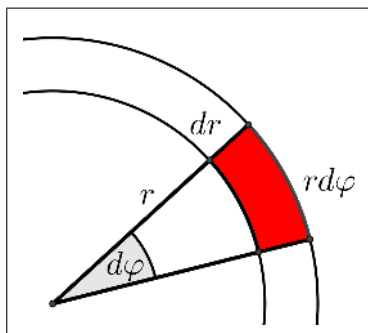


Abbildung 67: Differentielles Flächenelement $dr r d\varphi$ in polaren Koordinaten für die Berechnung einer Kreisfläche.

Dies ist wiederum die aus den Gln. 5.2 und 5.3 bekannte Koordinatentransformation zwischen gedrehten kartesischen Bezugssystemen.

Länge (Linienintegral). Aus den Gln. 44.1-44.3 erhält man die Längenelemente durch partielle Differentiation, beispielsweise

$$dx = \frac{\partial x}{\partial r} dr + \frac{\partial x}{\partial \vartheta} d\vartheta + \frac{\partial x}{\partial \phi} d\phi = \sin \vartheta \cos \phi dr + r \cos \vartheta \cos \phi d\vartheta - r \sin \vartheta \sin \phi d\phi.$$

Fläche (Flächenintegral). Um die Fläche eines Kreises mit dem Radius r_K zu berechnen, kann man Polarkoordinaten (r, φ) mit dem in Abb. 67c skizzierten differentiellen Flächenelement $d\mathcal{A} = d^2r = dr r d\varphi$ heranziehen:

$$\mathcal{A} = \int_0^{r_K} \int_0^{2\pi} dr r d\varphi = \int_0^{2\pi} d\varphi \int_0^{r_K} r dr = 2\pi \frac{1}{2} r_K^2 = \pi r_K^2$$

Volumen (Volumenintegral). Orthogonale Koordinatensysteme sind solche, bei denen sich die Koordinatenlinien rechtwinklig schneiden. Zu den drei wichtigsten orthogonalen Koordinatensystemen gehört das kartesische Koordinatensystem mit geradlinigen Koordinaten und die Systeme mit Zylinderkoordinaten und Kugelkoordinaten, die krummlinigen Koordinaten haben. In Tab. 10 die Längen- und Flächenelemente für die wichtigsten drei orthogonalen Koordinatensysteme angegeben.

Es ist also beispielsweise:⁴⁶

Das Volumen kann man damit auf unterschiedliche Weise berechnen:

$$V = \int dV = \iiint d^3r = \iiint dx dy dz = \iiint \varrho d\varrho d\varphi dz = \iiint r^2 dr \sin \vartheta d\vartheta d\varphi. \quad (44.8)$$

⁴⁶Zur Unterscheidung verwende ich im Folgenden die Glyphe ρ für die Dichte und die Glyphe ϱ für die Radialkoordinate eines zylindrischen Koordinatensystems. Beide bezeichnen jedoch den gleichen „rho“ ausgesprochenen griechischen Buchstaben, und man muss bei mündlicher Kommunikation daher aus dem Kontext erschließen, was gemeint ist.

Tabelle 9: Längenelemente und das Volumenelement für die drei wichtigsten orthogonalen Koordinatensysteme

Koordinatensystem	Koordinaten	Längenelemente			Volumenelement
kartesische Koord.	(x, y, z)	dx	dy	dz	$dx dy dz$
Zylinderkoord.	(ϱ, φ, z)	$d\varrho$	$\varrho d\varphi$	dz	$\varrho d\varrho d\varphi dz$
Kugelkoord.	(r, ϑ, φ)	dr	$r \sin \vartheta d\varphi$	$r d\vartheta$	$r^2 \sin \vartheta dr d\vartheta d\varphi$

Tabelle 10: Längen- und Flächenelemente für die drei wichtigsten orthogonalen Koordinatensysteme

Koordinatensystem	Längenelemente			Flächenelemente		
kartesische Koord.	dx	dy	dz	$dy dz$	$dx dz$	$dx dy$
Zylinderkoord.	$d\varrho$	$\varrho d\varphi$	dz	$\varrho d\varphi dz$	$d\varrho dz$	$\varrho d\varrho d\varphi$
Kugelkoord.	dr	$r \sin \vartheta d\varphi$	$r d\vartheta$	$r^2 \sin \vartheta d\vartheta d\varphi$	$r dr d\vartheta$	$r \sin \vartheta dr d\varphi$

Bei der Ausführung dieser Dreifachintegrale bezüglich der unterschiedlichen Koordinatensysteme müssen die Integrationsgrenzen entsprechend gewählt werden.

Beispiel: Volumen von Rotationskörpern. Manchmal ist die Querschnittsfläche eines Körpers als Funktion einer Koordinate bekannt. Dann kann das Dreifachintegral auf ein einfaches Integral zurückgeführt werden. Das ist z.B. für rotationssymmetrische Körper der Fall. Hier ist es vorteilhaft, Zylinderkoordinaten zu wählen und die z -Achse entlang der Symmetrieachse zu legen. Das Volumen kann dann durch die Integration von scheibenförmigen Flächenelementen $\pi y^2(z) dz$ bestimmt werden:

$$V = \iiint \varrho d\varrho d\varphi dz = \int_{z_i}^{z_f} dz \int_0^{y(z)} \varrho d\varrho \int_0^{2\pi} d\varphi = \pi \int_{z_i}^{z_f} y^2(z) dz \quad (44.9)$$

Hierbei bezeichnet $y(z)$ den Radius des scheibenförmigen Elements an der Stelle z , z_i ist die z -Koordinate des Anfangspunkts des Körpers und z_f diejenige seines Endpunkts. Bei einer Kugel mit dem Radius r_K ist beispielsweise $y^2 = r_K^2 - z^2$, $z_i = -r_K$ und $z_f = r_K$. Daraus folgt durch einfache Integration von Gl. 44.9 für das Kugelvolumen: $V = \frac{4}{3}\pi r_K^3$.

Masse. Gl. 18.1 bleibt auch im Grenzfall infinitesimal kleiner Massenelemente dm und Volumenelemente dV richtig, d.h. man kann mittels

$$dm = \rho dV \quad (44.10)$$

von einem Massenelement auf ein Volumenelement umrechnen. Man kann also das Integral über Massenelemente in ein Integral über Volumenelemente konvertieren:

$$m = \int_m dm = \int_V \rho dV = \iiint \rho d^3r.$$

45. Geometrie eines Körpers

Form, Länge, Fläche und Volumen gehören zu den primären Größen, welche einen makroskopischen Körper beschreiben. Diese Beschreibung wird i.A. relativ zu einem Koordinatensystem ausgeführt. Bezüglich der Techniken zur Berechnung sei auf Lehrbücher der Mathematik verwiesen. Hier soll nur auf einige konzeptionellen Ideen eingegangen werden, die hinter solchen Berechnungen stecken.

Form. Um die Form eines Körpers zu charakterisieren, müssen wir sagen können, was zum Körper gehört, also innen ist, und was außen, also nicht dazu gehört. Das kann man durch die explizite Angabe von Punktemengen tun. Beispielsweise stellt die Punktemenge

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leq 1$$

ein Ellipsoid mit den Halbachsen a, b, c dar (Abb. 68a).

Eine wichtige Angabe ist der Rand bzw. die Oberfläche. Für eine Linie ist die Berandung durch die beiden Endpunkte gegeben, für eine Fläche durch eine Kurve und für ein Volumen geschieht dies durch Angabe seiner Oberfläche, also indem man alle Koordinatenpunkte angibt, welche zu seiner Oberfläche gehören. Für das Ellipsoid ist der Rand durch

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

gegeben, d.h. durch die Punkte der Oberfläche. Wenn man einen Schnitt durch diese Oberfläche legt, beispielsweise einen Schnitt mit $z = 0$, so erhält man eine Kurve, hier die durch

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{45.1}$$

dargestellte Ellipse. Sie wird für $a = b$ zu einem Kreis.

Linien kann man auch durch eindimensionale Parameterdarstellungen angeben, wie das bei der durch Gl. 40.5 dargestellte Ortskurve im Raum gemacht wurde. Für Flächen benötigt man entsprechend zweidimensionale Parameterdarstellungen.

Wir können eine Fläche im Dreidimensionalen aber auch dadurch spezifizieren, indem wir durch eine Funktion F mit $F(x, y, z) = 0$ eine der beiden Koordinaten durch die beiden anderen festlegen. Für die Oberfläche eines Ellipsoids wäre das beispielsweise die Funktion

$$F(x, y, z) = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1$$

Wenn man noch eine zweite solche Bedingungsgleichung $G(x, y, z) = 0$ mit einer Funktion G aufstellt und die beiden so beschriebenen Flächen eine Schnittmenge haben, wird dadurch eine Kurve im Raum beschrieben.

Länge (Linienintegral). Eine Länge misst man, indem man ein Lineal anlegt. Nun sind Lineale oft starre Maßstäbe. Wenn man damit die Länge einer geraden Strecke messen möchte, macht das keine Probleme. Aber was soll man tun, wenn man es mit einer gewundenen Straße oder einer gekrümmten Kurve im Raum zu tun hat? Da bleibt nichts anderes übrig, als die Strecke l in kleine Streckenabschnitte zu zerlegen, die man durch kleine Lineale in guter Näherung bestimmen kann. Die Gesamtstrecke erhält man, indem man die einzelnen Längenmessungen aufaddiert. Wenn man das mit differentiell kleinen Längenelementen dl tut, dann wird dies durch das bestimmte Integral

$$l = \int_A^E dl \quad (45.2)$$

geleistet. Man integriert die Streckenelemente dl vom Anfangspunkt A bis zum Endpunkt E .

Beispiel: Es seien die Punkte $P(x, y)$ der Kurve $y = \cosh x$ in der x - y -Ebene gegeben. Wie lang ist die Strecke vom Anfangspunkt $A = (0, 1)$ zum Punkt $E = (\ln 2, 1\frac{1}{4})$? Im euklidischen Raum ist das Längenelement durch $dl = \sqrt{dx^2 + dy^2}$ gegeben. Durch Differenzieren der Kurve erhält man $dy = \sinh x dx$ und mit $dl = \sqrt{1 + \sinh^2 x} dx$ ergibt sich:⁴⁷

$$l = \int_{(0,0)}^{(1,1)} dl = \int_0^{\ln 2} \sqrt{1 + \sinh^2 x} dx = \int_0^{\ln 2} \cosh x dx = \sinh(\ln 2) = 3/4$$

Fläche (Flächenintegral). Wie man die Fläche für die geometrischen Grundelemente Rechteck, Dreieck und Kreis berechnet oder das Volumen von Quader und Kugel, soll hier als bekannt vorausgesetzt sein. Die Berechnung von Flächen bzw. Volumina allgemeinerer Objekte führt man auf diese Grundelemente zurück und berechnen Sie mit den Methoden der Integralrechnung.

Für kartesische Koordinaten zerlegt man eine Fläche in differentielle Flächenelemente $dA = dx dy$ und schreibt für die Flächenberechnung symbolisch

$$A = \int dA = \iint dx dy. \quad (45.3)$$

Das DOPPELINTEGRAL „ \iint “ steht symbolisch dafür, dass zwei Summationen über infinitesimale Elemente zu bilden sind, nämlich einmal in y -Richtung, wo die Flächenelemente $dx dy$ zu Streifen aufintegriert wird, welche die feste Breite dx haben. Anschließend integrieren man die Fläche aller Streifen entlang der x -Richtung.

Beispiel 1: Fläche unter einer Kurve. Es sei die Fläche A unter einer Kurve $y = f(x)$ von $x = a$ bis $x = b$ zu berechnen (Abb. 68b).

⁴⁷Bei der Rechnung wurde verwendet: $\cosh x = (e^x + e^{-x})/2$ sowie $\cosh^2 x - \sinh^2 x = 1$.

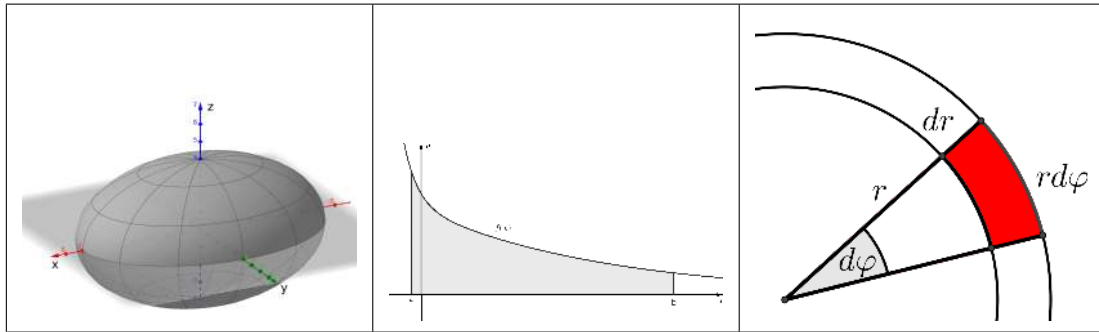


Abbildung 68: (a) Ellipsoid mit $a = 7$, $b = 5$ und $c = 4$. (b) Berechnung der Fläche unter einer Kurve durch eine Integration. (c) Differenzielles Flächenelement $dr r d\varphi$ in polaren Koordinaten für die Berechnung einer Kreisfläche.

$$\mathcal{A} = \iint d^2r = \iint dx dy = \int_a^b dx \int_0^{f(x)} dy = \int_a^b f(x) dx \quad (45.4)$$

Hier ist d^2r die Notation für ein allgemeines Flächenelement, das aus zwei Längenelementen besteht. Im Falle eines kartesischen Koordinatensystems gilt $d^2r = dx dy$. Im letzten Ausdruck von Gl. 45.4 wird das Doppelintegral auf ein Einfachintegral zurückgeführt, für das die Lösungsmethoden bekannt sind.

Volumen (Volumenintegral) Bei der Berechnung von Volumina wendet man die für die Fläche gelernten Methoden analog an: Man summiert über infinitesimale Volumenelemente auf.

$$V = \int dV = \iiint d^3r = \iiint dx dy dz. \quad (45.5)$$

Mit d^3r bezeichnen wir wiederum ein allgemeines Volumenelement im Dreidimensionalen.

46. Drehung makroskopischer Körper

46.1. Schwerpunktsystem: Schwerpunkt und Eigendrehimpuls

Ein aus n Massen m_i mit $i = 1, \dots, n$ bestehendes abgeschlossenes System charakterisiert man durch additive Erhaltungsgrößen: Neben der Masse $m = \sum m_i$ und dem Impuls $\vec{p}_S = \sum \vec{p}_i$ des Systems gehört auch der Drehimpuls dazu. Wenn man jedoch vom DREHIMPULS EINES SYSTEMS sprechen möchte, gibt es eine Schwierigkeit: Der Drehimpuls ist, wie aus Kap. 41.2 bekannt, von der Wahl des Bezugspunkts abhängig. Zur Definition des Drehimpulses eines Systems gehört also immer die Festlegung des Bezugspunkts hinzu. Dieser spezielle Bezugspunkt ist der SCHWERPUNKT \vec{r}_S . Im nichtrelativistischen

Grenzfall des relativistischen Drehimpulses ergibt sich der Schwerpunkt aus dem MASSENMOMENT $\sum m_i \vec{r}_i$:

$$\vec{r}_S = \frac{1}{m} \sum_{i=1}^n m_i \vec{r}_i \quad (46.1)$$

Ein Ruhesystem, dessen Nullpunkt im Schwerpunkt verankert ist, bezeichnet man als SCHWERPUNKTSYSTEM. Wenn man die zeitliche Ableitung von Gl. 46.1 bildet, ergibt sich für die Impulserhaltung $\vec{p} = \sum \vec{p}_i = 0$. Folglich ist das Schwerpunktsystem ein Spezialfall eines Ruhesystems und damit auch ein Inertialsystem (Kap. 11.5).⁴⁸ Mit diesem Bezugspunkt ist der Drehimpuls des Gesamtsystems

$$\vec{L}_S = \sum_{i=1}^n (\vec{r}_i - \vec{r}_S) \times (\vec{p}_i - \vec{p}_S) = \sum_{i=1}^n \vec{r}_i \times (\vec{p}_i - \vec{p}_S) \quad (46.2)$$

eindeutig und wird als EIGENDREHIMPULS des Systems bezeichnet. Eine Galaxie kann z.B. näherungsweise als ein abgeschlossenes System aufgefasst werden. Wenn man die Position aller Sterne und ihre Impulse kennt, ergibt sich aus Gl. 46.2 der Eigendrehimpuls der Galaxie. Verankert man den Nullpunkt eines Inertialsystems im Schwerpunkt, so ist $\vec{r}_S = \vec{p}_S = 0$ und daher ist

$$\vec{L}_S = \sum_{i=1}^n \vec{r}_i \times \vec{p}_i.$$

Der Drehimpuls bezüglich des Schwerpunktsystems ist gleich dem Eigendrehimpuls.

46.2. Rotationsenergie

Das Modell des starren Körpers. Im Folgenden wird das Modell eines Systems untersucht, bei dem alle Systembestandteile bei einer Drehbewegung mit der gleichen Winkelgeschwindigkeit $\vec{\omega}$ rotieren. Ein solches Modellsystem wird als STARRER KÖRPER bezeichnet. Für einen Körper, der wirklich starr wäre, d.h. bei dem sich die relativen Abstände seiner Bestandteile unter keinen Umständen ändern, wäre es exakt, aber reale Körper sind nie wirklich starr und die Idealisierung durch das Modell des starren Körpers ist oft „cum grano salis“ zu nehmen. Wenn man es einsetzen möchte, muss man immer zuerst abwägen, ob das Modell für die zu lösende Problemstellung eine adäquate Beschreibung ist.

Das Modell sollte man auch deshalb nicht allzu ernst nehmen, weil es nämlich gewissermaßen „Hardcore Nicht-Relativistik“ ist. Wenn man nämlich die Vorderfront eines starren Körpers bewegt, würde sich wegen der vorausgesetzten Starrheit zugleich seine Rückfront bewegen und damit wäre eine Signalgeschwindigkeit möglich, die größer als die Raum/Zeit-Konstante wäre. Da die Winkelgeschwindigkeit aller Systemteile konstant ist, würde die Geschwindigkeit $\vec{v} = \vec{\omega} \times \vec{r}$ mit zunehmenden Radius \vec{r} irgendwann die Raum/Zeit-Konstante c überschreiten. Ferner werden mit dem Eigendrehimpuls und dem

⁴⁸Das Ruhesystem wurde in Kap. 11.5 definiert. Anders als beim Schwerpunktsystem war der Nullpunkt des Inertialsystems dabei noch nicht festgelegt.

Schwerpunkt Begriffe eingeführt, in denen die Ortskoordinaten explizit auftreten, was in dieser Form mit der Relativitätstheorie unvereinbar ist.⁴⁹

Dennoch kann man das Modell auf überraschend viele reale Systeme anwenden. Beispielsweise kann man auch die Erde für gewisse Fragestellungen in guter Näherung als starren Körper auffassen – inklusive der Bäume, die auf ihr wachsen und deren Blätter, inklusive der Ozeane und sogar inklusive der Luft unseres Hörsaalraums, denn relativ zur Geschwindigkeit in der Größenklasse von 1000 km/h, mit welcher sich die Erdoberfläche auf unserem Breitengrad relativ zu ihrem momentanen Schwerpunktsystem bewegt, sind die Bewegungen der Bäume, Blätter, Gewässer oder der erdnahen Luft vernachlässigbar. Wenn man ein Honigglas oder eine Kaffeetasse nicht allzu schnell dreht, wird sich auch der darin befindliche Honig bzw. Kaffee zuletzt mit der gleichen Winkelgeschwindigkeit drehen wie das Glas oder die Tasse und man kann das Ganze dann ebenfalls als einen starren Körper beschreiben, obwohl am Honig und am Kaffee nichts wirklich starr ist.

Rotationsenergie und Trägheitsmoment. Bezüglich eines Ruhesystems rotiere ein starrer Körper mit der konstanten Winkelgeschwindigkeit $\vec{\omega}$ um eine Achse a . Ein Massenelement dm im Abstand ϱ_a von der Achse bewegt sich also mit einer Bahngeschwindigkeit $v = \varrho_a \omega$ und trägt mit $dE_{rot} = \frac{1}{2} dm v^2 = \frac{1}{2} (\varrho_a^2 dm) \omega^2$ zur kinetischen Energie der Rotationsbewegung bei. Man bezeichnet diese speziell auf die Rotation bezogene kinetische Energie als ROTATIONSENERGIE und erhält sie für das Gesamtsystem, indem man die Beiträge der einzelnen Massenelemente aufsummiert bzw. aufintegriert:

$$E_{rot} = \frac{1}{2} \Theta_a \omega^2 \quad (46.3)$$

Die Größe

$$\Theta_a = \int_V \varrho_a^2 dm \quad (46.4)$$

nennt man TRÄGHEITSMOMENT bezüglich der Achse a . Die Integration erfolgt über das Volumen V des Körpers.

Es sei b eine zu a parallele Achse, $d = |\vec{d}|$ der Abstand der Achsen und \vec{d} der orthogonal zu den Achsen stehende Abstandsvektor. Wegen $\vec{\varrho}_b = \vec{d} - \vec{\varrho}_a$ gilt

$$\Theta_b = \Theta_a + md^2 - 2\vec{d} \cdot \int_V \vec{\varrho}_a dm. \quad (46.5)$$

Der Schwerpunkt \vec{r}_S ist wichtig für die Zerlegung zwischen translatorischer kinetischer Energie und Rotationsenergie, die zwei unabhängige Möglichkeiten bzw. Freiheitsgrade der kinetischen Energie repräsentieren. Bei der Berechnung des Schwerpunkts

⁴⁹In der Relativitätstheorie sind Schwerpunkt und Eigendrehimpuls zum relativistischen Drehimpulstensor verknüpft.

$$\vec{r}_S = \frac{1}{m} \int_V \vec{r} dm \quad (46.6)$$

kann man den Ortsvektor \vec{r} von einem Bezugspunkt auf der Achse a aus berechnen und ihn in die Komponente \vec{r}_{\parallel} parallel zur Achse und in die Komponente $\vec{\varrho}_a$ orthogonal zur Achse zerlegen, d.h. $\vec{r} = \vec{r}_{\parallel} + \vec{\varrho}_a$. Dann gilt $\vec{d} \cdot \vec{r}_S = \frac{1}{m} \vec{d} \cdot \int_V \vec{\varrho}_a dm$ und somit

$$\Theta_b = \Theta_a + m d^2 - 2m \vec{d} \cdot \vec{r}_S. \quad (46.7)$$

Das Trägheitsmoment ist am kleinsten, wenn $\vec{d} \cdot \vec{r}_S = d r_S$. Das ist der Fall, wenn b eine Achse ist, die durch den Schwerpunkt geht. Dieser Anteil der Rotationsenergie im Schwerpunktsystem ist ein unabhängiger Freiheitsgrad der inneren Energie \mathcal{U} eines abgeschlossenen Systems.⁵⁰

Berechnung von Trägheitsmomenten. Wenn man das Trägheitsmoment Θ_S bezüglich einer durch den Schwerpunkt gehenden Achse kennt, so kann es mit Hilfe von Gl. 46.5 für eine beliebige Achse berechnet werden. Es genügt daher, für einen ausgedehnten Körper das Trägheitsmoment Θ_S bezüglich des Schwerpunkts zu berechnen. Aus Symmetriegründen bietet es sich an, diese Rechnung in Zylinderkoordinaten (ϱ, φ, z) durchzuführen, wobei die z -Achse in die Drehachse gelegt wird und somit die Radialkoordinate $\varrho = \varrho_S$ ist. Für homogene Körper ist die Dichte ρ konstant und daher folgt

$$\Theta_S = \int_V \varrho^2 dm = \rho \iiint_V \varrho^3 d\varrho d\varphi dz. \quad (46.8)$$

Für einen Kreiszylinder vom Radius r der Dicke $\Delta r \ll r$, einer Höhe h und der Masse $m = 2\pi r \Delta r h \rho$ ergibt sich direkt:

$$\Theta_S \approx 2\pi r^3 \Delta r h \rho = m r^2. \quad (46.9)$$

Für einen Zylinder bzw. eine Scheibe muss man die Beiträge infinitesimaler Kreiszylinder von 0 bis r aufintegrieren und erhält

$$\Theta_S = \frac{1}{2} \pi r^4 h \rho = \frac{1}{2} m r^2. \quad (46.10)$$

Für eine Halbkugel mit Radius r integriert man Scheiben der Dicke dz mit einem Radius $\sqrt{r^2 - z^2}$ von $z = 0$ bis $z = r$ auf und erhält

$$\Theta_S = \frac{1}{2} \rho \int_0^r \pi (r^2 - z^2)^2 dz = \frac{1}{2} \rho \pi \left[r^5 - \frac{2}{3} r^5 + \frac{1}{5} r^5 \right] = \frac{1}{5} m r^2. \quad (46.11)$$

Das Trägheitsmoment einer Vollkugel ist doppelt so groß und somit ist $\Theta_S = \frac{2}{5} m r^2$.

⁵⁰ Und somit auch ein Beitrag zu dessen Masse. Die Summe der Impulse ist im Schwerpunktsystem bekanntlich Null.

Teil VI.

Anhang

A. Newtons Hypothese der absoluten Zeit

Die Newtonsche Mechanik ist auf der Hypothese aufgebaut, dass die Zeit „absolut“ ist. Das bedeutet, dass sie von keinen äußeren Umständen abhängt. Insbesondere kann Newtons absolute Zeit dahin gedeutet werden, dass die auf Uhren angezeigte Zeit nicht von den Bewegungsumständen der Uhren abhängt, d.h. dass sie nicht davon abhängt, ob die Uhr bewegt wird oder nicht. Diese Hypothese besticht zunächst einmal durch ihre Einfachheit. Sie ist auch zulässig, denn sie erfüllt offensichtlich auch das Relativitätsprinzip. Führt es zur richtigen Naturbeschreibung, wenn man von den physikalisch einfachsten der zulässigen Hypothesen ausgeht? Ist das Einfachste zugleich das Richtige?

Überraschenderweise scheint das in der Physik tatsächlich meistens der Fall zu sein. Das liegt aber oft daran, dass man sich dafür entscheidet, sich auf die einfachste Theorie festzulegen, so lange sie nicht falsifiziert worden ist. Warum soll man sich mit schwierigeren Theorien herumplagen, wenn sie im Rahmen der Messunsicherheit gleich richtig sind, d.h. wenn die schwierigere Theorie nicht mehr leistet als die einfachere. Wenn mehrere Theorien die Fakten erklären, sollte man nach der Auffassung von Heinrich Hertz bzw. Ernst Mach die einfachste und somit denkökonomischste Theorie als die paradigmatische Theorie bzw. als das Standardmodell auswählen.

Die Newtonsche Hypothese der absoluten Zeit kann man einem experimentellen Test unterwerfen. Beispielsweise kann man eine Taschenuhr mit der Geschwindigkeit v gleichförmig von B nach C und wieder zurück transportieren. Dann sollte nach Newton die auf der mit Taschenuhr angezeigte Dauer $\tau(v)$ für die Rundreise exakt gleich der Dauer $\tau(0)$ sein, die man nach der Rückkehr für die Reise auf der lokalen Standuhr in B abliest.

Mit den experimentellen Möglichkeiten, die zur Zeit von Newton zur Verfügung standen, ergab sich im Rahmen der Messunsicherheit tatsächlich, dass $\tau(v) = \tau(0)$ und dabei blieb es auch für mehr als zwei Jahrhunderte. Für die Newtonsche Mechanik kommt man in Gl. also mit $K = 0$ bzw. $\gamma(v) = 1$ bequem aus. Das vereinfacht die Kinematik erheblich: Die Newtonsche absolute Zeit macht sowohl die Messung der Geschwindigkeit als auch die Synchronisation von Uhren sehr einfach. Denn wenn die Zeit nicht vom Bewegungszustand abhängt, kann man sie jederzeit auf einer bewegten Uhr ablesen. Setzt man in Gl. 8.9 $K = 0$, so erhält man für die Newtonsche Mechanik die sehr viel einfachere Formel

$$u = v + u' \tag{A.1}$$

für die Geschwindigkeitstransformation, nämlich ein Additionsgesetz für die Geschwindigkeit.

So lange die Newtonsche absolute Zeit nicht falsifiziert war, war es auch vernünftig, von ihr auszugehen. Die Newtonsche Mechanik war die ökonomischste der denkbaren

Theorien und wurde zum Standardmodell der Mechanik. Im Rahmen der Newtonschen Theorie stehen Raum und Zeit unverbunden nebeneinander und haben daher auch keine gemeinsame Metrik.

Die Falsifizierung der Newtonschen Hypothese der absoluten Zeit hat zwei Konsequenzen:

1. Raum und Zeit nicht mehr unabhängig voneinander, sondern zur Raumzeit vereinigt. Die Theorie der Raumzeit ist damit die erste Vereinigung zweier Theorien, der Theorie des Raumes (homogen, isotrop) und der Zeit (homogen) zu einer gemeinsamen (Relativitätsprinzip) Theorie über eine Strukturkonstante (c).
2. Mit ihr ist zugleich die komplette Newtonsche Mechanik falsifiziert, weil sie auf eben dieser Hypothese beruht.

B. Probleme des intergalaktischen Handels und deren künftige Lösung

In Star Trek gelten die Karemma als die mächtigsten Händler der Galaxis. Um irdische Elefanten für die Zoos diverser Planeten verkaufen zu können, benötigen sie die Daten dieses größten landlebenden Säugetiers. Die von terranischen Biologen ermittelten Daten könnten sie aus einem terranischen Tierhandels-Katalog entnehmen:

1 Elefant	Masse	$m = 4 \text{ t} = 4000 \text{ kg}$
	erwartete Lebensdauer	$\tau = 80 \text{ a} \approx 2.5 \cdot 10^9 \text{ s}$
	Länge	$\ell = 10 \text{ m}$
	Höhe	$h = 3 \text{ m}$

Auf Karemma wird man sich jedoch verzweifelt fragen, was man mit diesen Angaben anfangen soll, denn auf Karemma ist das gesetzlich vorgeschriebene Längenmaß das Mips, die Masse wird in Quart gemessen und die Zeit Karaaar. Wieviel Mips sind 1 m, wieviel Quart sind 1 kg und wieviel Karaaar sind 1 s? Praktischerweise befindet sich der Regierungssitz der „Vereinten Föderation der Planeten“ in Paris, wo auch das „Bureaux Interplanetaire des Poids et Mesures“ eingerichtet ist. Das muss sich nun erst einmal mit dem „Büro für Eichmaße auf Karemma“ über die Umrechnung der Maße verständigen.

Dazu ermitteln die Karemma das Zeitmaß genauso wie auf der Erde mit dem gleichen Cäsium-Isotop in der Einheit Karaaar. Damit ist die Umrechnung Sekunde/Karaaar bekannt. Beide Büros werden sinnvollerweise die Raum/Zeitkonstante c in ihren jeweiligen Maßeinheiten durch eine Definition festgelegt haben. Sobald sie diesen Zahlenwert dem jeweiligen anderen Büro bekanntgegeben haben, kann man Mips in Meter umrechnen.

Mit dem Urkilogramm in Paris ist das eine schwierigere Sache. Da muss man entweder eine Kopie des Urkilogramms zum Karemma-Planeteten transportieren, was wegen der langen Reisezeit eher inakzeptabel ist, oder aber ein neues intergalaktisches Maßsystem definieren. Derzeit ist man in der Tat bei der Physikalisch-Technischen Bundesanstalt



Abbildung 69: Silizium-Kugel für die Kilogramm-Definition (Avogadroprojekt, Physikalisch-Technischen Bundesanstalt in Braunschweig). Die Frage, ob sich damit die Basis für ein künftiges intergalaktisches Handelsabkommen legen lässt, konnte auf der letzten Conférence Générale des Poids et Mesures (Paris, 2014) noch nicht entschieden werden.

in Braunschweig im Rahmen des Avogadro-Projekts bestrebt, ein intergalaktisch verwendbares Einheits-Kilogramm zu definieren, welches durch die Masse einer bestimmten Anzahl von Siliziumatomen festgelegt werden soll (Abb. 69).

Nachdem nun das Einheitenproblem gelöst ist, wenden wir die Erkenntnisse der Speziellen Relativitätstheorie nun auf intergalaktischen Zoo-Business an. Auf der Erde sei ein Elefant im Zoo, der von den Maßen her den Daten des Tierhandelskatalogs entspricht. Ferner soll die Erde mit einer Geschwindigkeit von $v = (\sqrt{8}/3)c$ auf Karemma zurasen, und zwar entlang der Fuß-Kopf-Richtung des Elefanten. Captain Spock, der auf Vulkanien Physik studiert hat, kommt dann zu folgenden Schlüssen vom Standpunkt des Karemma-Planeten aus:

Da seine Länge orthogonal zur Bewegungsrichtung der Karemma gemessen wird, ergeben ihre Messungen – von Schwanzspitze zu Rüsselende – auch für sie einen 10 m langen Elefanten. Für die anderen Angaben rechnet er erst einmal den γ -Faktor aus:

$$\gamma = \frac{1}{\sqrt{1 - (8/9)}} = 3$$

Damit ergibt sich eine niedliche Höhe von $\ell' = 1\text{m}$ und für die vom Karemmaplaneten aus prognostizierte Lebensdauer eines auf der Erde lebenden Elefanten $\tau' = \gamma\tau = 240\text{a}$.⁵¹ Im Vergleich zu Längen und Zeiten hat die Masse die außergewöhnliche Eigenschaft, dass sie eine Invariante der Lorentztransformation ist. Deshalb hat der Elefant im irdischen Zoo auch bezüglich der Karemma eine Masse von 4 t.

Wenn der Elefant verkauft und in einen Zoo auf dem Karemma-Planeten verbracht ist. Dann hat er im dortigen Zoo exakt die gleichen Daten, die im irdischen Tierhandels-

⁵¹Bewegung wirkt sich für Elefanten gewissermaßen wie eine Frischhaltefolie aus. Darüber sollte die Lebensmittelindustrie einmal nachdenken ...

Katalog stehen. Für einen Karemna sieht ein Elefant im Karemna-Zoo exakt genauso aus, wie dieser Elefant für einen Erdenbewohner im irdischen Zoo ausgesehen hat. Aber wenn die Erdenbewohner nun von der Erde aus den Elefanten auf Karemna beobachten, lebt der für sie $\tau = 240$ Jahre und ist nur $\ell = 1$ m hoch: Auch Elefanten sind relativ!

C. Das Zwillingsparadoxon

Das Zwillingsparadoxon gehört zu den etwas schwierigeren Problemen der Speziellen Relativitätstheorie. Es lässt sich im Rahmen der Kinematik nicht auflösen, sondern erst im Rahmen der Dynamik, weil man dabei den gerade eingeführten Begriff des Schwerpunktsystems (Kap. 46.1) braucht.

Was ist das Zwillingsparadoxon? Die Definition der Gleichzeitigkeit bzw. die Synchronisation der Uhren spielt eine wichtige Rolle für die in Kap. 8.6 angesprochene Zeitdilatation. Aber die Situation ist und bleibt nun einmal symmetrisch, wie man es von der Relativitätstheorie auch erwarten sollte. Daher kann man kinematisch nicht sagen, welche Uhr eine dilatierte Zeit aufweist. Das ZWILLINGSPARADOXON bringt den Einwand, den man schon früh gegen die Spezielle Relativitätstheorie erhoben hat, auf den Punkt: Zwei Zwillinge D und E mögen sich zur Zeit $t = t' = 0$ am gleichen Ort $x = x' = 0$ befinden. Das sei das Ereignis A. Nun bewege sich E von D weg und kehrt wieder zu D zurück. Aus der Sicht von D sollte er wegen der Zeitdilatation jünger sein als D. Aus der Sicht von E, so wird nun argumentiert, dass sich D wegbewegt und kehrt wieder zu E zurückkehrt. Also müsste D der jüngere sein. Da aber aber objektiv festgestellt werden kann, wer jünger ist, liegt ein Paradoxon vor. Enthält die Relativitätstheorie einen Widerspruch, der sie disqualifiziert?

Zunächst einmal kann man die Theorie auf Konsistenz prüfen. Man hat ein Startereignis, ein Umkehrereignis und ein Ankunftsereignis. Der Zwilling E bewegt sich auf dem Hinweg zum Umkehrpunkt mit der Geschwindigkeit v und ruht in einem entsprechenden Ruhesystem S' . Auf dem Rückweg bewegt er sich mit der Geschwindigkeit $-v$ und ruht in einem entsprechenden Bezugssystem S'' . Das Startereignis findet für beide Zwillinge bezüglich ihres jeweiligen Ruhesystems S bzw. S' am gleichen Ort statt. Das Umkehrereignis findet bezüglich S und S'' am gleichen Ort statt. Legt man die Entfernung ℓ des Umkehrereignisses bezüglich S fest und ermittelt den Zeitpunkt der Umkehr dort auf einer relativ zu S ruhenden synchronisierten Uhr, so ergibt sich die gleiche Zeitdilatation, egal ob man die Rechnung in S, S' oder S'' ausführt: Der Zwilling E ist stets der jüngere. Das kann man nachrechnen oder es auch lassen, weil das Ergebnis einfach so sein muss: Die Relativitätstheorie wurde so konstruiert, dass sich für die Hintereinanderschaltung von Lorentztransformationen konsistente Ergebnisse ergeben.

Damit kann man das Problem nicht aus der Welt schaffen, dass man zum entgegengesetzten Ergebnis kommt, nämlich dass D stets der jüngere ist, wenn man den Ablauf aus der Sicht von E analysiert: D bewegt sich zunächst mit $-v$ relativ zu S' , d.h. sein Ruhesystem ist S'' . Am Umkehrpunkt in der Entfernung $\ell' = -\ell$ wechselt er ins Ruhesystem

S und kehrt damit zu Zwilling E zurück. Da auch Beschleunigungen relativ sind, gibt es vom kinematischen Standpunkt keinen Unterschied und man erhält einen Widerspruch. Meist wird das Zwillingsparadoxon jedoch in einer Weise folgendermaßen formuliert: *D und E sind Zwillingsbären. Eines Tages verlässt E die Höhle, um zu einem fernen Stern zu reisen. Anschließend kehrt E zu D in die Höhle zurück.*

Da sowohl D als auch E biologische Uhren darstellen, ist die Frage, wer jünger geblieben ist bzw. bei wem die Zeitdilatation auftritt, objektiv entscheidbar. Wenn die Geschwindigkeit hoch genug ist, ist der Altersunterschied so groß, dass dieser Befund zweifelsfrei durch eine Messung festgestellt werden kann.

So wie das Zwillingsparadoxon hier formuliert ist („verlässt ... die Höhle, ... kehrt ... zurück“), wird von D implizit angenommen, dass er während des ganzen Vorgangs stets in ein und demselben Galileisystem verblieben ist. Die Auflösung des Paradoxons scheint dann sehr einfach: der bewegte Bär E wird der jüngere sein, denn er hat während des Vorgangs das Galileisystem gewechselt. Die Situation für D und E scheint also nicht symmetrisch zu sein. Das liegt aber daran, dass der Vorgang nicht „neutral“ formuliert wurde. Daher soll nun eine „neutral“ formulierte kinematische Geschichte folgen:

D und E sind identische Zwillingsbären, die mutterseelenallein im leeren Weltall leben. Eines Tages beobachtet D, dass sich E von ihm wegbewegt und nach einer Weile zu ihm zurückkommt. An diesem Tag beobachtet E, dass sich D von ihm wegbewegt und zu ihm nach einer Weile wieder zurückkehrt.

Wer hier eine beschleunigte Bewegung ausgeführt hat, bzw. wer in einem Galileisystem verblieben ist und wer nicht, kann in dieser Formulierung der Geschichte nicht entschieden werden, denn auch Beschleunigung ist ein relativer Begriff und bezieht sich auf ein Bezugssystem. Damit ist auch nicht mehr von vornherein klar, wer umkehrt und wer während des ganzen Vorgangs in ein und demselben Galileisystem verbleibt. Die zentrale Frage ist also, welche Uhr wirklich als ruhend anzusehen ist. Für die beiden Bären kann man zwei extreme Standpunkte einnehmen, die beide richtig zu sein scheinen:

1. Die Situation der identischen Zwillinge ist kinematisch völlig symmetrisch. Eine absolute Beschleunigung gibt es nicht. Folglich müssen bei einem Treffen beide biologische Uhren dasselbe anzeigen.
2. Ein Zwilling ist der tatsächlich ruhende. Der andere Zwilling wurde daher definitiv beschleunigt, also geht bei letzterem die Uhr langsamer. Bei einem späteren Treffen stellt man auf ihren Uhren eine Zeitdifferenz fest.

Wie auch immer man das Problem dreht oder wendet, man kann das Zwillingsparadoxon im Rahmen der Kinematik einfach nicht klären.

Wie kann man das Zwillingsparadoxon auflösen? Nun soll auch die dynamische Seite des Paradoxons betrachtet werden. Wenn eine zunächst ruhende Uhr sich zu irgendeinem Zeitpunkt in Bewegung setzt, dann ist das nur möglich, wenn eine Wechselwirkung eingetreten ist, die den Bewegungszustand geändert hat. Die Uhren sollen nun ein abgeschlossenes System bilden und somit Impulserhaltung gewährleisten. Man betrachtet das gemeinsame Ruhesystem, das die Uhren haben, bevor die Wechselwirkung einsetzt.

Nach der Wechselwirkung verläuft dann auf allen an der Wechselwirkung beteiligten und somit bewegten Uhren eine Zeit, die relativ zur Zeit dilatiert ist, die im Ruhesystem abläuft.

Auf das Beispiel der beiden Zwillingbären im leeren Weltall angewandt, verhält es sich also so: Wenn diese identisch sind und sich aufgrund einer Wechselwirkung gegenseitig abstoßen, läuft die Lebenszeit für beide Bären genau gleich ab. Das ist von der Symmetrie her schon offensichtlich. Diese Zeit ist gegenüber einer dritten Uhr dilatiert, die im Ruhesystem abläuft. Wenn ein Bär auf der Erde verbleibt, sein Zwilling zu einem fernen Stern fliegt und zurückkehrt, dann ist der erstere wegen der hohen Masse der Erde nahezu im Ruhesystem verblieben und nach der Rückkehr des Zwilling folglich der ältere Bär.

So weit ist das Zwillingsproblem nun für lineare Bewegungen aufgeklärt. Neben den linearen Bewegungen muss man aber auch Rotationsbewegungen betrachten, denn auch Drehungen sind relativ. Wie bereits Ernst Mach einmal formuliert hat [10]:

„Bleibt man aber auf dem Boden der Tatsachen, so weiß man bloß von relativen Räumen und Bewegungen. Relativ sind die Bewegungen im Weltsystem, von dem unbekanntem und unberücksichtigten Medium des Weltraums abgesehen, dieselben nach der ptolemäischen und nach der kopernikanischen Auffassung. Beide Auffassungen sind auch gleich richtig.“

Aus rein kinematischer Sicht kann man daher nicht sagen, ob Bär A, Bär B oder beide Bären beschleunigt wären, wenn sie sich umeinander drehen würden. Rein kinematisch kann man daher nicht klären, wie es sich mit der Zeitdilatation verhält. Denn wenn es im Weltraum nur eine Sonne und eine Erde gäbe, wie sollte man denn feststellen können, ob sich die Erde um die Sonne dreht oder die Sonne um die Erde? Wenn es nur den Mond, die Sonne und die Erde aber keine Planeten gäbe, dann wäre ptolemäische Weltbild sogar zweckmäßiger als das kopernikanische. Andererseits, ob die relativistische Uhr auf der Erde oder der Sonne langsamer tickt, ist ein physikalisch messbarer Effekt, der sich z.B. darin äußern würde, dass die radioaktiven Zerfallszeiten auf der Erde und der Sonne nicht exakt gleich sein können.

Bezieht man jedoch die Dynamik mit ein, so gibt es für ein abgeschlossenes System ein spezielles ausgezeichnetes Ruhesystem. Das ist jenes, bezüglich dem der gesamte Eigen Drehimpuls null ist: das Schwerpunktsystem. Die Zeit muss also auf das Schwerpunktsystem als Referenz bezogen werden. Das Schwerpunktsystem ist ein spezielles Ruhesystem (Kap. 11.5) mit einer gegebenen Eigenzeit. Gegenüber dieser Eigenzeit ist die Zeit auf allen anderen Uhren dilatiert. Wenn man auf den gesamten Kosmos als abgeschlossenes System Bezug nehmen könnte, so könnte man sagen:

*Das generische Bezugssystem der Speziellen Relativitätstheorie ist das
Baryzentrum des Kosmos.*

In der Praxis ist das nicht brauchbar, weil man das Baryzentrum des Kosmos nicht kennt. Wenn jedoch ein Teilsystem des Kosmos mit dem Rest der Welt nicht wechselwirkt, dann kann man es als abgeschlossenes System betrachten und das Zeitmaß auf sein Baryzentrum beziehen. Wenn man ein System vielleicht nicht exakt, aber doch näherungsweise

als abgeschlossenes System betrachten kann, dann definiert sein Baryzentrum in guter Näherung das Schwerpunktsystem, auf dem die Eigenzeit abläuft.

Diese Überlegungen sind in die Definition der *temps atomique barymetrique* eingeflossen. Dies ist eine internationale Atomzeitskala, die sich auf das Baryzentrum unseres Sonnensystems bezieht. Die hier ablaufende Zeit ist sozusagen unsere beste Approximation an die unerreichbare absolute Atomzeitskala, die sich auf das Baryzentrum des Kosmos beziehen müsste. Wenn ein Bär im Baryzentrum des Sonnensystems verbliebe und sich der Zwillingbär für einige Zeit auf einen anderen Planeten unseres Sonnensystems begibt, so ist der erstere stets der ältere, wenn die Zwillinge wieder im Baryzentrum zusammenkommen.

Anders als zu Newtons Zeiten verfügt man heute über Uhren, die eine so hohen Zeitauflösung haben, dass man die Zeitdilatation nachweisen kann. 1971 wurden im sogenannten Hafele-Keating-Experiment Flugzeuge auf eine Rundreise geschickt, die eine Atomuhr an Bord hatten. Die Reisezeit dieser Borduhr wurde mit der Reisezeit auf einer Atomuhr verglichen, die auf dem Boden zurückgeblieben war. Für diese Uhren mit einer unterschiedlichen Bewegungsgeschichte zeigte sich zweifelsfrei, dass die Reisezeiten nicht gleich ausfallen, denn die auf dem Boden stationierte Uhr war auf der Erde fixiert und blieb quasi im Schwerpunktsystem. Aus den Daten des Experiments kann man aus Gl. 8.27 bzw. aus

$$c = \frac{\Delta t(\ell)}{\sqrt{\Delta t^2(0) - \Delta t^2(\ell)}} |v| \quad (\text{C.1})$$

die Raum-/Zeitkonstante c bestimmen oder, je nachdem wie man es sehen will, die Einheit Meter ableiten. Die Auswertung des Experiments ist nicht ganz so einfach wie hier vereinfacht dargestellt, weil auch die Gravitation eine Rolle spielt.

D. Relativistische Federn

Es sei hier nur kurz am Beispiel einer Hookeschen Feder (s. Kap. 12.4.1) angerissen, wie man phänomenologisch beschriebene Energiebeiträge, also potentielle Energien, in die Relativitätstheorie einbauen kann. Ihre Masse $m(x)$ setzt sich aus der Masse $m_0 = m(x=0)$ der entspannten Feder und der Massenänderung zusammen, die beim Spannen der Feder infolge der aufgenommenen Energie auftritt:

$$m(x) = m_0 + \frac{1}{2c^2} kx^2 \quad (\text{D.1})$$

Die innere Energie der Feder ist demnach

$$\mathcal{U}(x) = m_0 c^2 + \frac{1}{2} kx^2 \quad (\text{D.2})$$

Wenn die Feder sich relativ zu einem Bezugssystem mit einem Impuls p bewegt, ist ihre Gesamtenergie durch

$$E(p, x) = \sqrt{p^2 c^2 + \mathcal{U}^2(x)} = \sqrt{p^2 c^2 + m_0^2 c^4 + m_0 c^2 kx^2 + k^2 x^4 / 4} \quad (\text{D.3})$$

gegeben. In der Newtonschen Näherung kann man den letzten Term unter der Wurzel vernachlässigen. Unter Verwendung von $\sqrt{1+x} \approx 1+x/2$ ergibt sich in nichtrelativistischer Näherung

$$E(p, x) \approx m_0 c^2 + \frac{1}{2m_0} p^2 + \frac{1}{2} k x^2.$$

In der nichtrelativistischen Näherung werden Energieumsätze zwischen verschiedenen Formen der kinetischen oder potentiellen Energiebeiträge betrachtet. Da es dabei nur um Änderungen der Energie geht, ignoriert man dann oft den konstanten Energiebeitrag $m_0 c^2$.

E. Legendre-Transformation

Obwohl die Legendre-Transformation für Funktionen einer Variablen diskutiert werden kann (und der Einfachheit halber hier auch so diskutiert wird), spielt sie meist erst im Kontext von Funktionen mehrerer Variabler eine Rolle. Das ist der Grund, warum sie hier unter das Thema der Funktionen mehrerer Variabler eingeordnet wurde.

Mit f sei eine Funktion bezeichnet, welche einer Variablen $x \in L$ einer Definitionsmenge L einen Funktionswert $f(x)$ zuordnet: $f : x \mapsto f(x)$ und g sei eine Funktion, welche einer Variablen $s \in M$ einer anderen Definitionsmenge M einen Funktionswert $g(s)$ zuordnet: $g : s \mapsto g(s)$.

Zunächst besteht weder eine Beziehung zwischen den Variablen x und s noch zwischen den Funktionen f und g . Nun führen wir durch die Forderung

$$s = s(x) = \frac{df(x)}{dx}$$

eine Beziehung zwischen den Variablen x und s ein. Diese Gleichung stellt eine Koordinatentransformation dar, welche besagt, wie man die Koordinate s aus der Definitionsmenge M erhält, wenn die Koordinate x gegeben ist. Wir wollen annehmen, dass diese Koordinatentransformation bijektiv ist, d.h. dass genauso gut für jeden Wert s der Wert $x = x(s)$ eineindeutig berechnet werden kann. Durch die Funktion f wird die zu x KONJUGIERTE VARIABLE s etabliert.

Immer noch ist für $g(s)$ eine beliebige Funktion möglich. Mit ihr lässt sich durch

$$y = y(s) = \frac{dg(s)}{ds}$$

ebenfalls eine zu s konjugierte Variable y etablieren. Wegen der Beliebigkeit der Funktion g ist klar, dass i.A. $y \neq x$. Nur wenn man eine ganz spezielle Funktion $g(s)$ wählt, werden x und s gegenseitig konjugierte Variable zueinander sein.

Die Forderung, dass x und s gegenseitig konjugierte Variable zueinander sein sollen, läuft also auf die Forderung hinaus, dass der Funktion f eine ganz spezielle (konjugierte) Funktion g zugeordnet wird. Eine Zuordnung $T : f \mapsto g$, die dem kompletten Graphen f den Graphen der Funktion g zuordnet, stellt eine Transformation auf dem Funktionenraum

dar. Im vorliegenden Fall wird diese Transformation als **LEGENDRE-TRANSFORMATION** bezeichnet. Sie gehört neben der Fourier-Transformation und der Laplace-Transformation zu den wichtigsten Transformationen auf dem Funktionenraum, die in sehr vielen Gebieten der Physik eine Anwendung finden.

Wenn zwischen zwei Funktionen $f(x)$ und $g(s)$ die symmetrische Beziehung

$$f(x) + g(s) = xs \tag{E.1}$$

besteht, dann sind sie Legendre-Transformierte zueinander. Die Variable s ist bezüglich der Funktion f die konjugierte Variable zu x und umgekehrt ist die Variable x bezüglich g die konjugierte Variable zu s , d.h. es besteht die symmetrische Beziehung:

$$s = s(x) = \frac{df(x)}{dx} \tag{E.2}$$

$$x = x(s) = \frac{dg(s)}{ds} \tag{E.3}$$

Das kann man folgendermaßen einsehen: Sei s z.B. die zu x konjugierte Variable. Dann erhält man durch Differentiation von Gl. E.1 nach x :

$$\begin{aligned} \frac{df(x)}{dx} + \frac{dg(s(x))}{dx} &= 1s + x \frac{ds}{dx} \\ s + \frac{dg}{ds} \frac{ds}{dx} &= s + x \frac{ds}{dx} \\ x &= \frac{dg}{ds} \end{aligned}$$

Daraus folgt, dass umgekehrt auch x die zu s konjugierte Variable ist. Das bedeutet, dass x und s ein Paar von zueinander konjugierten Variablen darstellen, wenn f und g ein Paar von Funktionen sind, die zueinander Legendre-transformiert sind, d.h. in der durch Gl. E.1 gegebenen Beziehung zueinander stehen.

Beispiel 1: Als Beispiel sei $H(p) = p^2/2m$ betrachtet.⁵² Die zu p konjugierte Variable v ergibt sich aus $v = v(p) = dH/dp = p/m$ und infolgedessen ist die Umkehrtransformation $p(v) = mv$. Setzt man sie in H ein, so erhält man die Funktion $L(v) = H(p(v)) = mv^2/2$. Man kann leicht zeigen, dass $H(p)$ und $L(v)$ gerade zwei Funktionen sind, die gegenseitige Legendre-Transformierte sind, denn es gilt

$$H(p) + L(v) = \frac{p^2}{2m} + \frac{m}{2}v^2 = pv$$

Infolgedessen ist $p = p(v) = dL/dv = mv$ auch die zu v konjugierte Variable.

Beispiel 2: Die innere Energie $\mathcal{U}(S, V)$ und die Enthalpie $\mathcal{H}(S, \mathbf{p})$ sind Legendre-Transformierte bezüglich der konjugierten Variablen Volumen V und Druck \mathbf{p} (s. Kap. 27.2). Die Variable Entropie ist von der Legendre-Transformation nicht betroffen.

⁵²Quadratische Funktionen stellen eine Ausnahme dar. Die Legendre-Transformierte einer quadratischen Funktion ist selbst eine quadratische Funktion (aber der konjugierten Variablen).

F. Hamilton- und Lagrangefunktion

Wenn man für den Feder-Oszillator den Impuls p als extensive Variable für die Aufstellung der kinetischen Energie wählt, dann ist die sich daraus und aus der potentiellen Energie ergebende Gesamtenergie

$$H(p, x) = E_{kin}(p) + E_{pot}(x) = \frac{1}{2} \frac{p^2}{m} + \frac{1}{2} kx^2 \quad (\text{F.1})$$

eine Funktion der beiden Zustandsvariablen Impuls p und Auslenkung x . Eine Darstellung der Energie, in welcher der Impuls die unabhängige Zustandsvariable des Anteils der kinetischen Energie ist, nennt man HAMILTONFUNKTION.

Die möglichen Zustände, die bei einer stationären (konstanten) Gesamtenergie E_0 durchlaufen werden können, also die Zustände, die einem idealen harmonischen Oszillator zur Verfügung stehen, ergeben sich aus der Gleichung

$$H(p, x) = E_0, \quad (\text{F.2})$$

d.h. die Zustände liegen im Zustandsraum auf einer Ellipse (Abb. 48c), d.h. auf einer geschlossene Kurve, die mit der Zeit periodisch durchlaufen wird.

Eine Änderung der Hamiltonfunktion ist durch das vollständige Differential

$$dH = \partial E_{kin} + \partial E_{pot} = \left(\frac{\partial H}{\partial p} \right)_x dp + \left(\frac{\partial H}{\partial x} \right)_p dx \quad (\text{F.3})$$

gegeben. Es ist die Summe der beiden $\partial E_{kin} = \left(\frac{\partial H}{\partial p} \right)_x dp$ und $\partial E_{pot} = \left(\frac{\partial H}{\partial x} \right)_p dx$, welche die Änderung der kinetischen bzw. potentiellen Energie darstellen. Die partiellen Ableitungen

$$v = \left(\frac{\partial H}{\partial p} \right)_x \quad (\text{F.4})$$

$$F = - \left(\frac{\partial H}{\partial x} \right)_p \quad (\text{F.5})$$

stellen die KONJUGIERTEN ZUSTANDSVARIABLEN dar. Es handelt sich um intensive Zustandsvariable, denn sie ergeben sich gemäß Gl. F.4 bzw. F.5 als Ableitungen einer extensiven Größe nach extensiven Zustandsvariablen. Die zum Impuls p konjugierte Zustandsvariable ist die Geschwindigkeit v . Die zur Auslenkung x konjugierte Zustandsvariable ist die an der Masse angreifende und sie beschleunigende Kraft F . Wenn man von Gl. F.1 ausgeht, ergibt sich:

$$v = p/m \quad (\text{F.6})$$

$$F = F_{Hooke} = -kx \quad (\text{F.7})$$

Gl. F.6 und F.7 stellen einen Zusammenhang zwischen den Zustandsvariablen p , x , v und F her. Solche Gleichungen nennt man ZUSTANDSGLEICHUNGEN. Im vorliegenden Fall

handelt es sich um besonders einfache Zustandsgleichungen: die Geschwindigkeit ist einfach nur dem Impuls proportional und die HOOKESCHE FEDERKRAFT F_{Hook} (eine konservative Kraft!) nur der Auslenkung. Zustandsgleichungen ermöglichen, von den extensiven Variablen auf intensive Variable umzurechnen. Wegen $v = \dot{x}$ und der Newtonschen Bewegungsgleichung $F = \dot{p}$ ergeben sich aus Gl. F.4 und Gl. F.5 die HAMILTONSCHEN BEWEGUNGSGLEICHUNGEN:

$$\dot{x} = \frac{\partial H}{\partial p} \quad (\text{F.8})$$

$$\dot{p} = - \frac{\partial H}{\partial x} \quad (\text{F.9})$$

Diese gelten viel allgemeiner als für das Beispiel des Feder-Oszillators, für das wir sie hier hergeleitet haben. Für ein freies relativistisches Teilchen ist die Hamiltonfunktion beispielsweise $H(p) = \sqrt{m^2 c^4 + c^2 p^2}$. Die zum Impuls konjugierte Geschwindigkeitsvariable ergibt sich gemäß Gl. F.8 zu:

$$v = \dot{x} = \frac{\partial H}{\partial p} = \frac{\partial}{\partial p} \sqrt{m^2 c^4 + c^2 p^2} = \frac{pc^2}{\sqrt{m^2 c^4 + c^2 p^2}} = \frac{pc^2}{H(p)} \quad (\text{F.10})$$

Wenn die Hamiltonfunktion gegeben ist, kann man die Geschwindigkeit also leicht ausrechnen, denn sie ist gerade die konjugierte Variable des Impulses. Im Experiment kann man die Geschwindigkeit aber i.A. leichter bestimmen als den Impuls. Daher wäre es günstiger von einer Funktion $\mathcal{L}(v)$ der Geschwindigkeit auszugehen, deren konjugierte Variable der Impuls ist. Diese Funktion $\mathcal{L}(v)$ ist gerade die Legendre-Transformierte der Hamiltonfunktion und heißt LAGRANGEFUNKTION. Gemäß Gl. E.1 ergibt sich die Lagrangefunktion gemäß

$$\mathcal{L}(v) = p(v)v - H(p(v))$$

Setzt man hier die aus Gl. F.10 folgende Umkehrfunktion $p(v) = mv/\sqrt{1-v^2/c^2}$ ein, erhält man die relativistische Lagrangefunktion

$$\mathcal{L}(v) = -mc^2 \sqrt{1 - v^2/c^2}$$

eines freien Teilchens. Von der Lagrangefunktion $\mathcal{L}(v)$ her betrachtet ist der kanonische Impuls

$$p = \frac{d\mathcal{L}}{dv} \quad (\text{F.11})$$

gerade die konjugierte Variable zur Geschwindigkeitskoordinate v .

Im nichtrelativistischen Grenzfall $v \ll c$ erhält man

$$\mathcal{L}(v) \approx \frac{1}{2}mv^2 - mc^2.$$

Da in alle relevanten Beziehungen nur Ableitungen eingehen, ist ein konstantes \mathcal{U} unerheblich. Falls die innere Energie nicht konstant ist, sondern einen potentiellen Energiebeitrag $E_{pot}(x)$ enthält (s. Kap. D), gewinnt man im Rahmen der Newtonschen Näherung (!) mit $v = \dot{x} \ll c$ die Lagrangefunktion aus

$$\mathcal{L}(\dot{x}, x) \approx E_{kin}(\dot{x}) - E_{pot}(x). \quad (\text{F.12})$$

Es ist also nicht schwer, die Lagrangefunktion anzugeben, wenn die phänomenologischen Ausdrücke für die kinetische und potentielle Energie bekannt sind, denn in der nicht-relativistischen Näherung ist sie einfach die Differenz aus kinetischer und potentieller Energie.

Literatur

- [1] R.K.P. Zia, E.F. Redish und S.R. McKay: *Making sense of the Legendre Transform*, arXiv:0806.1147v2
- [2] <http://books.google.at/books?id=mhLVHR5QAQkC>
- [3] Galileo, Galilei: "Dialog über die beiden hauptsächlichsten Weltsysteme, das Ptolemäische und das Kopernikanische", hier nur auszugsweise und redigiert wiedergegeben.
- [4] <http://www.physics.umd.edu/perg/abp/TPProbs/Problems/OP/OP42.htm>
- [5] Hans Joachim Schlichting: „Sonntaler – Abbilder der Sonne.“ In: Praxis der Naturwissenschaften - Physik. 43/4, Nr. 19.
- [6] Frank Close: „Neutrino“, Oxford University Press, New York 2010. S. 20ff und 160
- [7] W.I. Lenin, „Die Krise der modernen Physik“ in „Materialismus und Empiriekritizismus“ (1909).
- [8] L. Boltzmann: „Über eine These Schopenhauers“ (1905).
- [9] Roger Penrose: „The Road to Reality“ (2004)
- [10] E. Mach: „Die Mechanik in ihrer Entwicklung“ (1883)
- [11] Clemens Schaefer, Max Päsler: „Einführung in die theoretische Physik“, 1. Band (1970)
- [12] U. Helmich, <http://www.u-helmich.de/index.html>

Index

- äußere Energie, 99
- Abbildung, 31
Abbildungsgleichung, 42
Aberration, sphärische, 56
abgeschlossenes System, 92
Abklingkonstante, 128
Ablenkwinkel, 24
Achse, optische, 34
Adhäsion, 152, 155
Adiabatengleichung, 193
Aggregatzustände, 142
Akkumulator, 217
Amperemeter, 219
Amplitude, 125
Anfangsbedingung, 84
Anode, 217
Ansatz, 75
aperiodischer Grenzfall, 129
Aperturblende, 55
Arbeit, 138
Astigmatismus, 56
astronomische Einheit, 16
Atomuhr, 65
Auftrieb, 143
Azimutwinkel, 16
- Bahngeschwindigkeit, 253
Basis, 235
Batterie, 216
Beleuchtungsstärke, 10
Beschleunigung, 72, 85
Bewegung, beschleunigte, 72
Bewegung, geradlinige, 67
Bewegung, gleichförmige, 67
Bewegungsgleichung, Hamiltonsche, 280
Bewegungsgleichung, Newtonsche, 136
Bezugssystem, 59
Bild, 31
Bild, physikalisches, 31
Bildungsenergie, 151
- Bildweite, 41
Blindleistung, 132
Bogenmaß, 9
Breckkraft, 46
Brechung, 25
Brechungsgesetz, Snellius'sches, 29
Brechungsindex, 26
Brechzahl, 26
Brennstoffe, 200
Brennstoffzelle, 216
Bruch, 155
- Candela, 10
Celsius, 183
chemische Elemente, 200
Clausius-Clapeyron-Gleichung, 207
Corioliskraft, 259
Coulomb (Einheit), 218
Coulombmeter, 219
- Dämpfung, 129
Dauer, 64
Dehnung, 145
Dichte, spezifische, 143
Dielektrizitätskonstante, 221
Differential, totales, 158
Dimension, 235
Dissipation, 113
Drehimpuls, 247
Drehimpuls eines Systems, 266
Drehmoment, 247
Drehsinn, 24
Druck, 147, 176
Druckausgleich, 148
- Eigendrehimpuls, 267
Eigenschwingung, 133
Einfallsebene, 24
Einheitsvektor, 237
Einstein, Albert, 72, 78
Einsteinsynchronisation, 66
Elastizität, 144

Elastizitätsmodul, 145
 elektrische Spannung, 213
 elektrisches Potential, 210
 Elektrizität, 208
 Elektrode, 216
 Elektrolyse, 217
 Elektrolyt, 216
 Elektroskop, 209
 Elektrostatik, 214
 Energie, 96
 Energie, chemische, 199
 Energie, elektrische, 210
 Energie, potentielle, 108
 Energie, thermische, 167, 175
 Energiedichte, chemische, 200
 Enthalpie, 151
 Enthalpie, freie, 199
 Entropie, 168
 Ereignis, 65
 Ereignisintervall, 79
 Erhaltungsaussage, 91
 Erhaltungsgröße, 91
 Expansionskoeffizient, Joulescher, 188

Federkonstante, 110
 Federkraft, Hookesche, 280
 Flächennormale, 24
 Flüssigkeit, inkompressible, 150
 Fluid, 142, 152
 Fluid, ideales, 146
 Freiheitsgrad, 160
 Frequenz, 125
 Funktion, periodische, 64

Gain, 130
 Galileiklassen, 72
 Galileisysteme, 72
 Gas, ideales, 203
 Gaskonstante, universelle, 202
 Gegenstandsweite, 41
 Generator, thermoelektrischer, 215
 Geschwindigkeit, 66
 Geschwindigkeit, mittlere, 67
 Gesetz, Avogadrosches, 201

Gesetz, Hookesches, 110
 Gesetz, Ohmsches, 230
 Gibbs-Energie, 199
 gleichförmige Beschleunigung, 85
 gleichförmige Kreisbewegung, 248
 Gleichgewicht, thermisches, 178
 Gleichgewicht, thermodynamisches, 170, 184
 Gleichgewichtstemperatur, 179
 Gleichgewichtszustand, 144
 Gleichgewichtszustand, mechanischer, 139
 Gleichzeitigkeit, 67
 Gleitreibung, 157
 Gleitreibungskoeffizient, 157
 Größe, extensive, 161
 Größe, physikalische, 7
 Gravitationsgesetz, Newtonsches, 254
 Gravitationskonstante, 256
 Grenzfall, Newtonscher, 106
 Grenzfall, quasistatischer, 175
 Grenzfall, relativistischer, 106
 Grenzflächenenergie, 152

Hamiltonfunktion, 279
 heiß, 178
 Heizwert, 200
 Homogenität, 58
 hydraulischer Transformator, 149
 Hypothese, physikalische, 5
 Hypothesen, 2

idealer Isolator, 229
 idealer Leiter, 229
 Inertialsysteme, 98, 100
 Inertialsysteme, Newtonsche, 257
 innere Energie, 99
 Invariante, 76
 Irreversibilität, 168
 Isolation, elektrische, 214
 Isotropie, 59

Joule, 96

Körper, starrer, 267
 kalt, 178

Kapazität, elektrische, 221
 Kathode, 217
 Kelvin, 183
 Kilogramm, 95
 Kinematik, 58
 Kinetik, 83
 Kinetik, erster Ordnung, 114
 Kinetik, nullter Ordnung, 114
 kinetische Energie, 102
 Kohäsion, 155
 Kollimation, 13
 Kompressibilität, 149
 Kompressionsmodul, 150
 Konkavspiegel, 43
 Kontinuitätsgleichung, 116
 Kontinuum, 141
 Kontinuumsmodell, 142
 Konvexspiegel, 43
 Koordinatensystem, 59
 Koordinatensystem, rechtshändig, 61
 Koordinatenvektor, 235
 Kraft, 135
 Kraft, externe, 135
 Kraft, konservative, 135
 Kreisfrequenz, 125
 Kreisprozess, 149, 161, 195
 Kreisprozess, Carnotscher, 182
 Kriechfall, 128
 Kugelkoordinaten, 60

 Längeneinheit, internationale, 7
 Laborsystem, 103
 Ladung, elektrische, 210
 Ladungserhaltung, 211
 Ladungsträger, 227
 Ladungstransport, elektrischer, 226
 Ladungstransport, elektrolytischer, 216
 Ladungstransport, elektronischer, 216
 Ladungstrennung, 211
 Lagrangefunktion, 280
 Legendre-Transformation, 278
 Leistung, 116
 Lichtfluss, 10
 Lichtstärke, 10

 Lineal, 8
 Linearkombination, 235
 Linse, 45
 Linse, sphärische, 45
 Linsenmacher-Gleichung, 46
 Lorentzgleichungen, 78
 Lorentzkontraktion, 243
 Lupe, 54

 Masse, 95
 Masse, molare, 202
 Masse, schwere, 255
 Massendefekt, 104, 105
 Massenerhaltung, 107
 Massenerhaltungssatz, 200
 Massenmoment, 267
 Massenpunkt, 141
 Materie, 98
 Materie, isotrope, 142
 Materie, kondensierte, 142
 Median, 19
 Messen, 17
 Messinstrument, 17
 Messung, 17
 Messunsicherheit, 20
 Messwert, 20
 Meter, 7, 79
 Mikroskop, 55
 Mol, 202
 Molmasse, 202
 Molwärme, 202
 Momentangeschwindigkeit, 84

 Newton, 136
 Newtonsche Gleichung, 40
 Nichtgleichgewichtszustand, 139
 nichtrelativistische Näherung, 106
 normal, 15
 Normaldruck, 183

 Objektiv, 54
 offenes System, 92
 Ohm, 231
 Ohmmeter, 232
 Okular, 54

orthogonal, 15
 Orthogonalität, 237
 Orthonormalbasis, 237
 Oszillator, 124
 Oszillator, harmonischer, 126
 Oszillator, idealer harmonischer, 124
 Oszillatoren, gekoppelte, 132

 Parallelschaltung, 220
 paraxial, 47
 Parsec, 17
 partielle Ableitung, 158
 partielle Differentiale, partielle, 279
 Pascal, 147
 Pendel, mathematisches, 126
 Periode, 125
 Periodendauer, 125
 Phase, 125, 142
 Phasenübergang, 205
 Phasenraumdiagramm, 161
 plastische Deformationen, 144
 Platte, planparallele, 26
 Poissonzahl, 145
 Potential, 159
 Potential, chemisches, 203
 Potentialausgleich, 213
 Potenzreihe, 109
 Präparation, 13
 Primärzelle, 217
 Prinzip, Archimedisches, 143
 Prisma, 26
 Prozess, adiabatischer, 174
 Pseudoskalar, 251
 Punkt, kritischer, 205
 Punktteilchen, 141
 Pyroelektrizität, 215

 Raum, linearer, 235
 Raum-Zeit-Diagramm, 67
 Raum/Zeit-Konstante, 77
 Raumwinkel, 9
 Reaktion, endotherme, 104
 Reaktion, exotherme, 104
 rechtshändig, 61

 reell, 48
 Reflexion, 23
 Reflexionsgesetz, 24
 Reibungselektrizität, 208
 Reibungsgesetz, Stokessches, 158
 Reibungskraft, 135
 relativ, 59, 66
 Relativitätsprinzip, 71
 Resonanzfall, 131
 Rotationsenergie, 268
 Ruhesystem, 99

 Schatten, 33
 Schermodul, 146
 Scherung, 146
 Schmelzenthalpie, spezifische, 205
 Schmelzkurve, 207
 Schmelztemperatur, 207
 Schwebung, 133
 Schwereparameter, 112
 Schwerpunkt, 266
 Schwerpunktsystem, 267
 Schwingfall, 129
 Schwingung, 124
 Schwingung, stationäre, 130
 Schwingungsgleichung, 127
 Sekundärzelle, 217
 Sekunde, 65
 Serienschaltung, 219
 SI-System, 8
 Skalare, 68, 252
 Skalarprodukt, 236
 Spezielle Relativitätstheorie, 72
 Spiegel, 35
 Spiegelbild, 36
 Standardabweichung, 20
 Standardbedingungen, 185
 Standardmesswert, 20
 stationär, 143
 Sterad, 9
 stigmatische Abbildung, 33
 Stoß, 89
 Stoß, elastischer, 103
 Stoß, quasielastischer, 104

Stoffmenge, 142, 201
 Strahldivergenz, 13
 Strom, 227
 Stromstärke, elektrische, 227
 Substanz, 88
 Symmetrie, 71
 Symmetrietransformationen, 260
 systematischer Messfehler, 17
 Systemzeit, 67

 Tangentialebene, 42
 Teilchen, 89
 Teleskop, 54
 Temperatur, 168, 176
 Theorie, 2
 Thermodynamik, Hauptsätze, 169
 Thermodynamik, phänomenologische, 167
 Thermoelement, 215
 Thermometer, 179, 181
 Thermometrie, 179
 Thermoskop, 179
 Torsion, 146
 Trägheitsgesetz, 257
 Trägheitskraft, 257
 Trägheitsmoment, 268
 Transformation, lineare, 237
 Transmission, 24
 Treibstoffe, 200
 Triangulation, 15
 Tripelpunkt, 182, 207

 Uhr, 64
 Ulk, 242
 Urkilogramm, 96

 Vektor, 235
 Vektor, axialer, 240, 252
 Vektor, polarer, 240
 Vektor-Kreuzprodukt, 240
 Vektoren, 68
 Vektorraum, 235
 Verbindungen, chemische, 200
 Verdampfungsenthalpie, 205
 Vergleichsmethoden, 9
 Verlustleistung, 132

 Verlustleistung, dissipative, 115
 Verstärkungskoeffizient, 130
 virtuell, 48
 Viskosität, 157
 Volt, 219
 Voltmeter, 220
 Voltmeter, elektrostatisches, 220
 Volumen, spezifisches, 162
 Vorgang, irreversibler, 114

 Wärme, 173
 Wärmekapazität, isobare, 188
 Wärmekapazität, isochore, 188
 Wärmekapazität, spezifische, 188
 Wärmekraftmaschinen, 195
 Wärmepumpen, 195
 Wärmestrom, 227
 Wärmewiderstand, 230
 Waage, 143
 waagrecht, 15
 Watt, 116
 Wechselwirkung, 88
 Weltlinien, 68
 Widerstand, elektrischer, 231
 Widerstand, ohmscher, 231
 Winkelbeschleunigung, 248, 253
 Winkelgeschwindigkeit, 248, 253
 Wirkleistung, 132
 Wirkungsgrad, 137
 Wirkungsgrad, Carnotscher, 194, 196

 Zeitdilatation, 81
 Zentrifugalkraft, 259
 Zentripetalbeschleunigung, 250
 zufällig, 18
 Zustand, 88, 160
 Zustandsdiagramm, 161
 Zustandsfunktion, 160
 Zustandsgleichung, 279
 Zustandsgleichung, allgemeine, 191
 Zustandsgleichung, Van der Waalsche, 191
 Zustandsgleichungen, 176
 Zustandsgröße, 88
 Zustandsgröße, extensive, 90, 161

Zustandsgröße, intensive, [163](#)
Zustandsgrößen, intensive, [161](#)
Zustandskoordinate, [160](#)
Zustandsraum, [160](#)
Zustandsvariable, [108](#), [160](#)
Zwillingsparadoxon, [273](#)